



Universidade Federal do Rio de Janeiro
Observatório do Valongo

TOPCAT for dummies

Aluna: Rayssa Guimarães Silva
Professora: Karín Menéndez-Delmeste

Prefácio

Esse guia é baseado em duas aulas do curso Astronomy from Archival Data da IAU, em especial a aula dada pelo Mark Taylor, criador do TOPCAT. Também foi inspirado pela minha experiência com a XIX SVO School do Observatório Virtual Espanhol. O que eu tirei de ambas experiências é que tutoriais acabam sendo muito específicos e não exploram as capacidades do TOPCAT com a devida profundidade. Ao mesmo tempo, aulas ou apresentações que focam muito em apresentar o que pode ser feito não nos ajudam a ultrapassar a barreira inicial de aprender a usar uma nova ferramenta.

Disso nasceu esse guia em um formato meio híbrido. Escrevi seções explorando especificamente as funcionalidades mais básicas e as mais usadas, mas também mencionei algumas coisas a mais para dar uma dimensão adicional das coisas que o TOPCAT pode oferecer. Meu maior objetivo com esse guia é que ele dê a você uma ideia de como o TOPCAT pode ser útil no seu trabalho, mas também te mostre por onde começar. Sobre esse último ponto, tentei exemplificar ao máximo como fazer cada uma das tarefas de forma independente de um tutorial.

Espero que esse guia te dê coragem de desbravar o manual (extremamente bem documentado!) de quase 600 páginas do TOPCAT quando for necessário porque ele e você podem fazer coisas realmente incríveis juntos.

Sumário

1	Introdução	3
2	Instalação e conceitos básicos	5
3	Tabelas	7
3.1	Abrir tabelas locais e remotas	7
3.2	Acessar e visualizar informações	8
3.2.1	Estatísticos	9
3.3	Modificar e selecionar	10
3.3.1	Expressões algébricas	10
3.3.2	Conversão de coordenadas	10
3.3.3	Criar subconjuntos	11
3.3.4	Filtros	12
3.4	Salvar	13
4	Gráficos	15
4.1	Histogramas	16
4.2	Planos	18
4.3	Projeções	22
4.4	Cubos e esferas	23
4.5	Séries históricas	24
5	Juntar e comparar tabelas	25
5.1	Concatenação	25
5.2	Crossmatching	26
5.2.1	Interno	26
5.2.2	Externo	27
5.2.3	Gráficos	28
6	Comunicação com outros programas	30
6.1	Exemplo: enviando catálogos para o SAOImage DS9	30
A	Observatórios Virtuais e TOPCAT	32
A.1	Simple Image Access (SIA) e Simple Spectral Access (SSA)	32
A.2	Table Access Protocol (TAP)	34
A.3	Vizier	37
B	STILTS	39

1 Introdução

O que é o TOPCAT (**T**ool for **OP**erations on **C**atalogues **A**nd **T**ables) e o que ele pode fazer? A resposta curta é que ele é um programa que faz tudo o que você quiser com tabelas¹. A resposta longa é que ele é um visualizador e editor de tabelas interativo que foi construído de forma a incluir diversas coisas que facilitam a vida de um astrônomo e lida especialmente bem com a exploração de tabelas muito grandes.

O TOPCAT é indicado para exploração interativa de dados, mas é especialmente útil para uma análise rápida de dados com as quais não temos familiaridade, pois ele é capaz de lidar com metadados (como os headers de FITS). Isso nos permite simplesmente abrir uma tabela – mesmo que não saibamos nada sobre ela – e conseguir visualizar e analisar os dados. Portanto, a filosofia do TOPCAT é nos poupar de todas as coisas mecânicas e entediadas que os astrônomos precisam fazer quando estão trabalhando com dados tabulares, normalmente catálogos, de forma que o astrônomo possa pensar em fazer, de fato, astronomia.

Algumas das coisas que ele faz de forma eficiente:

- Acessa tabelas grandes (milhões de linhas/centenas de colunas) rapidamente;
- Lê e salva tabelas em diversos formatos;
- Calcula estatísticos das colunas;
- Junta tabelas através de algoritmos de *crossmatching*;
- Adiciona colunas a partir de expressões algébricas realizadas com outras colunas;
- Seleciona subconjuntos de dados;
- Cria e exporta diversos tipos de gráficos;
- Se comunica com outros programas e serviços de observatórios virtuais.

O objetivo desse guia é abordar alguns desses pontos, mas antes de partir para as explicações, vale mencionar algumas coisas que o TOPCAT não faz²:

- Não lida com imagens ou espectros;
- Scripts;
- Tabelas *muito* grandes;
- Todos os formatos ASCII conhecidos pela humanidade;
- Escrever seus artigos por você 😊.

Alguns comentários sobre os pontos anteriores: ds9 e Aladin são boas ferramentas para manipulação de imagens. Como é um programa interativo, não há como escrever scripts, mas existe um pacote de *scripting* associado ao TOPCAT: o STILTS (**S**tarlink **T**ables **I**nfrastructure **L**ibrary **T**ool **S**et). Não é possível carregar tabelas muito grandes, como as bilhões de linhas do catálogo completo do Gaia, mas é possível filtrar grandes catálogos diretamente pelo TOPCAT e trabalhar com esse subconjunto. Apesar de aceitar muitos tipos de formatos, o TOPCAT não trabalha com todos e às vezes será necessário converter as tabelas por outros programas (como o Astropy). Por fim, o TOPCAT não vai fazer sua astronomia por você porque ele não faz nada realmente inteligente: ele não vai dizer o que é uma estrela ou o que é uma galáxia, nem vai determinar a constante de Hubble

¹Sério, o slogan é literalmente “*Does what you want with tables.*”

²Tiradas de uma apresentação do criador

por você. Ele apenas fornece as ferramentas para você poder manipular os dados e tentar entender o que eles estão dizendo.

2 Instalação e conceitos básicos

A instalação do TOPCAT pode ser feita de alguns jeitos diferentes dependendo do sistema operacional³, mas o jeito mais simples é usar a distribuição *stand alone* – um único arquivo java que contém todo o programa. Todas as demonstrações desse guia foram escritas usando a versão 4.8 do TOPCAT (de 21 de março de 2021) com as especificações da Fig. 1.

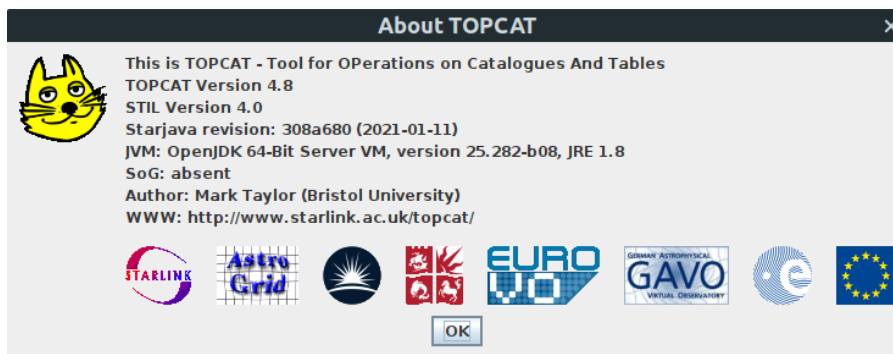


Figura 1: Especificações da versão do TOPCAT usada nesse guia. Essas informações podem ser obtidas no botão de ajuda da barra do TOPCAT (ver Fig. 3).

Independente da distribuição e do sistema operacional, a janela da Fig. 2 será a janela principal do TOPCAT e funciona como uma espécie de sumário de arquivos e processos do programa. Por padrão, todas as tabelas carregadas no programa são listadas na caixa do lado esquerdo. No lado direito há alguns controles básicos sobre as tabelas individuais: Sort Order (ordem das linhas) e Row Subset (qual subconjunto da tabela desejamos utilizar). Na parte inferior direita, na caixa Clients, estão todos os programas abertos que podem receber ou enviar dados para o TOPCAT (isso inclui outras janelas do TOPCAT!).

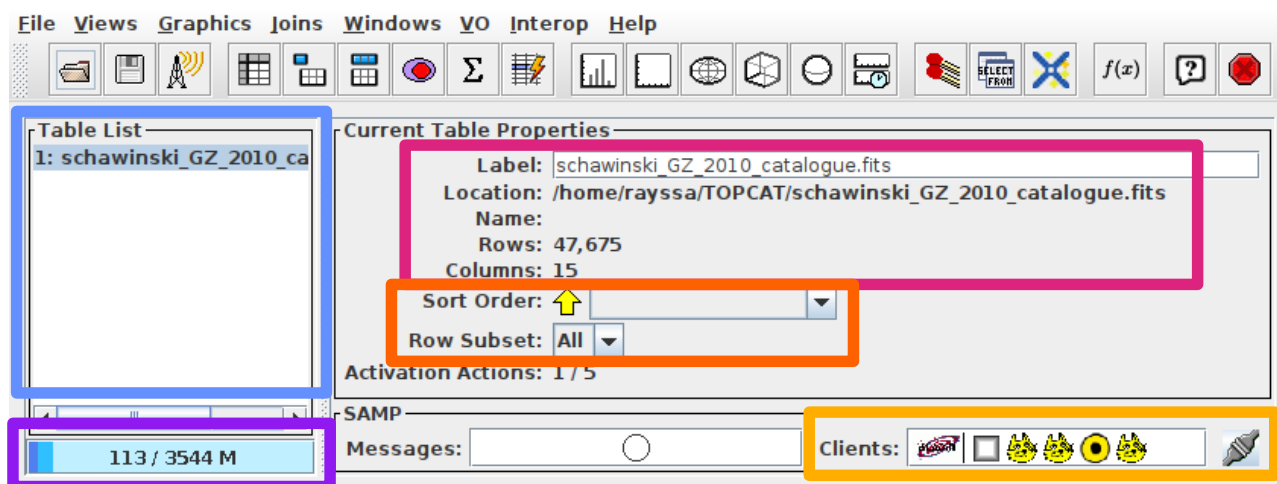


Figura 2: Janela de controle do TOPCAT. *Lista de tabelas carregadas e disponíveis, memória usada, informações básicas da tabela carregada, filtros e ordenação, programas conectados disponíveis.*

Na parte superior da janela principal, há uma barra de menu similar a da Fig. 3. Nessa barra estão as tarefas mais comuns feitas com o TOPCAT. A maior parte dos botões abre novas janelas

³Informações específicas em <http://www.star.bris.ac.uk/~mbt/topcat/#install>

que têm suas próprias barras de menu. Para descobrir o que cada botão faz em específico basta passar o mouse por cima e esperar uma frase de informação (também conhecida como *tooltip*) aparecer.



Figura 3: Barra de menu do TOPCAT. Cada caixa representa um grupo de botões que controla algum aspecto do programa. Da esquerda para a direita: **controles de ler, salvar e enviar tabelas**; **controles de exibição de dados e metadados da tabela**; **controles de visualização e criação de gráficos**; **controles de crossmatch de tabelas e de acesso a bases de dados remotas**; **botões de informação sobre as operações aritméticas, ajuda e o botão de fechar o programa**.

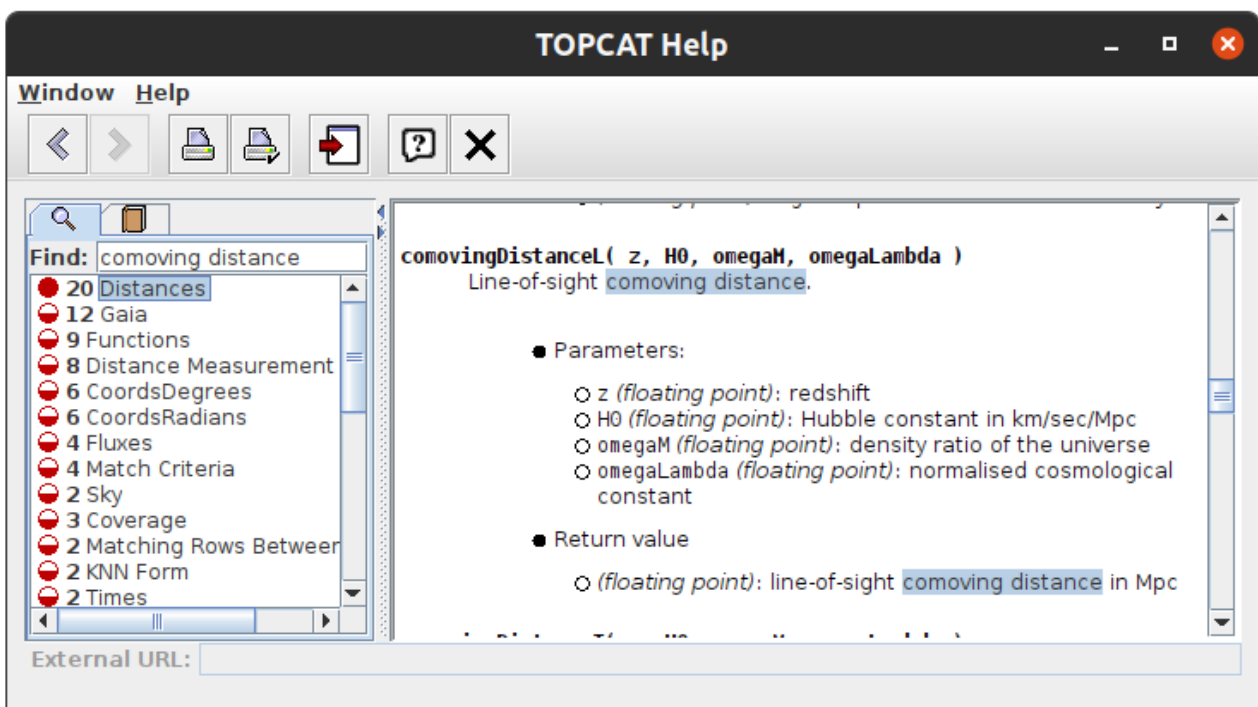



Figura 4

Por fim, a janela que se tornará sua grande amiga ao usar o TOPCAT é a janela de ajuda (Fig. 4). Ela pode ser acessada pelo menu em Help ou clicando no botão . A melhor funcionalidade dessa janela é a busca livre por termos na primeira aba (com a lupa) e o menu hierárquico na aba seguinte. Essa busca é um ótimo lugar para descobrir se o TOPCAT consegue fazer algo que você precisa.

3 Tabelas

3.1 Abrir tabelas locais e remotas

O primeiro passo para usar o TOPCAT (além de abrir o programa, claro) é carregar uma tabela. Essas tabelas podem ser locais – tabelas que estão salvas no seu computador – ou tabelas remotas – tabelas que são acessadas via serviços de observatórios virtuais. Para uma tabela local, carregá-la é tão simples quanto clicar no primeiro botão da Fig. 3. Uma nova janela (Fig. 5) aparecerá e nela você pode procurar o arquivo usando o Filestore Browser, o System Browser ou digitando o caminho completo no campo Location. Na Fig. 2 a tabela carregada é uma tabela local⁴, evidente pelo endereço no Location. Essa mesma tabela poderia ter sido carregada remotamente. Para isso, basta colocar o link para a tabela no campo de Location, como na Fig. 7.

Existem diversas outras formas de carregar tabelas remotas (Apêndice A), mas aqui só abordaremos a busca de cone (📍).

Ao clicar na busca de cone, uma nova janela aparecerá (Fig. 6). A busca pode ser feita por palavras-chave no campo keywords e o botão ao lado controla como as palavras serão combinadas; AND significa que todas devem estar presentes e OR significa que apenas algumas são obrigatórias. Na Fig. 6, buscamos por um catálogo de QSOs do Sloan Digital Sky Survey (SDSS) digitando as palavras-chave 'sdds qso' e clicamos em Find Services. Após os catálogos que atendem às palavras-chaves serem encontrados, devemos escolher um e dizer ao programa qual o cone de objetos deve ser trazido do catálogo. Aqui, escolhemos o catálogo “SDSS QSO DR7 and DR9 (D’Isanto+, 2018)” selecionamos o segundo link (Cone URL terminado em /dr7a?). Finalmente, limitamos os dados a um cone de 10 graus centrado nas coordenadas de Virgo.

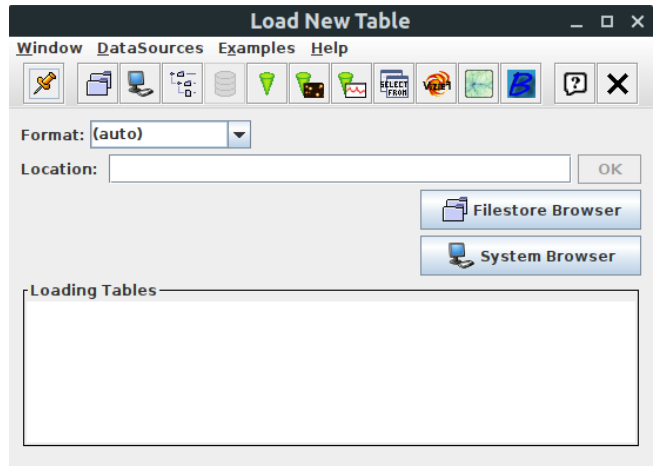


Figura 5

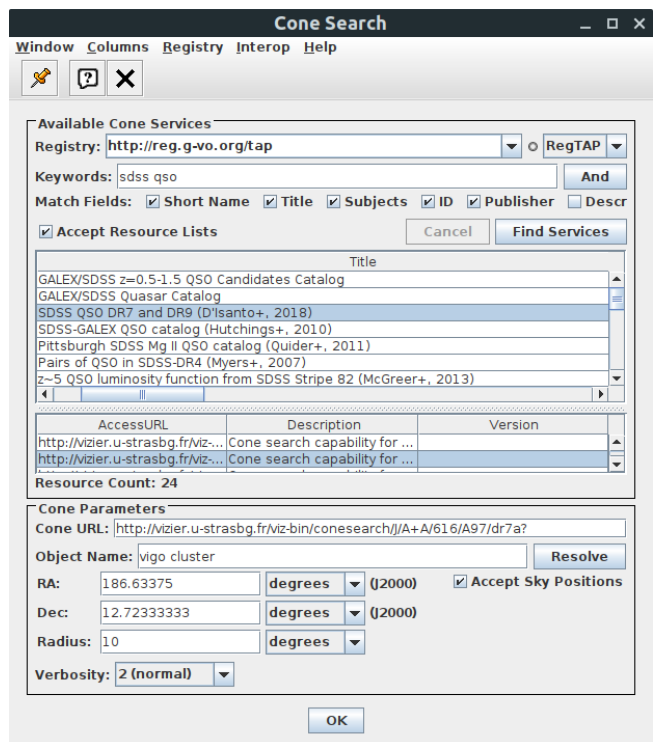
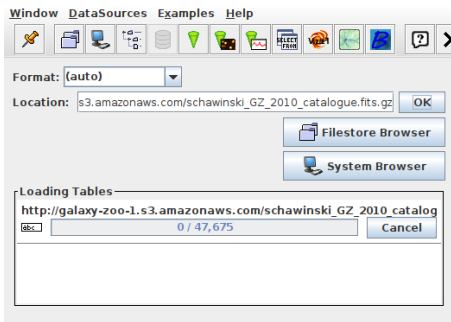
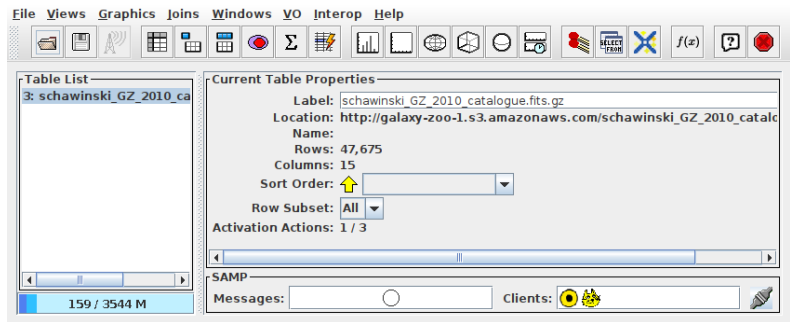


Figura 6

⁴Tabela de AGN host galaxies do Galaxy Zoo em <https://data.galaxyzoo.org/>



(a)



(b)

Figura 7

3.2 Acessar e visualizar informações

TOPCAT permite visualizar rapidamente três informações sobre uma tabela: dados, metadados da tabela e metadados das colunas. Cada uma dessas visualizações abre uma janela diferente (Fig. 8) e elas podem ser rapidamente acessadas clicando nos botões da barra de menu (Fig. 3).

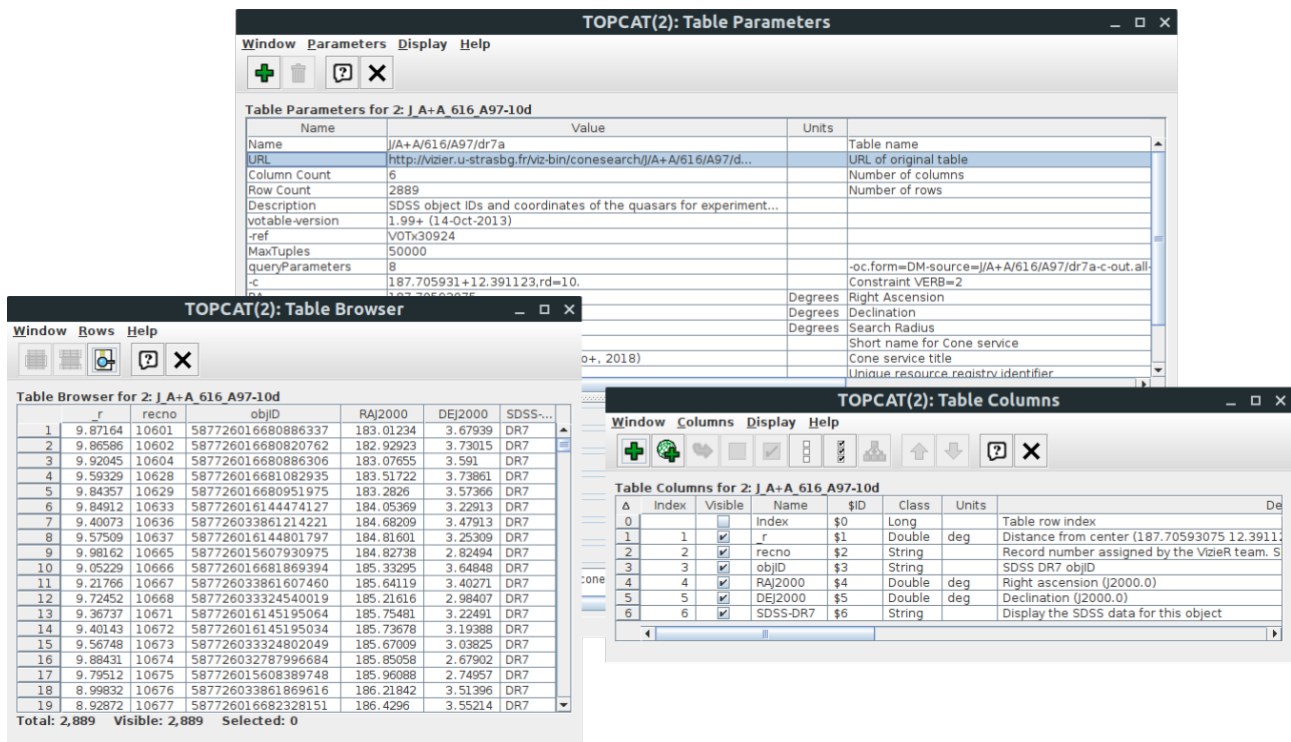


Figura 8: As três janelas que aparecem ao clicar nos três primeiros botões do segundo conjunto da Fig. 3.

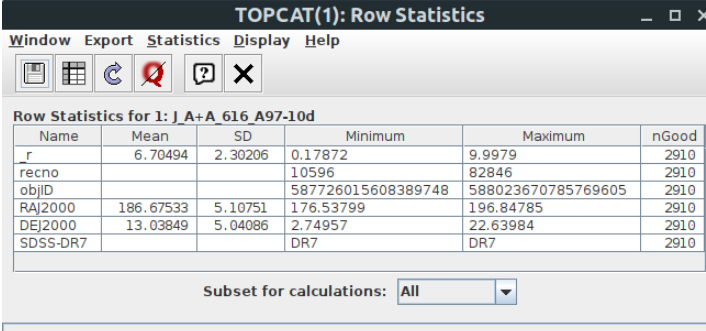
A visualização dos dados (📊) abre uma janela como a do lado esquerdo na Fig. 8 e é simplesmente uma visualização das linhas e colunas da tabela. A visualização dos metadados da tabela (📄), ao fundo no centro da Fig. 8, mostra os metadados que se aplicam à toda a tabela. Esse tipo de visualização é especialmente útil para tabelas em formatos FITS ou VOTables, pois são capazes de conter uma estrutura complexa de informação. Por fim, a visualização dos metadados das colunas (📄), do lado direito na Fig. 8, mostra informações sobre cada coluna. Na janela dessa visualização podemos controlar quais colunas queremos deixar visíveis e novas colunas.

3.2.1 Estatísticos

Outra informação que o TOPCAT oferece são os estatísticos das colunas. Para exibi-los, basta clicar em Σ na barra de menu. Os valores calculados são exibidos em uma nova janela (Fig. 9a) na forma de uma tabela. Cada linha representa uma coluna na tabela original e cada coluna apresenta um estatístico. Esses estatísticos são calculados considerando as linhas não vazias da coluna original, o número dessas linhas é dado pela coluna nGood. As seguintes colunas são mostradas por padrão:

-
- Name: O nome da coluna na tabela original;
- Mean: O valor médio das células não vazias. Para colunas booleanas esta é a proporção de células não vazias que são verdadeiras (True);
- SD: O desvio padrão da população das células não vazias;
- Minimum: O valor mínimo. Para colunas não numéricas, se as linhas puderem ser ordenadas, o valor 'menor' será mostrado. Por exemplo, células com texto mostrarão a primeira entrada definida por ordem alfabética;
- Maximum: Igual ao mínimo, mas mostra os maiores valores;
- nGood: O número de células não vazias.

Vários itens adicionais de informações estatísticas também são calculados, mas as colunas que os exibem ficam ocultas por padrão. Para exibi-las, clicamos em Display e selecionamos as colunas, e a tabela de estatísticos será atualizada, como na Fig. 9b. Para descobrir exatamente o que cada opção significa, acesse o manual em [?](#).



TOPCAT(1): Row Statistics

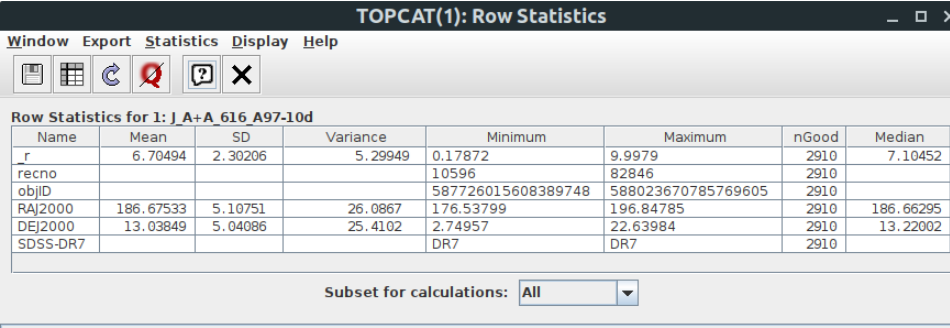
Window Export Statistics Display Help

Row Statistics for 1: J A+A 616 A97-10d

Name	Mean	SD	Minimum	Maximum	nGood
_r	6.70494	2.30206	0.17872	9.9979	2910
recno			10596	82846	2910
objID			587726015608389748	588023670785769605	2910
RAJ2000	186.67533	5.10751	176.53799	196.84785	2910
DEJ2000	13.03849	5.04086	2.74957	22.63984	2910
SDSS-DR7			DR7	DR7	2910

Subset for calculations: All

(a)



TOPCAT(1): Row Statistics

Window Export Statistics Display Help

Row Statistics for 1: J A+A 616 A97-10d

Name	Mean	SD	Variance	Minimum	Maximum	nGood	Median
_r	6.70494	2.30206	5.29949	0.17872	9.9979	2910	7.10452
recno				10596	82846	2910	
objID				587726015608389748	588023670785769605	2910	
RAJ2000	186.67533	5.10751	26.0867	176.53799	196.84785	2910	186.66295
DEJ2000	13.03849	5.04086	25.4102	2.74957	22.63984	2910	13.22002
SDSS-DR7				DR7	DR7	2910	

Subset for calculations: All




(b)


Figura 9

3.3 Modificar e selecionar

3.3.1 Expressões algébricas

As ações dessa seção normalmente serão feitas através de expressões algébricas. Com essas expressões é possível criar novas colunas a partir de outras, filtrar linhas das colunas para criar subconjuntos ou gráficos. A sintaxe é parecida com C, mas não é necessário saber nada de antemão porque a ajuda oferece uma lista das funções e exemplos de como usá-las.


Para exemplificar como usar essas expressões, vamos usar a tabela de QSOs das últimas duas seções. Primeiro, vamos ver como modificar a tabela criando novas colunas. Para isso, clicamos no botão de exibir os metadados das colunas () e, na nova janela, teremos duas opções: criar uma nova coluna a partir de colunas existentes () ou adicionar um novo par de coordenadas a partir de coordenadas existentes (). Para exemplificar a primeira opção, vamos converter a coluna `_r` de graus para sexagesimais.

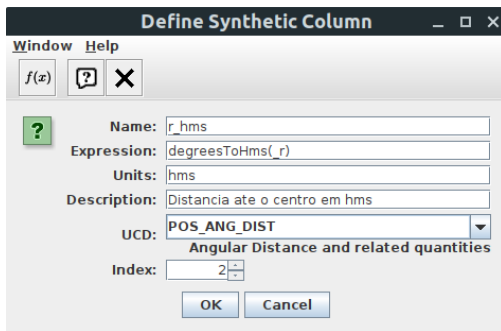
O primeiro passo é clicar em  na janela de metadados das colunas. Na nova janela (Fig. 10a) é onde vamos preencher as informações da nova coluna e os campos mais importantes são Name, Expression (é o único campo obrigatório) e Index. O Index controla a posição onde a coluna será inserida na tabela. Aqui usamos 2 para inseri-la ao lado da coluna `_r`, como pode ser visto na Fig. 10b. Em Expression, escrevemos a expressão algébrica `degreesToHms(_r)` para converter a coluna `_r` para as novas unidades.

Você deve estar se perguntando como vai descobrir qual a expressão que precisa usar. Na Fig. 10a, ao clicar no ícone $f(x)$, somos levados para a janela da Fig. 11, onde podemos encontrar informações sobre as funções (e até mesmo criar novas). De forma muito resumida, encontraremos:

- Operadores aritméticos padrão (+, -, /, *)
- Expressões condicionais (q? A: b)
- Funções matemáticas padrão (abs, max, round, sin, cos, pow, ...)
- Coordenadas do céu (graus, sexagesimais, distâncias angulares)
- Astrometria (propagação de época com/sem erros, ...)
- Distâncias cosmológicas (redshift, distância de luminosidade, lookback time, ...)
- Fluxos (Magnitudes Johnson AB, Jansky)
- Conversões de tempo (ISO8601, MJD, Julian, Besselian)
- ... e mais (e é extensível)

3.3.2 Conversão de coordenadas

Há um outro tipo de conversão mais específica acessível pela segunda opção na janela de metadados das colunas: a conversão de coordenadas. Ao selecionar o botão , uma nova janela abrirá (Fig. 12) e poderemos fazer a conversão de um par de colunas de posição para outros sistemas de coordenadas ou para o mesmo sistema, mas com outras unidades. Na Fig. 12, convertemos a unidade de graus para sexagesimais sem mudar o sistema. O resultado da conversão está na Fig. 13.



(a)

	r	r_hms	recno	objID	RAJ2000	DEJ2000	SDSS...	RA:
1	9.91353	00:39:39	10596	587726016680689706	182.60056	3.64717	DR7	12:10:
2	9.72688	00:38:54	10601	587726016680886337	183.01234	3.67939	DR7	12:12:
3	9.71032	00:38:50	10602	587726016680820762	182.92923	3.73015	DR7	12:11:
4	9.8954	00:39:35	10603	587726016680755244	182.75889	3.59974	DR7	12:11:
5	9.78624	00:39:09	10604	587726016680886306	183.07655	3.591	DR7	12:12:
6	9.49833	00:38:00	10628	587726016681082935	183.51722	3.73861	DR7	12:14:
7	9.7312	00:38:55	10629	587726016680951975	183.2826	3.57366	DR7	12:13:
8	9.83121	00:39:19	10633	587726016144474127	184.05369	3.22913	DR7	12:16:

(b)

Figura 10

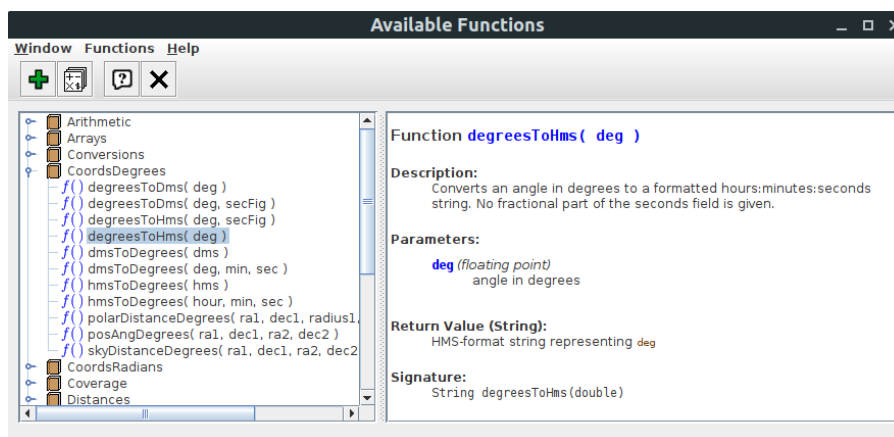


Figura 11

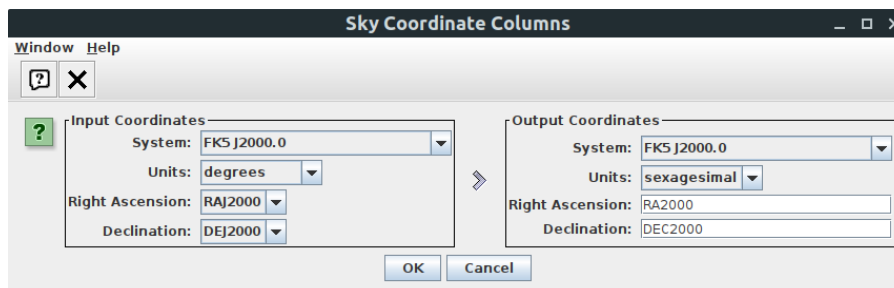




Figura 12

3.3.3 Criar subconjuntos

Existem algumas formas de selecionar dados no TOPCAT para criar subconjuntos. As principais são a seleção direta de linhas da tabela, a criação de um filtro a partir de uma expressão algébrica e selecionar pontos em um gráfico. Para usar a primeira, precisamos abrir a visualização dos dados da tabela como na Fig. 10b. Os botões de criação de subconjuntos só ficarão visíveis quando pelo menos uma linha for selecionada. É possível selecionar mais linhas ao pressionar a tecla **Shift** durante a seleção. Com as linhas selecionadas, temos duas opções: criar um subconjunto com essas linhas  ou um subconjunto com todas as linhas não selecionadas .

TOPCAT(2): Table Browser

Window Rows Help







Table Browser for 2: J A+A_616_A97-10d

	r	recno	objID	RAJ2000	DEJ2000	SDSS...	RA2000	DEC2000
1	9.91353	10596	587726016680689706	182.60056	3.64717	DR7	12:10:24.13	+03:38:49.8
2	9.72688	10601	587726016680886337	183.01234	3.67939	DR7	12:12:02.96	+03:40:45.8
3	9.71032	10602	587726016680820762	182.92923	3.73015	DR7	12:11:43.02	+03:43:48.5
4	9.8954	10603	587726016680755244	182.75889	3.59974	DR7	12:11:02.13	+03:35:59.0
5	9.78624	10604	587726016680886306	183.07655	3.591	DR7	12:12:18.37	+03:35:27.6
6	9.49833	10628	587726016681082935	183.51722	3.73861	DR7	12:14:04.13	+03:44:19.0
7	9.7312	10629	587726016680951975	183.2826	3.57366	DR7	12:13:07.82	+03:34:25.2
8	9.83121	10633	587726016144474127	184.05369	3.22913	DR7	12:16:12.89	+03:13:44.9
9	9.44351	10636	587726033861214221	184.68209	3.47913	DR7	12:18:43.70	+03:28:44.9
10	9.63941	10637	587726016144801797	184.81601	3.25309	DR7	12:19:15.84	+03:15:11.1
11	9.16554	10666	587726016681869394	185.33295	3.64848	DR7	12:21:19.91	+03:38:54.5
12	9.37217	10667	587726033861607460	185.64119	3.40271	DR7	12:22:33.89	+03:24:09.8
13	9.83973	10668	58772603324540019	185.21616	2.98407	DR7	12:20:51.88	+02:59:02.7
14	9.53813	10671	587726016145195064	185.75481	3.22491	DR7	12:23:01.15	+03:13:29.7

Total: 2,910 Visible: 2,910 Selected: 1

Figura 13

3.3.4 Filtros

A segunda forma de selecionar um subconjunto de dados é criar um filtro. Para isso, clicamos no botão  na barra de menu e uma nova janela abrirá (Fig. 14a). Nessa janela há algumas opções que nos permitem selecionar em quais linhas da tabela o filtro será aplicado: todas as linhas , n linhas no início da tabela , n linhas no final da tabela , linhas alternadas  ou o complemento das linhas selecionadas . Como exemplo, vamos selecionar todas as galáxias que estejam dentro de um raio de 5° , então vamos usar todas as linhas.

TOPCAT(2): Row Subsets

Window Subsets Display Interop Help

Row Subsets for 2: J A+A_616_A97-10d

ID	Name	Size	Fraction
1	All	2910	100%

(a)

Define Row Subset

Window Help

$f(x)$? X

Subset Name: closest

Expression: $r < 5$

OK Cancel

(b)

TOPCAT(1): Row Subsets


Window Subsets Display Interop Help


Row Subsets for 1: J A+A_616_A97-10d


ID	Name	Size	Fraction	Expression
1	All	2910	100%	
2	closest	701	24%	$r < 5$
3	not_closest	2209	76%	

(c)


Figura 14

Ao clicar em , uma nova janela (Fig. 14b) abrirá e poderemos escrever a expressão para o filtro. Nomeamos como 'closest' e filtramos pela distância ao centro ' $r < 5$ '. Filtros mais

complexos podem ser criados juntando expressões algébricas. Exemplos podem ser encontrados na [seção 7.9 do manual do TOPCAT](#) ou clicando em $f(x)$ na barra de menu e acessando o manual pelo botão  na nova janela.

Na Fig 14c (a mesma janela na Fig. 14a), o conjunto de linhas que atendem o filtro estará listado abaixo da tabela completa (All) com as informações de qual fração da tabela original esse filtro representa e qual a expressão foi usada para criá-lo. É possível alterar a expressão do filtro (refiltrar) clicando duas vezes sobre o campo Expression. A terceira linha é o conjunto complementar de linhas obtido ao clicar em .

3.4 Salvar

Para salvar as tabelas que estão abertas na sessão do TOPCAT, clicamos em . A nova janela (Fig. 15a) consiste em duas partes: a parte superior é usada para determinar qual ou quais tabelas serão salvas; a parte inferior determina onde elas serão salvas.

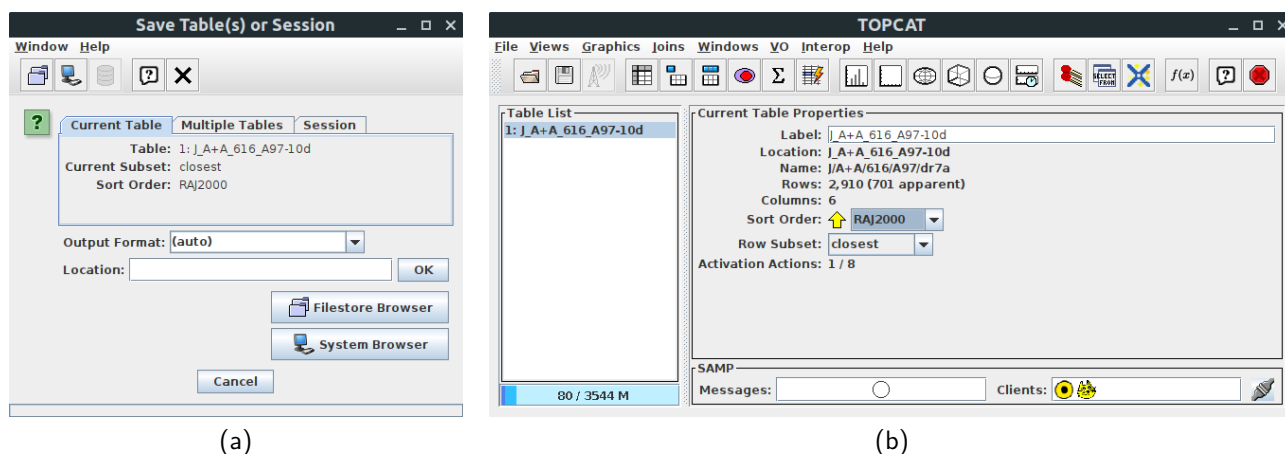


Figura 15

A aba Current Table salva a tabela selecionada na janela do controle do TOPCAT (Fig. 15b) exatamente da forma como ela é exibida. Isso significa que o TOPCAT considera quaisquer modificações feitas nos dados, seja para exibição (ocultar colunas, mudar a ordem das linhas), quando para inclusão de colunas. No caso de haver subconjuntos, apenas o conjunto selecionado em Row Subset será salvo. No caso de vários conjuntos, pode ser uma boa ideia criar uma nova coluna booleana na tabela para indicar a quais outros subconjuntos as linhas pertencem. Outra forma de salvar vários subconjuntos é usando a aba Session.

A opção Session para salvar a sessão nos permite salvar muitas das informações sobre as tabelas carregadas em sua sessão TOPCAT atual. Ao contrário das opções Current Table e Multiple Tables, a tabela é salva por completo, independente das alterações para exibição. Os itens salvos incluem:

- Todas as linhas, não apenas aquelas no subconjunto atual
- Todas as colunas visíveis e ocultas, junto com sua ordem e status de visibilidade
- Todos os subconjuntos de linhas definidos
- A ordem de classificação atual







- O subconjunto de linhas atual

As tabelas são salvas como um arquivo multi-table fit plus (recomendado) ou VOTable. Esses são arquivos normais de múltiplas tabelas que qualquer leitor de FITS ou VOTable pode ler, mas contém algumas informações específicas que podem ser lidas pelo TOPCAT. O resultado é que se você salvar um arquivo desta forma e depois carregá-lo de volta no TOPCAT, as tabelas que ele carrega aparecerão no mesmo estado de quando você as salvou, em termos de subconjuntos definidos e atuais, ordem das linhas, visíveis e colunas invisíveis, e assim por diante.




Por fim, a opção Multiple Tables serve para salvar várias tabelas no mesmo arquivo. Se um formato de saída tipo FITS ou VOTable for usado, quando o arquivo for carregado novamente no TOPCAT, todas as tabelas salvas serão carregadas individualmente. Para escolher quais tabelas serão salvas, basta marcar as caixas de seleção. Mais uma vez, apenas as colunas que não estejam ocultadas do subconjunto atual serão salvas.

4 Gráficos

O TOPCAT tem seis tipos de gráficos

-  Histogramas;
-  Gráfico 2D/Plano;
-  Projeções;
-  Gráfico 3D/Cubo;
-  Gráfico 3D/Esfera;
-  Séries históricas.

Esses gráficos ajudam muito a explorar os dados, especialmente por todos os gráficos serem responsivos – qualquer mudança nas tabelas atualiza automaticamente os gráficos. Os pontos no gráfico não precisam corresponder exatamente às colunas da tabela, pois o TOPCAT aceita expressões algébricas. Os gráficos podem ser exportados em PDF, EPS, PNG, etc e é possível usar a sintaxe do \LaTeX para escrever os títulos e as legendas. Por fim, os gráficos podem ser reproduzidos através dos comandos do STILTS (ver Apêndice B).

Antes de prosseguir com a descrição de como criar e modificar cada um desses tipos de gráficos, vamos dar uma olhada na anatomia de uma janela de gráficos (Fig. 16). As janelas de gráficos são divididas verticalmente em duas partes: um painel gráfico no topo, onde vemos o gráfico em si, e um painel de controle embaixo, onde configuramos o gráfico. O painel gráfico possui uma barra de menu que dá acesso a mais uma forma de criar subconjuntos a partir da interação com o gráfico (, , ).

O painel de controle é dividido horizontalmente em duas partes também: do lado esquerdo há uma lista de controles fixos e ativos; e do lado direito ficam os painéis de detalhes de cada um desses controles. Esses painéis são diferentes para cada controle, mas costumam ter diversas abas. Embaixo do painel, há uma barra que contém informações sobre a posição do cursor no gráfico (Position), o número de pontos que está de fato no gráfico contra o número total de pontos da tabela (Count), a ajuda sobre a [navegação no gráfico](#) com os botões do mouse, e a barra de progresso.

A barra no topo do painel de controle controla a adição ou remoção dos controles ativos e muda dependendo do tipo de gráfico que estamos fazendo. Independente do tipo de gráfico, todos vão compartilhar um conjunto fixo de controles:

- Frame: controla a área, bordas e título do gráfico;
- Legend: controla o estilo e a localização da legenda;
- Axes: controla os eixos do gráfico (valores mínimo e máximo, escala, nomes, etc), mas varia

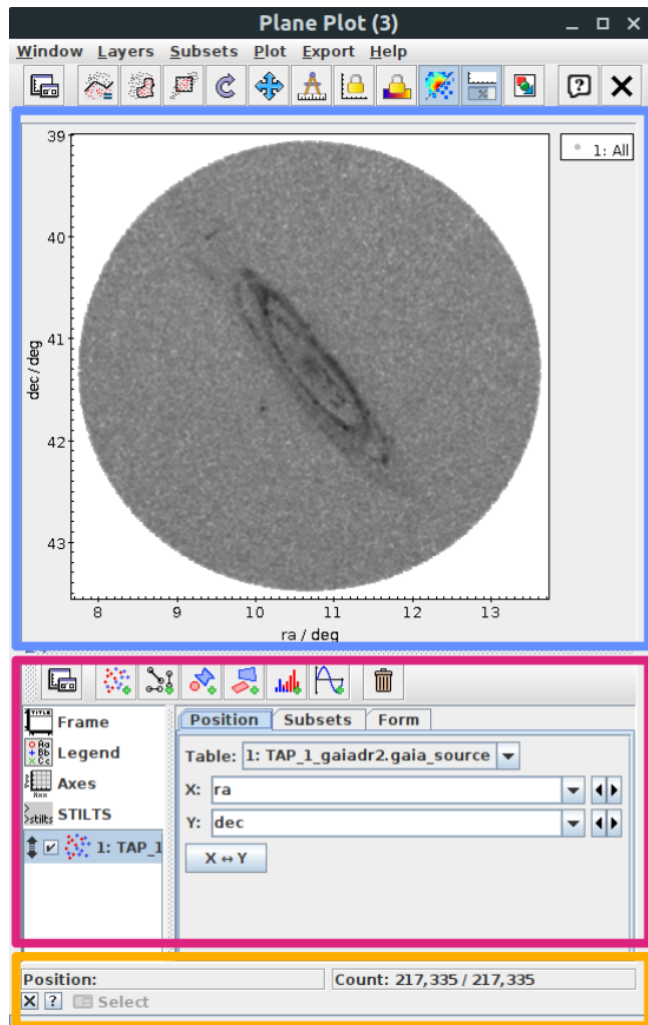


Figura 16


de um tipo de gráfico para outro;




- STILTS: mostra o comando que deverá ser usado pelo STILTS para reproduzir o gráfico atual.


Quanto aos controles ativos, na barra podemos adicionar diferentes “camadas” ou partes de um gráfico. Por exemplo, se temos um gráfico de dispersão, podemos adicionar um novo conjunto de dados de uma outra tabela em forma de linha ou histograma. Um controle ativo que muitos gráficos compartilham é o Aux Axis que controla tudo sobre mapas de cores.

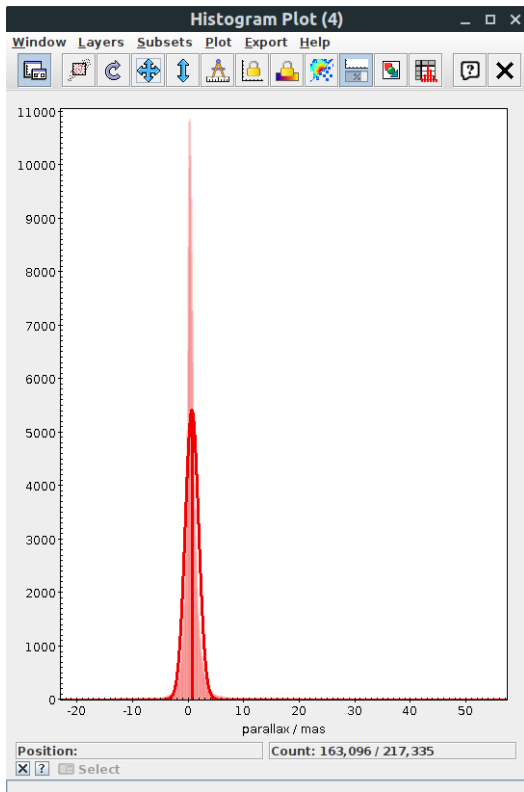
A última coisa que vale mencionar é o menu de exportação (Export) dessa janela. Nele é possível exportar não só o gráfico em diversas extensões, mas também o comando STILTS e tabelas de dados que tenham sido criadas para produzir os gráficos (como histogramas).

4.1 Histogramas

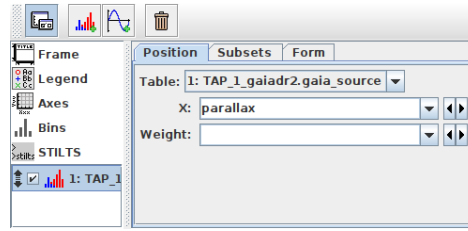
Criar um histograma (ou qualquer gráfico, na verdade) no TOPCAT é tão simples quanto clicar no botão  na barra de menu. Como exemplo, vamos criar um histograma das paralaxes estelares da tabela obtida na seção A.2. O primeiro passo após clicar no botão é especificar, no painel de controle (Fig. 17b), qual a quantidade que queremos. Nesse caso é a coluna parallax.

Na Fig. 17bc podemos ver o controle ativo  Bins no painel de controle. Esse controle é específico dos histogramas e configura diversos aspectos da binnagem dos dados. Na Fig. 17c colocamos os bins com a menor largura que o TOPCAT permitiu. No controle do histograma () , na aba Forms, podemos adicionar outros tipos de gráficos para o mesmo conjunto de dados. Vamos tentar adicionar uma gaussiana com o  Add Gaussian. Na própria aba Form, mais abaixo, o programa nos dá os parâmetros da gaussiana ajustada.

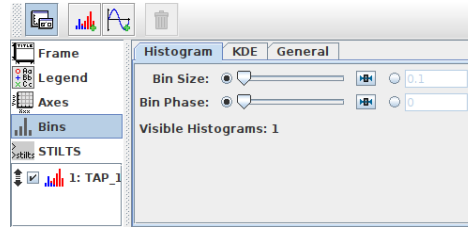
Outro tipo de gráfico que o TOPCAT faz é o “densograma”-- uma representação ao longo do eixo horizontal da densidade suavizada dos valores usando um mapa de cores. É como uma estimativa de densidade kernel (KDE) e a função usada pode ser alterada na aba KDE do controle  Bins. Na Fig. 18a temos um densograma da distribuição de RAs dos dados ao redor de Andrômeda. Sobre o histogramas estão três linhas que representam o centro e limites grosseiros das bordas dos braços espirais da galáxia. Fica claro que os picos na distribuição, representados pelo rosa no densograma, correspondem bem a essas posições. As Fig. 18b e Fig. 18c mostram as configurações principais para obter o gráfico.



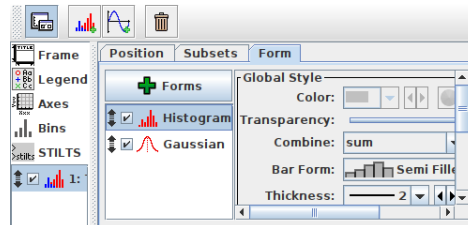
(a)



(b)

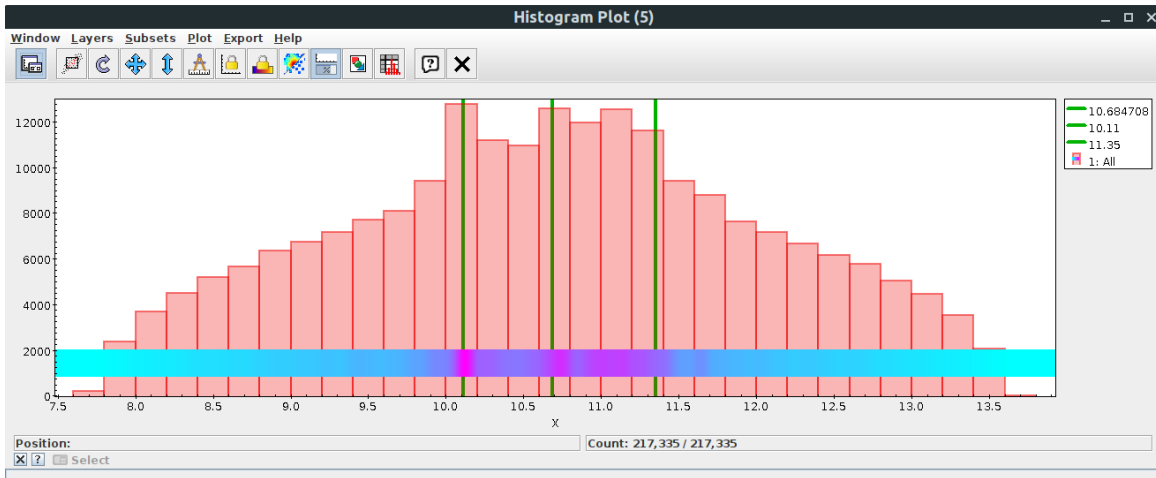


(c)

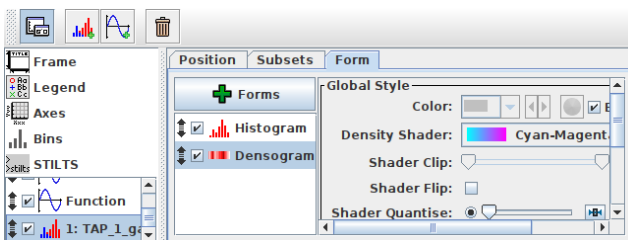


(d)

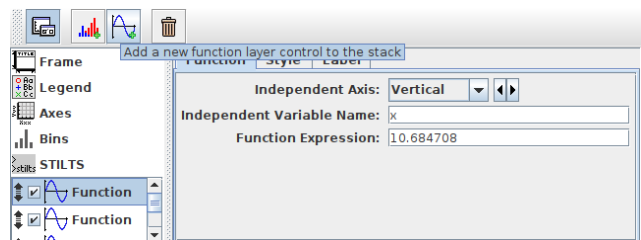
Figura 17



(a)




(b)



(c)


Figura 18

4.2 Planos

A opção Plane Plot () cria gráficos bidimensionais a partir de duas colunas, uma para X e outra para Y. Os controles específicos desse tipo de gráfico são:

- Mark: controla formato e tamanho do marcador
- Size: especifica o tamanho dos marcadores de acordo com uma coluna
- SizeXY: mesmo que Size, mas com duas colunas
- Vector: insere vetores com tamanhos XY definidos por duas colunas
- Error Bars: insere barras de erro
- XYEllipse: insere elipses (ou outras formas geométricas) definidas por três colunas
- XYCorr: insere elipses (ou outras formas geométricas) de erros e a correlação entre eles
- Polygon: insere polígonos com um número arbitrário de vértices especificados por um número arbitrário de colunas
- Line: insere uma linha que liga os dados
- Linear Fit: faz um ajuste linear de mínimos quadrados e insere a linha correspondente no gráfico
- Label: insere texto perto de cada ponto do gráfico
- Contour: insere contornos de densidade
- Grid: agrega os pontos do gráfico em células retangulares (um histograma bidimensional)
- Fill: preenche a área abaixo (ou acima ou ao lado) dos dados
- Quantile: insere uma linha com um dado quantil


Grande parte desses controles também existe para outros tipos de gráficos.

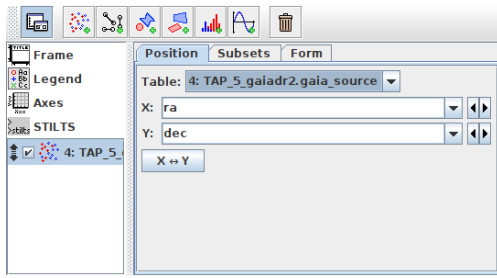
Como primeiro exemplo, vamos fazer um gráfico simples do nosso campo de visão. Clicamos em Plane Plot  e selecionamos as colunas de ra e dec (Fig. 19a) nos controles. Na aba Subsets (Fig. 19b) podemos mudar a legenda e a cor dos subconjuntos presentes, nesse caso, vamos selecionar cinza claro. Por fim, na aba Form (Fig. 19c) podemos alterar os controles de aparência do marcador. A imagem resultante está na Fig. 19d.

Para o segundo exemplo, vamos fazer um diagrama cor magnitude. Primeiro, devemos converter a magnitude aparente em absoluta. Do módulo da distância e da definição de paralaxe temos:

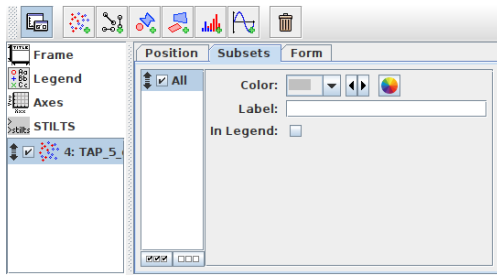
$$m - M = 5 \log d - 5 \implies M = m + 5 - 5 \log \left(\frac{1000}{p} \right),$$

onde p está em mili segundos de arco. Portanto, para X teremos a cor (BP - RP) e para Y teremos $\text{phot_g_mean_mag} + 5 - 5 * \log_{10}(1000/\text{parallax})$ (Fig. 20a). Note que o eixo Y foi invertindo usando os controles de Axes. Esse tipo de gráfico aceita dimensões adicionais, como cor. Na Fig. 20b, colorimos os pontos do gráfico de acordo com o menor valor (Combine) da magnitude g (Weight) em um dado pixel do gráfico. Os controles do mapa de cor ficam no controle Aux Axis e escolhemos a escala de um histograma por contraste (Fig. 20c).

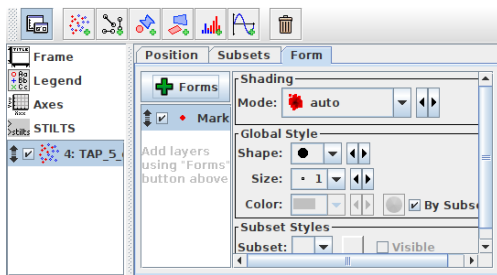
Um exemplo da responsividade do TOPCAT é exibição de um subconjunto no gráfico. Na Fig. 20b, podemos ver que a faixa de magnitudes g menor do que ~ 15 , forma uma sequência principal no diagrama. Podemos criar um subconjunto  com o filtro $\text{phot_g_mean_mag} < 15$. Esse



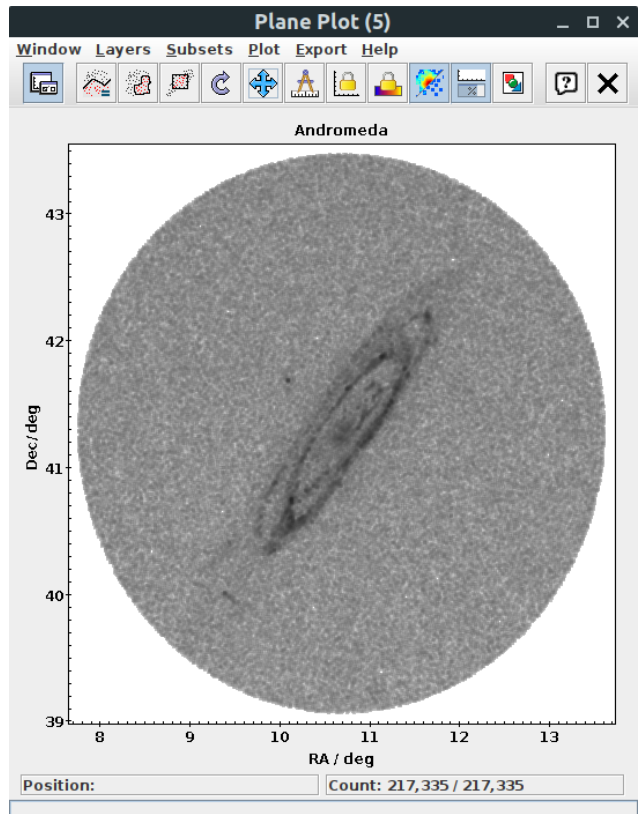
(a)



(b)




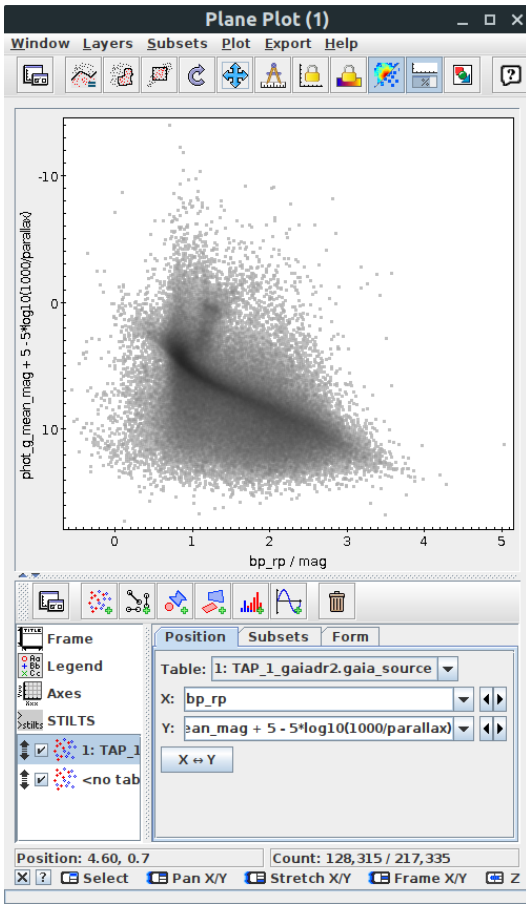
(c)



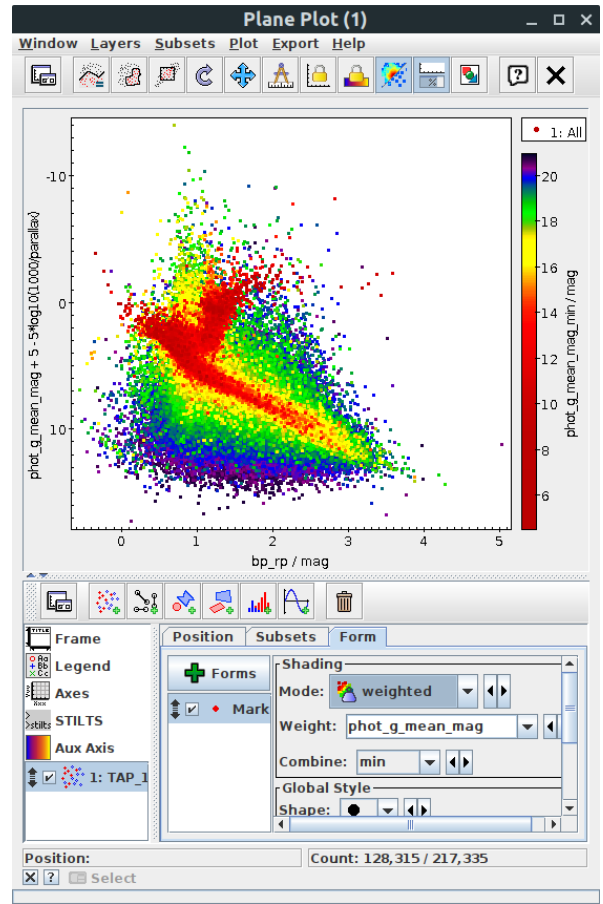
(d)

Figura 19

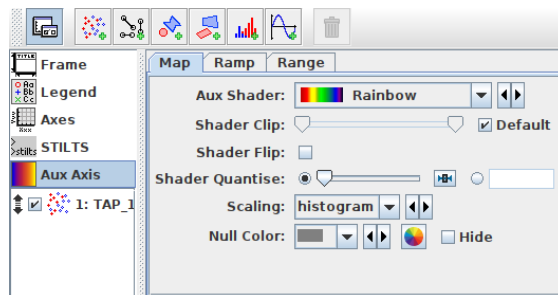
subconjunto (CMD) está em vermelho na Fig. 21a. Existem algumas formas de criar subconjuntos graficamente, na Fig. 21b usamos o botão  para criar um subconjunto à mão livre, os pontos correspondentes estão em amarelo na Fig. 21a.



(a)

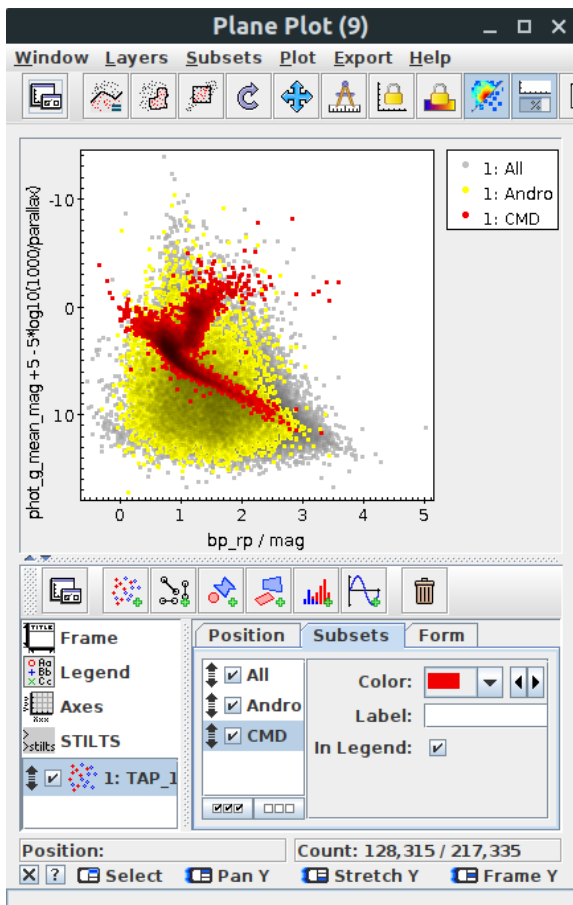


(b)

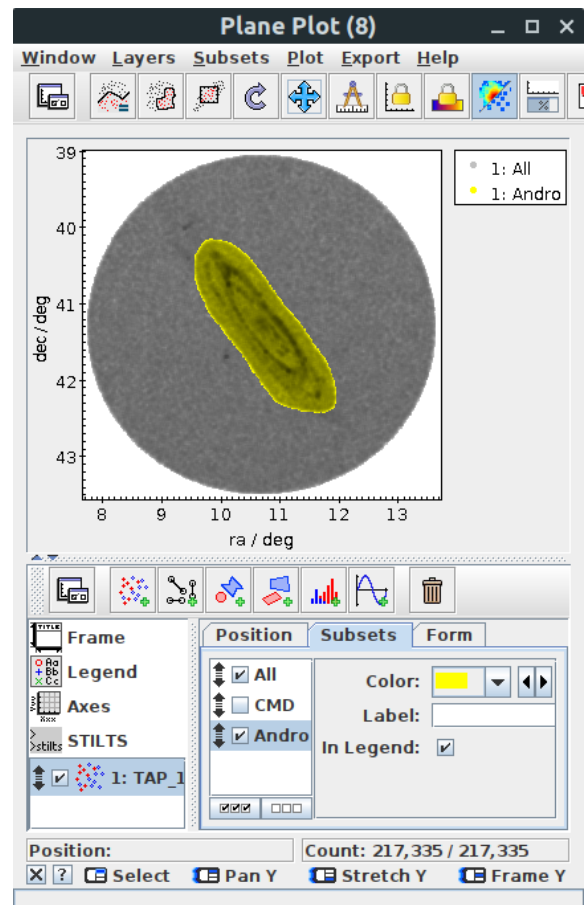


(c)

Figura 20



(a)



(b)


Figura 21

4.3 Projeções

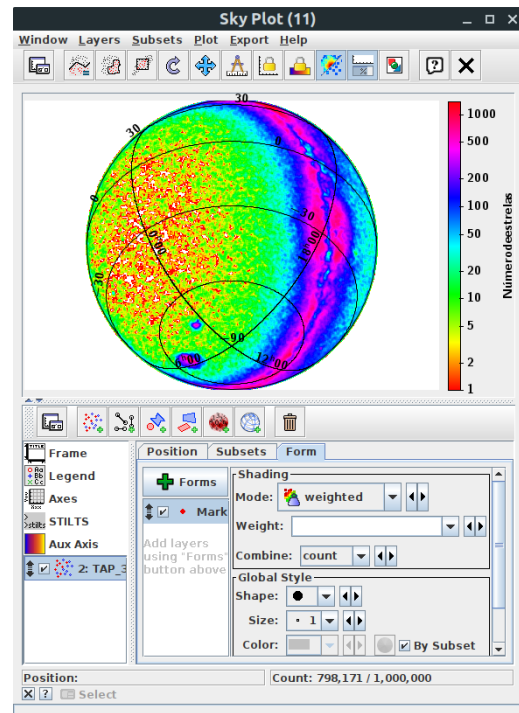
Bem parecido com o Plane Plot, o Sky Plot também cria gráficos bidimensionais, mas com dados de entradas de posições (longitude e latitude) na esfera celeste. Esse tipo de gráfico é otimizado para facilmente converter entre diferentes sistemas de coordenadas e diferentes tipos de projeções.

Para exemplo, vamos usar dados do Gaia obtidos como na seção A.2, mas com o seguinte comando ADQL

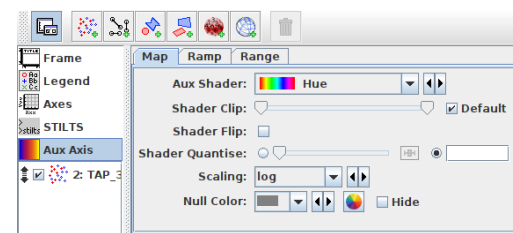
```
SELECT TOP 1000000 ra, dec, parallax
FROM gaiadr2.gaiia_source
ORDER BY random_index
```

Ao clicar em , o TOPCAT vai criar um gráfico a partir de duas colunas de posição e o sistema em que essas coordenadas são dadas (Data Sky System). Esse sistema é apenas o das coordenadas dos dados de entrada, o sistema e a projeção exibidas pelo gráfico são controlados em outro lugar! Assim como o Plane Plot, esse tipo de gráfico também aceita dimensões adicionais e colorimos pelo logaritmo da quantidade de estrelas (Fig. 22b), resultando na Fig. 22a.

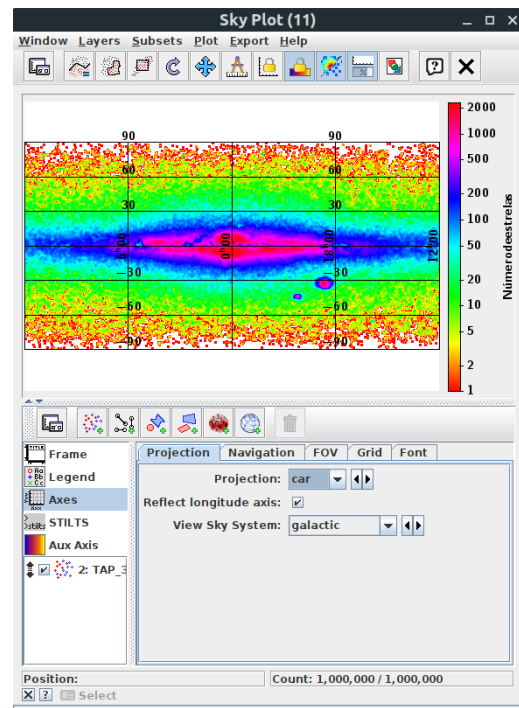
Para especificar o sistema e a projeção nas quais desejamos visualizar os dados, devemos ir no controle Axes e na aba Projection. Na Fig. 22a os dados estão no mesmo sistema que a tabela (equatorial) e na projeção padrão do TOPCAT (seno ou Sin). Há outros tipos de projeções: Aitoff (Ait) e Plate Carée (Car com lon=0 no centro do plano e Car0 com lon=0 nas bordas). Na Fig. 22, usamos o sistema galáctico na projeção cilíndrica (Car). Fica claro que o maior número de estrelas está concentrado no plano da galáxia (latitude = 0) e nas Nuvens de Magalhães.



(a)



(b)



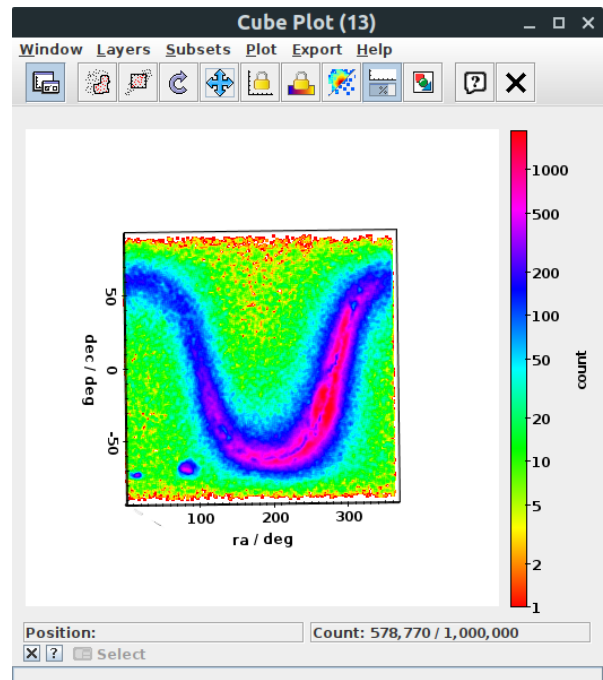
(c)

Figura 22

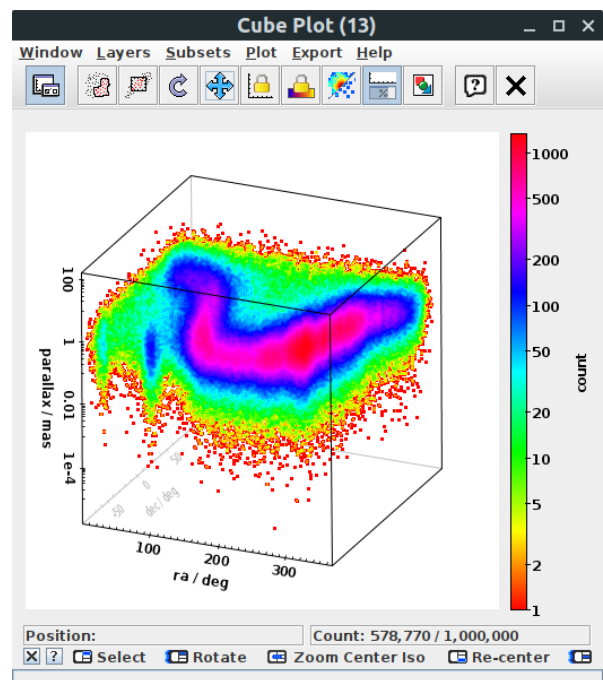
4.4 Cubos e esferas

Tanto o Cube Plot (📦) quanto o Sphere Plot (🌐) são os tipos de gráficos tridimensionais do TOPCAT. A principal diferença entre eles é o tipo de coordenada usada: no cubo, as coordenadas são cartesianas (X, Y, Z) e na esfera as coordenadas são latitude, longitude e raio. Navegar esse tipo de gráfico pode ser bem confuso, então vale a pena dar uma olhada na seção de [navegação](#) do TOPCAT.

Aqui temos dois exemplos desses gráficos com os dados do Gaia da seção anterior. A Fig. 23 é um gráfico tipo cubo e a Fig. 24 é um gráfico tipo esfera. Na Fig. 23a o gráfico foi girado de forma a mostrar apenas o plano RA vs Dec e na Fig. 23b podemos ver uma dimensão adicional com o logaritmo da paralaxe. Isso nos mostra, por exemplo, que as Nuvens de Magalhães são regiões com os menores paralaxes (distantes). Na Fig. 24b estão as mesmas variáveis, mas agora o paralaxe é representado como um raio ao invés de uma altura. Exibindo o subconjunto traçado direto na projeção do céu (Fig. 24a) fica claro que a região das nuvens realmente tem uma distribuição diferente de paralaxes quando comparada ao resto do céu.

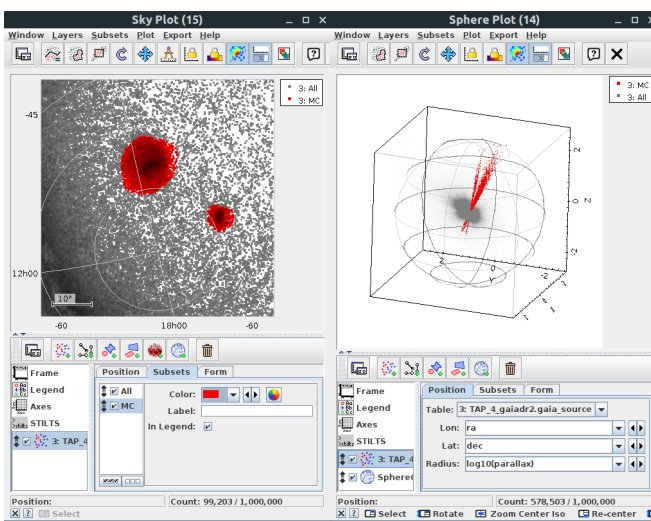


(a)



(b)

Figura 23





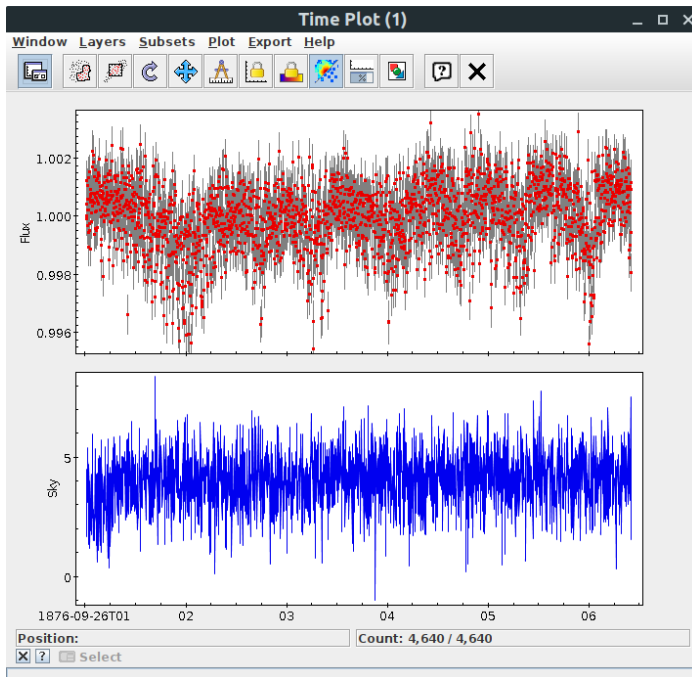
(a)

(b)

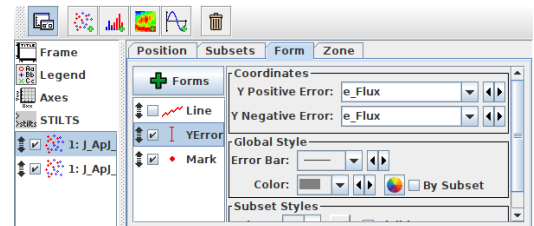
Figura 24

4.5 Séries históricas

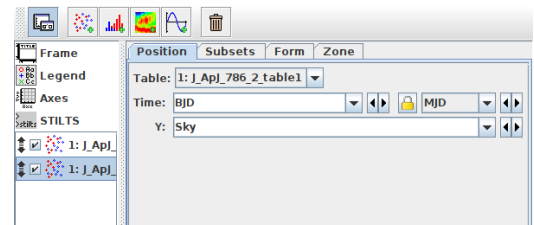
O Time Plot () faz gráficos de séries temporais. Os dados não se limitam a variáveis unidimensionais ao longo do tempo, esse tipo de gráfico também é capaz de representar espectros através do controle de espectrograma . Uma particularidade desse tipo de gráfico é que podemos ter várias "zonas" em uma mesma janela, isso significa que podemos fazer gráficos de variáveis diferentes que compartilham o mesmo eixo temporal, como na Fig. 25⁵.



(a)




(b)



(c)

Figura 25

O primeiro passo depois de clicar em  é definir a coordenada de tempo no campo Time (Fig. 25c). É importante selecionar o formato correto da coordenada temporal, ele pode ser

- DecYear: anos desde 0AD
- MJD: Data Juliana Modificada (dias desde 17 de novembro de 1858)
- JD: Dia Juliano (dias desde 1 de janeiro de 4713 AC)
- Unix: segundos desde a meia-noite de 1 de janeiro de 1970
- Iso8601: string de data e hora ISO 8601


O TOPCAT tentará adivinhar o formato correto a ser usado, mas se você pode olhar os metadados da tabela e definir manualmente.

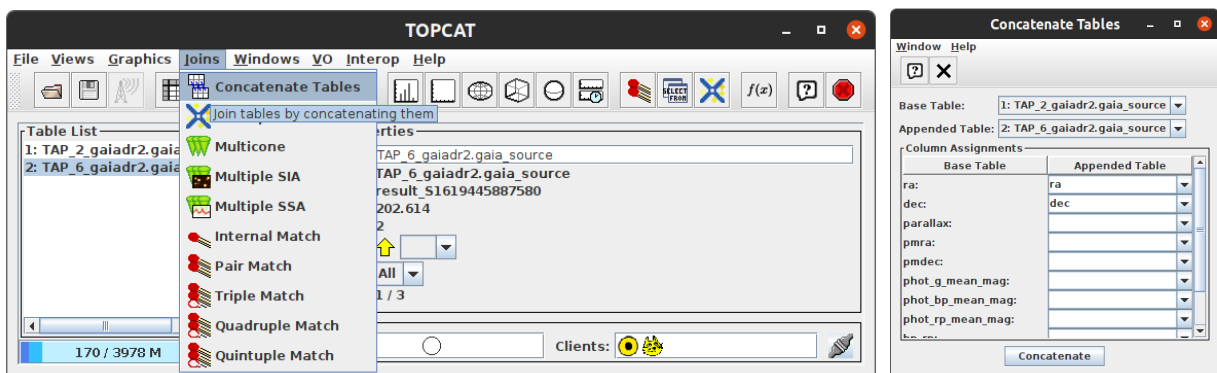
⁵Catálogo "Spitzer photometric time series of HD 97658" no Vizier (palavra-chave: J/ApJ/786/2)

5 Juntar e comparar tabelas

O TOPCAT nos permite realizar algumas operações envolvendo duas ou mais tabelas. Aqui chamei de “juntar e comparar”, pois há algumas formas diferentes de unir as tabelas. A união de cima para baixo (concatenação) vai simplesmente colar uma tabela na outra com base em algumas colunas correspondentes. Uniões “laterais” (*matching*) são mais complicadas porque dependem da ordem das linhas, então são realizadas de acordo com algum critério.

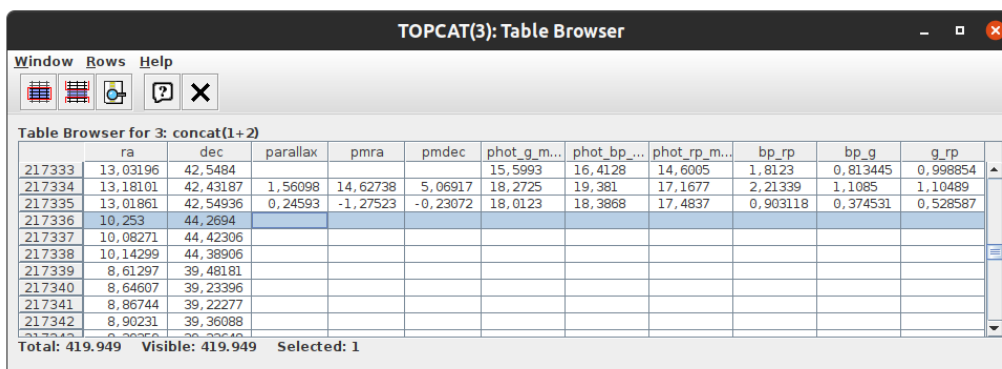
5.1 Concatenação

Para unir duas tabelas, clicamos no menu Joins e selecionamos a opção Concatenated Tables  (Fig. 26a). Na nova janela (Fig. 26b) podemos selecionar quais colunas da tabela original (coluna Base Table) correspondem às colunas da tabela que será acrescentada (coluna Appended Table). A tabela resultante terá o mesmo número de colunas da tabela original, mesmo que a tabela anexada fique com colunas vazias (Fig. 26c).



(a)

(b)



	ra	dec	parallax	pmra	pmdec	phot_g_m...	phot_bp_...	phot_rp_m...	bp_rp	bp_g	g_rp
217333	13.03196	42.5484				15.5993	16.4128	14.6005	1.8123	0.813445	0.998854
217334	13.18101	42.43187	1.56098	14.62738	5.06917	18.2725	19.381	17.1677	2.21339	1.1085	1.10489
217335	13.01861	42.54936	0.24593	-1.27523	-0.23072	18.0123	18.3868	17.4837	0.903118	0.374531	0.528587
217336	10.253	44.2694									
217337	10.08271	44.42306									
217338	10.14299	44.38906									
217339	8.61297	39.48181									
217340	8.64607	39.23396									
217341	8.86744	39.22277									
217342	8.90231	39.36088									


(c)

Figura 26

5.2 Crossmatching

O *matching* ou *crossmatching* pode ser de dois tipos: (i) interno, quando todos os arquivos que desejamos combinar estão carregados no TOPCAT; ou (ii) externo, quando uma ou mais tabelas são grandes demais para serem carregadas. Independente do tipo, os dois compartilham a mesma complicação comum desse tipo de união: como criar uma correspondência entre linhas independente da ordem em que elas aparecem nas tabelas? Na maioria das vezes, essa correspondência será feita por um ou mais critérios (*match criteria*) que podem envolver strings, posições no céu ou distâncias.

5.2.1 Interno

Para fazer um *match* interno podemos usar tanto o menu Join quanto o botão Pair match . Ambos vão nos levar para uma nova janela muito parecida com a da Fig. 27. Nessa janela podemos selecionar

- Quais tabelas devem ser combinadas;
- Quais são os critérios de correspondência;
- O que fazer com as linhas que não têm correspondência;
- Quais linhas incluir na tabela resultante.

De forma resumida, as opções de algoritmos de correspondência são

- Céu (Sky): todos os algoritmos de tipo céu vão usar posições, distâncias ou regiões no céu para testar a correspondência;
- Valor exato (Exact Value): determina uma correspondência exata entre um par de colunas;
- N-dimensionais: compara posições em um espaço cartesiano N-dimensional e são necessárias N colunas de cada tabela para criar a correspondência.

O último campo de seleção da janela (Output Rows) controla como e quais linhas serão incluídas na tabela final. O seletor Match Selection controla o que acontece quando várias de linhas de uma tabela correspondem a uma única linha na outra. Existem quatro opções:

- All matches: cada correspondência entre as duas tabelas é incluída no resultado e as linhas de ambas as tabelas de entrada podem aparecer várias vezes no resultado;
- Best match, symmetric: os melhores pares são selecionados de uma forma que qualquer linha de entrada que apareça em um par de resultados é desqualificada de aparecer em qualquer outro par de resultados, então cada linha de ambas as tabelas de entrada aparecerá em no máximo uma linha no resultado;
- Best match for each Table 1 row: apenas as melhores correspondências da tabela 2 para cada linha da tabela 1 aparecerão no resultado. Cada linha da tabela 1 aparecerá no máximo uma vez no resultado, mas as linhas da tabela 2 podem aparecer várias vezes.

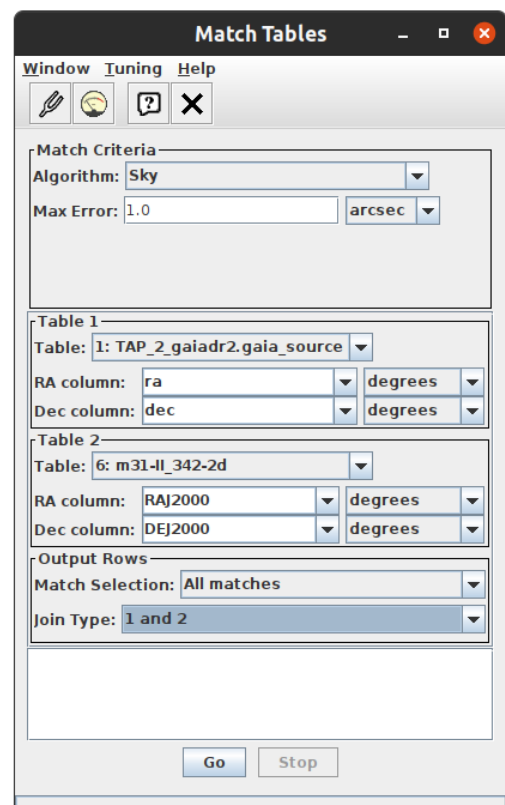


Figura 27

- Best match for each Table 2 row: mesmo que o anterior, trocando as tabelas.

A “melhor” correspondência geralmente significa a “mais próxima”. As diferenças entre as várias opções de seleção vão ser importantes nos casos em que o algoritmo de correspondência pode não ter um bom desempenho. Por exemplo, em campos com muitas estrelas, a distância entre os objetos em uma ou ambas as tabelas é próxima ou menor do que o raio de correspondência especificado. Nesse caso, uma das opções assimétricas geralmente é a mais apropriada.

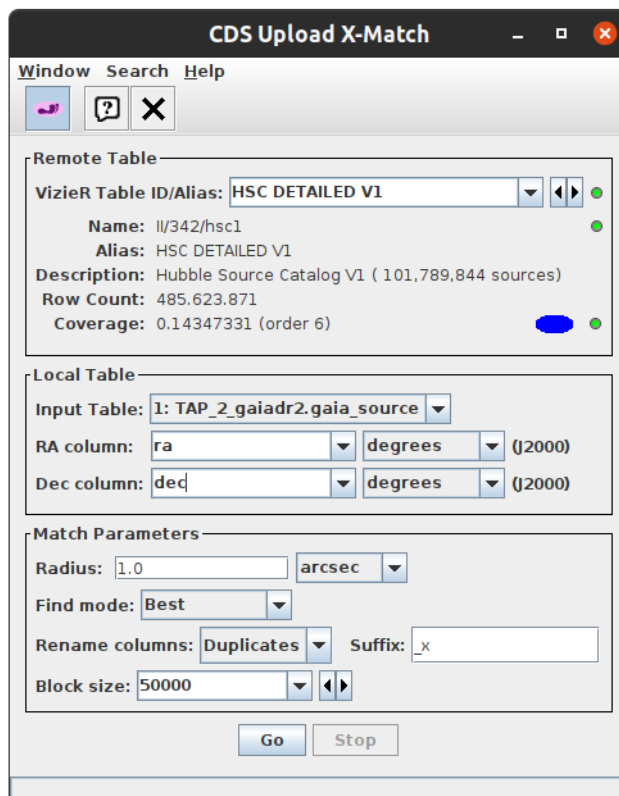
Por fim, o tipo de união (Join Type) permite escolher quais linhas de saída vão estar na tabela final.

- 1 and 2: a tabela de saída contém apenas as linhas que estão em ambas as tabelas de entrada de modo que cada linha de saída representa um par correspondente real;
- All from 1: todas as linhas com correspondência estão presentes na saída, mais as linhas sem correspondência da tabela 1;
- All from 2: mesmo que o anterior para a tabela 2;
- 1 or 2: todas as linhas das duas tabelas, com e sem correspondência, aparece na saída
- 1 not 2: todas as linhas da primeira tabela que não têm correspondências na tabela 2;
- 2 not 1: mesmo que o anterior para a tabela 2;
- 1 xou 2: é o “ou exclusivo” e a saída contém apenas linhas da primeira tabela que não têm correspondências na tabela 2 e vice-versa.

5.2.2 Externo

O crossmatching externo é feito entre uma tabela local e tabelas remotas disponibilizadas por serviços de Observatórios Virtuais ou similares. O serviço mais versátil é o X-Match, provido pelo CDS (Centre de Données astronomiques de Strasbourg), que dá acesso a crossmatching com tabelas do VizieR e do SIMBAD. Nesse caso, realizamos o crossmatching pelo menu VO com a opção CDS Upload X-Match (X). Na nova janela (Fig. 28), o primeiro bloco de seleção controla as opções da tabela remota. Em VizieR Tables ID/Alias podemos selecionar em qual catálogo desejamos buscar nossas fontes. O segundo bloco, Local Tables, controla quais são as colunas de posição das fontes na nossa tabela local. Por fim, no bloco de Match Parameters, temos os seguintes seletores e suas opções

- Radius: a distância máxima entre as posições da nossa tabela local e da tabela remota que um par pode ter para ser con-



(a)

Figura 28


siderado um match;

- Find mode: determina a forma como o resultado é gerado e usado
 - Best: cria um tabela com uma linha para cada linha da tabela local que tenha um match (as linhas sem match não são incluídas);
 - All: mesma coisa que Best, mas simétrico – cada linha de cada tabela só pode aparecer uma vez;
 - Each: cria uma nova tabela com a mesma ordem e número de linhas que a tabela local e cada match da tabela remota, caso exista, é incluído na linha;
 - Best Remote: cria uma tabela com todas as linhas da tabela remota que tenha um match com a tabela local – para usar esse modo o Block size deve ser maior do que o número de linhas da tabela remota;
 - Add Subset: nenhuma tabela nova é criada, mas sim um subconjunto das linhas que tem pelo menos uma contraparte na tabela remota;
- Rename columns: caso as colunas das tabelas tenham o mesmo nome, as colunas da tabela remota podem ser renomeadas de acordo com o sufixo especificado, mas também é possível renomear as colunas com o sufixo sempre;
- Block size: número de linhas que o serviço processa de uma vez.

Há outras opções de crossmatching, como o multi-SIA e o multi-SSA. Essas opções nos permitem buscar arquivos de imagem e espectros remotos. Seu uso é bem similar ao SIA e SSA para uma única fonte descritos na seção A.1, com a diferença de que podemos usar as colunas da tabela como múltiplas posições para objetos.

5.2.3 Gráficos

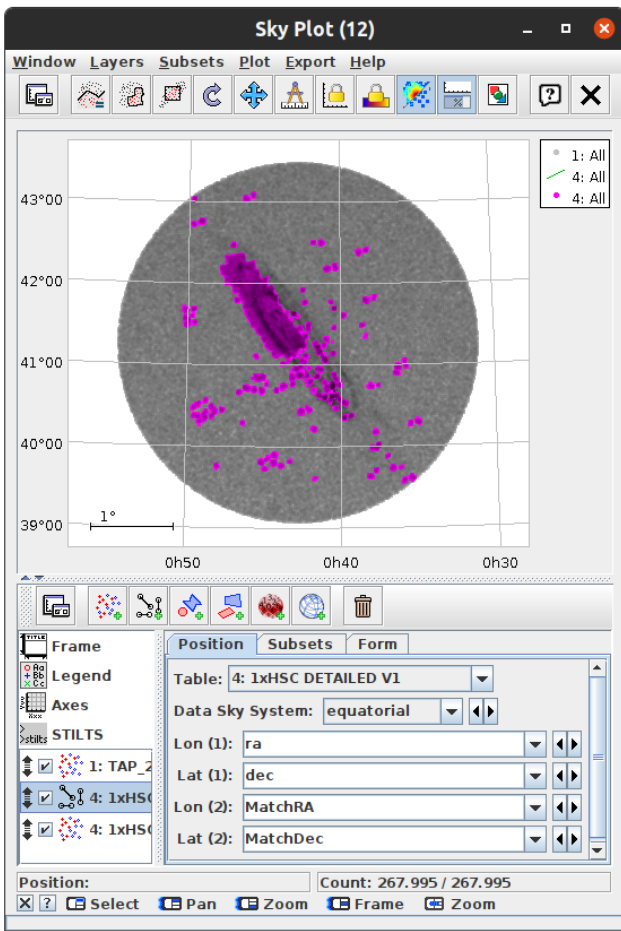
É natural querer visualizar graficamente o resultado do crossmatching. Para crossmatchings internos, o TOPCAT irá exibir uma janela (Fig. 29) com a opção de criar o gráfico quando houver sucesso. Para os externos, precisamos criar os gráficos “do zero”. A base de tudo é um Sky Plot (seção 4.3) com os dois conjuntos de dados. Na Figura 30a, esses são o primeiro e terceiro gráficos na lista. Esse gráfico por si só já dá uma ideia da proporção de pontos que tinham contrapartes na outra tabela.

Para visualizar quais são as contrapartes de cada ponto, selecionamos mais um controle do gráfico, uue traça linhas entre pares de pontos com o botão . Precisamos indicar ao TOPCAT quais são as colunas das posições e as colunas das contrapartes. No caso da Figura 30a, as posições estão nas típicas colunas de ra e dec e as posições das contrapartes estão nas colunas MatchRA e MatchDec, mas os nomes na sua tabela podem ser diferentes. O resultado (com muito zoom) é a Figura 30b, na qual cada par fica ligado por uma linha.

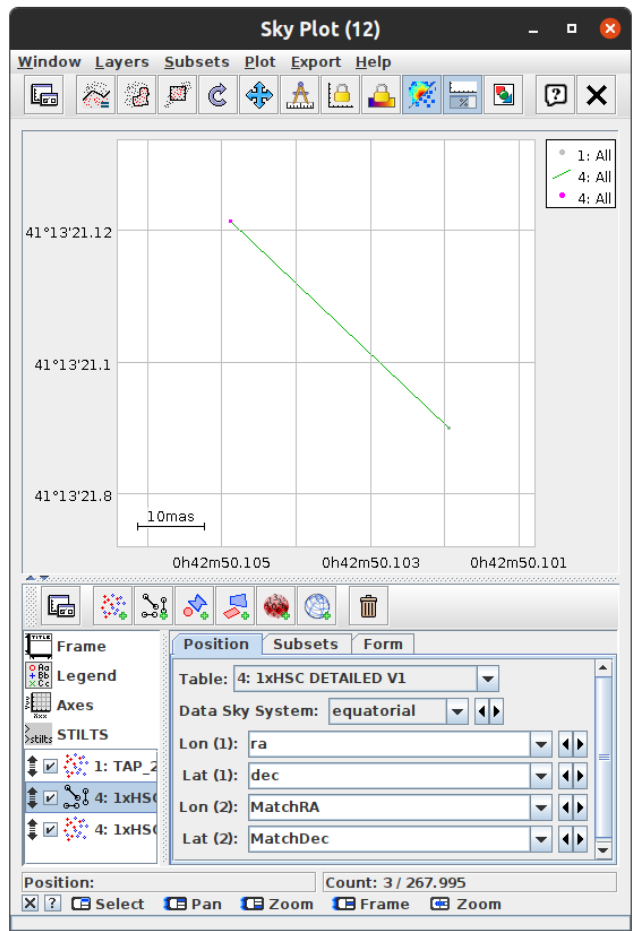


(a)

Figura 29



(a)



(b)

Figura 30

6 Comunicação com outros programas

TOPCAT é capaz de se comunicar com outros programas usando um protocolo de mensagem: via SAMP (**S**imple **A**pplications **M**essaging **P**rotocol) ou PLASTIC (**P**latform for **A**STronomical **I**nter**C**onnection). Basicamente, todos os programas estão conectados a um processo de *hub* através do qual o TOPCAT pode enviar ou receber informações (também chamadas de mensagens) de outros programas. Alguns dos programas com os quais o TOPCAT pode se comunicar são

- ferramentas de análise de imagens: SAOImage DS9, GAIA, Aladin;
- ferramentas de análise de tabelas: VisIVO, STILTS, outras janelas do TOPCAT;
- ferramentas de análise de espectros: SPLAT, VOSpec
- ferramentas de visualização do céu: Aladin, World Wide Telescope, VirGO;
- linguagens de script: Python via astropy;
- Outros!

6.1 Exemplo: enviando catálogos para o SAOImage DS9

Para exemplificar, vamos usar o SAOImage DS9 para demonstrar como é feito o envio de um catálogo para uma ferramenta de exibição de imagem. Usaremos o catálogo e a imagem do aglomerado de Coma⁶ mostrados na Fig. 31. O catálogo foi carregado no TOPCAT e a imagem foi carregada no SAOImage DS9.

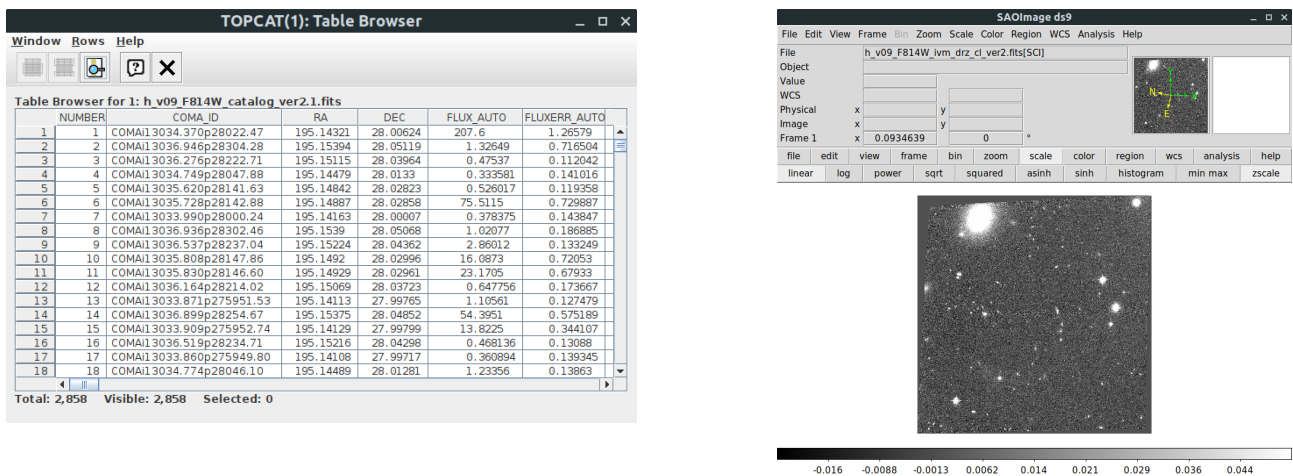


Figura 31

O catálogo contém todas as fontes identificadas na imagem e podemos enviá-lo do TOPCAT para o ds9 de duas formas. A primeira, é clicar no botão de Broadcast (📡) e enviar a tabela para todos os programas conectados ao TOPCAT. A segunda é usar o menu Interop e selecionar Send Tables to... (📄) e selecionar um programa específico, nesse caso, o ds9. Isso

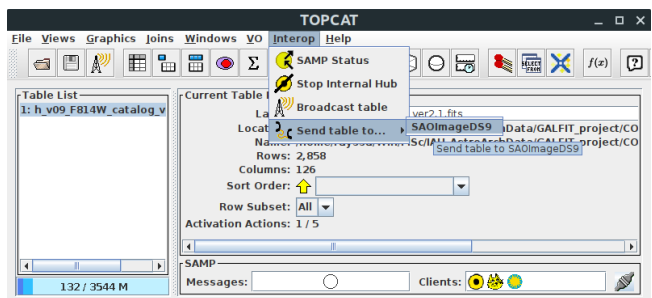
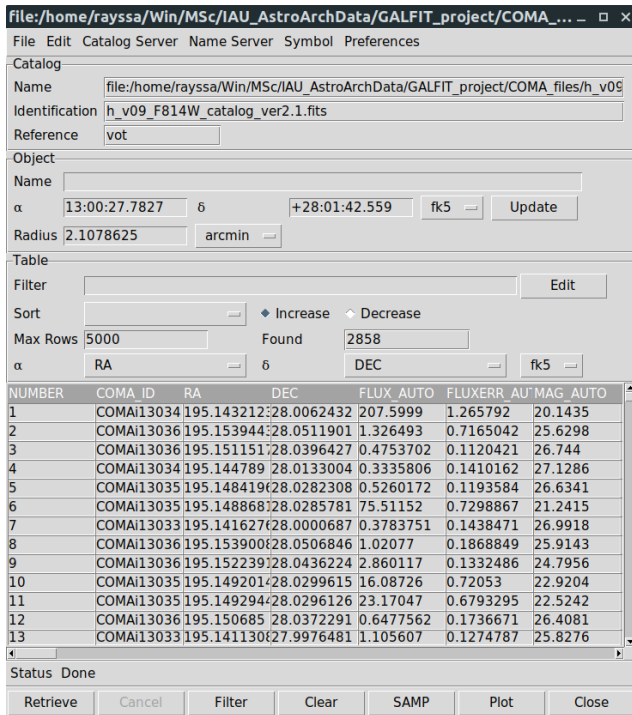


Figura 32

⁶COMA2-2 em <https://archive.stsci.edu/prepds/coma/datalist2.2.html>

irá abrir uma nova janela (Fig. 33a) com o catálogo, mas agora no ds9.

Ao clicar em Plot nessa janela, as posições das fontes serão desenhadas na imagem que já estava aberta no ds9. (Fig. 33b). Ao invés de enviar a tabela inteira, podemos enviar também subconjuntos, basta criar um novo subconjunto e selecionar o Row Subset antes de enviar a tabela ao programa.



(a)

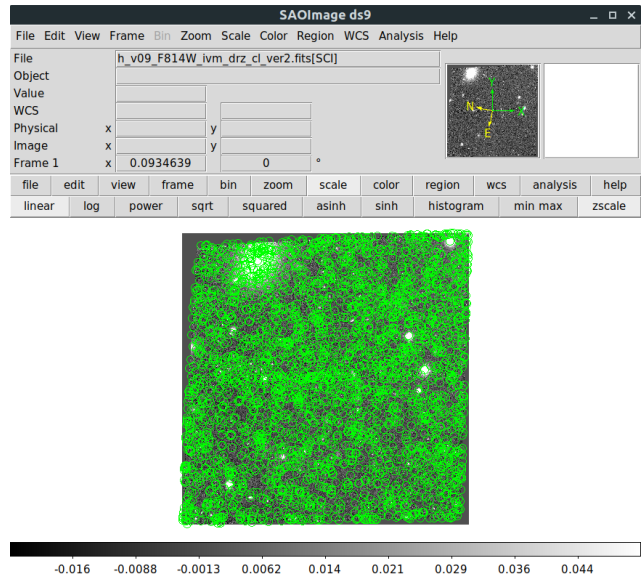


Figura 33

Apêndices

A Observatórios Virtuais e TOPCAT

Um observatório virtual (VO, do inglês *virtual observatory*) é uma coleção de arquivos de dados interoperáveis e de ferramentas de software que utilizam a internet para formar um ambiente de pesquisa científica no qual programas de pesquisa astronômica podem ser conduzidos. Da mesma forma que um observatório real é composto por telescópios com uma coleção de instrumentos astronômicos exclusivos, o VO consiste em uma coleção de centros de dados, cada um com coleções exclusivas de dados astronômicos, sistemas de software e recursos de processamento.

O principal objetivo dos VOs é permitir a distribuição e acesso transparente aos dados astronômicos em todo o mundo. Isso permite que os cientistas descubram, acessem, analisem e combinem dados observacionais e de laboratório provenientes de coleções de dados heterogêneas de forma fácil.

O TOPCAT possui uma série de protocolos que nos permitem acessar diretamente os dados de VOs. Apresentamos aqui alguns dos serviços não discutos na Seção 3.1.

- Cone search 📍: busca pelos dados de uma determinada região do céu em um catálogo remoto;
- Simple Image Access (SIA) 📷: busca dados de imagem (geralmente no formato FITS) em uma determinada região do céu a partir de um repositório remoto de dados de imagem;
- Simple Spectral Access (SSA) 📊: busca dados espectrais, geralmente em uma determinada região do céu, de um repositório de observações espectrais ou de modelos teóricos;
- Table Access Protocol (TAP) 📄: realiza consultas de forma livre a um banco de dados remoto usando uma linguagem semelhante ao SQL;
- Serviços do CDS: Simbad, VizieR 🌐, X-Match, Hips2fits
- Simulação Virgo-Millennium 🌌: realiza consultas SQL diretas à família de serviços que hospeda os bancos de dados da simulação cosmológica;
- Base de dados BaSTI (Bag of Stellar Tracks and Isochrones) 📊: realiza consultas diretas ao banco de dados BaSTI hospedado pelo INAF-Teramo Astronomical Observatory.

O manual do TOPCAT descreve detalhadamente como acessar cada um dos serviços e quais suas limitações. Aqui aborderemos o SIA, SSA, TAP e o Vizier.

A.1 Simple Image Access (SIA) e Simple Spectral Access (SSA)

Apesar do nome, o que o SIA faz é nos trazer uma tabela com uma coluna na qual cada linha representa uma imagem que pode ser baixada de um site remoto. Portanto, uma das colunas contém o URL da imagem e as outras colunas contêm metadados associados, como formato, tamanho, extensão da imagem e assim por diante. Para exemplificar como realizar uma consulta, vamos considerar a galáxia NGC 1097 observada no levantamento Spitzer Infrared Nearby Galaxies Survey (SINGS).

Para iniciar a consulta, clicamos no botão 📷 na Fig. 5. Na nova janela (Fig. 34a) preenchemos a palavra-chave SINGS no campo Keywords para buscar pelos dados do levantamento. Mais abaixo

precisamos preencher a região que deve ser buscada. Nesse caso, colocamos o nome da NGC 1097 e clicamos em Resolve para buscar as coordenadas. O tamanho angular foi definido como zero para que cubra a posição escolhida seja selecionada. Por fim, escolhemos imagens em formato FITS no seletor Image Format. As informações resumidas da tabela resultante estão na Fig. 34b.

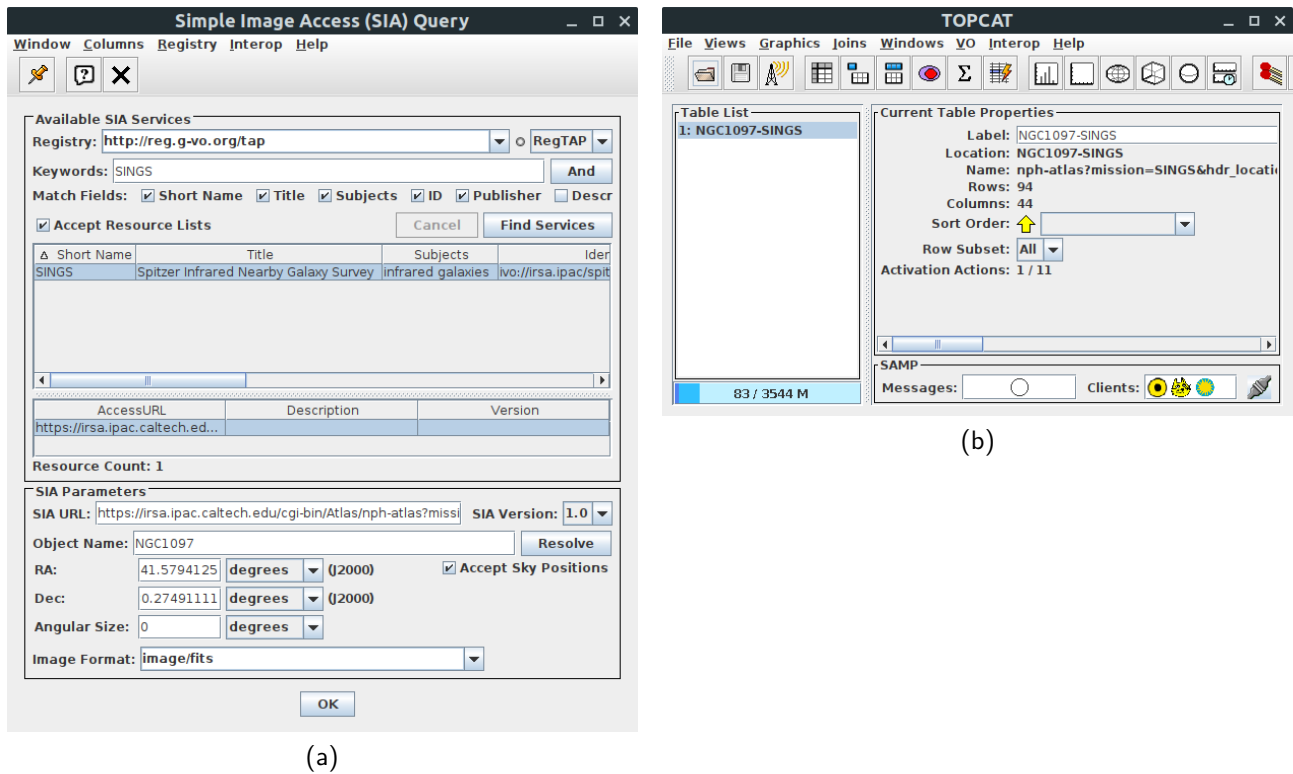


Figura 34

TOPCAT(1): Table Browser



Window Rows Help

Table Browser for 1: NGC1097-SINGS

	instrumen...	band_name	fname	sia_url
1	IRAC	IRAC1	galaxies/ngc1097/IRAC/ngc1097_v7.phot.1_wt.fits	https://irsa.ipac.caltech.edu:443/da...
2	IRAC	IRAC4	galaxies/ngc1097/IRAC/ngc1097_v7.phot.4.fits	https://irsa.ipac.caltech.edu:443/da...
3	IRAC	IRAC1	galaxies/ngc1097/IRAC/ngc1097_v7.phot.1.fits	https://irsa.ipac.caltech.edu:443/da...
4	IRAC	IRAC4	galaxies/ngc1097/IRAC/ngc1097_v7.phot.4_wt.fits	https://irsa.ipac.caltech.edu:443/da...
5	IRAC	IRAC3	galaxies/ngc1097/IRAC/ngc1097_v7.phot.3.fits	https://irsa.ipac.caltech.edu:443/da...
6	IRAC	IRAC2	galaxies/ngc1097/IRAC/ngc1097_v7.phot.2_wt.fits	https://irsa.ipac.caltech.edu:443/da...
7	IRAC	IRAC3	galaxies/ngc1097/IRAC/ngc1097_v7.phot.3_wt.fits	https://irsa.ipac.caltech.edu:443/da...
8	IRAC	IRAC2	galaxies/ngc1097/IRAC/ngc1097_v7.phot.2.fits	https://irsa.ipac.caltech.edu:443/da...
9	MIPS	MIPS24	galaxies/ngc1097/MIPS/ngc1097_mips24_crop_v5-0_wt.fits	https://irsa.ipac.caltech.edu:443/da...
10	MIPS	MIPS160	galaxies/ngc1097/MIPS/ngc1097_mips160_image_v5-0.fits	https://irsa.ipac.caltech.edu:443/da...
11	MIPS	MIPS24	galaxies/ngc1097/MIPS/ngc1097_mips24_image_v5-0_wt.fits	https://irsa.ipac.caltech.edu:443/da...
12	MIPS	MIPS70	galaxies/ngc1097/MIPS/ngc1097_mips70_image_v5-0.fits	https://irsa.ipac.caltech.edu:443/da...
13	MIPS	MIPS160	galaxies/ngc1097/MIPS/ngc1097_mips160_crop_v5-0.fits	https://irsa.ipac.caltech.edu:443/da...
14	MIPS	MIPS70	galaxies/ngc1097/MIPS/ngc1097_mips70_image_v5-0_wt.fits	https://irsa.ipac.caltech.edu:443/da...
15	MIPS	MIPS SED	galaxies/ngc1097/MIPS/ngc1097_mipssed_cube_unc_v3-0.fi...	https://irsa.ipac.caltech.edu:443/da...
16	MIPS	MIPS160	galaxies/ngc1097/MIPS/ngc1097_mips160_image_v5-0_wt.fits	https://irsa.ipac.caltech.edu:443/da...
17	MIPS	MIPS70	galaxies/ngc1097/MIPS/ngc1097_mips70_crop_v5-0.fits	https://irsa.ipac.caltech.edu:443/da...
18	MIPS	MIPS24	galaxies/ngc1097/MIPS/ngc1097_mips24_crop_v5-0.fits	https://irsa.ipac.caltech.edu:443/da...
19	MIPS	MIPS24	galaxies/ngc1097/MIPS/ngc1097_mips24_image_v5-0.fits	https://irsa.ipac.caltech.edu:443/da...

Total: 94 Visible: 94 Selected: 1

Figura 35


A tabela carregada tem 44 colunas, para mostrar menos colunas, clicamos em  e desselecionamos todas as colunas com o botão  na nova janela e selecionamos um pequeno conjunto de 5 colunas que pode ser visto na Fig. 35.

Essa tabela pode ser filtrada por instrumento ou banda observada, por exemplo, antes de ser salva. A coluna mais útil aqui é `sia_url`, pois tem os endereços onde as imagens podem ser baixadas. Isso pode ser feito manualmente copiando os endereços ou através de scripts com o Astropy.

O SSA é bem parecido com o SIA na forma de acesso e na tabela de saída. A principal diferença é que deixar o campo de tamanho angular em branco pode dar alguns problemas. Outra diferença é que é possível deixar os campos de RA e Dec em branco para não restringir a busca. Isso é especialmente útil no caso de serviços de espectros teóricos.

A.2 Table Access Protocol (TAP)


O TAP é uma forma de acessar dados de observatórios virtuais que é especialmente útil quando o volume de dados que desejamos acessar é grande. O acesso é feito usando uma linguagem específica: o ADQL (**Astronomical Data Query Language**). Para exemplificar como realizar uma consulta, nos baseamos na seção 4.3 e 4.4 de *A Beginner's Guide to Working with Astronomical Data*, portanto usaremos o mesmo exemplo e obteremos os dados observados com o GAIA para estrelas na região da galáxia de Andrômeda.

Para iniciar uma busca, clicamos em  na janela da Fig. 5. Na janela da Fig. 36a devemos selecionar um serviço de observatório virtual, no nosso caso, usaremos os dados do GAIA. Ao clicar duas vezes sobre o nome, seremos levados à janela da Fig. 36b, onde podemos acessar algumas informações sobre os catálogos disponíveis nesse serviço.

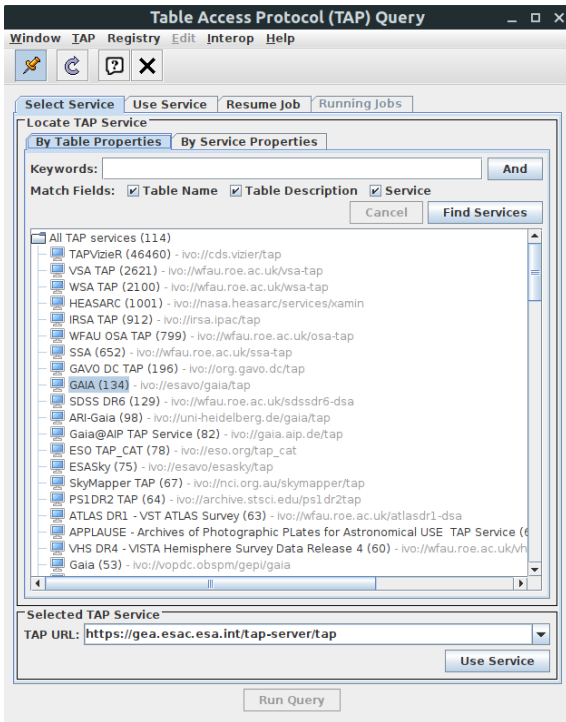
Para acessar os dados de fato, precisamos escrever uma consulta especificando quais dados queremos. Essas consultas são escritas em ADQL⁷, a que usaremos é:

```
SELECT ra, dec
FROM gaiadr2.gaia_source
WHERE 1=CONTAINS(POINT('ICRS', ra, dec), CIRCLE('ICRS', 10.684708, 41.268750, 2.2))
```

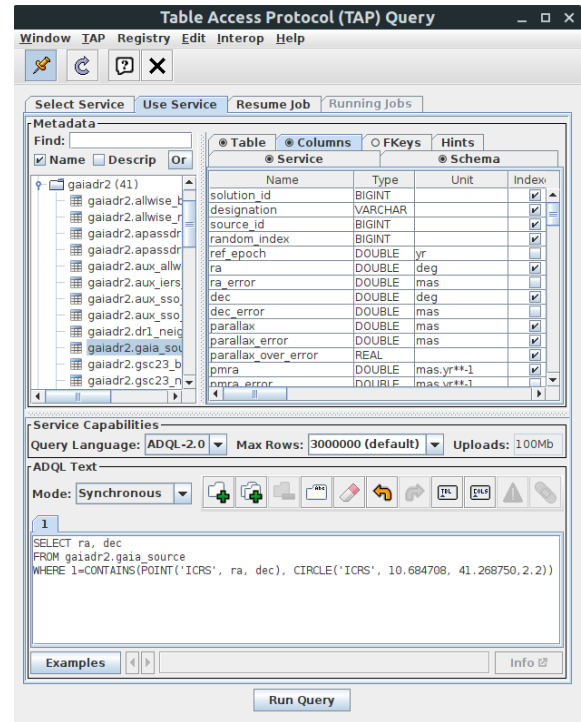
O que essa consulta faz é selecionar as colunas RA e Dec do catálogo `gaiadr2.gaia_source` sempre que a condição do WHERE for verdadeira. O WHERE determina que as coordenadas RA e Dec devem estar contidas em um círculo de 2.2° centrado nas coordenadas da galáxia de Andrômeda.

Podemos aumentar a complexidade da consulta de forma fácil com alguns botões dessa janela. O botão , por exemplo, permite selecionar colunas do catálogo e incluir seus nomes na consulta, como mostrado na Fig. 37. A consulta final fica:

⁷Para uma noção básica sobre SQL (a linguagem geral que originou o ADQL), o Software Carpentry oferece um conjunto de lições em inglês e a documentação específica do ADQL pode ser encontrada na página da Aliança Internacional de Observatórios Virtuais (IVOA).



(a)



(b)

Figura 36

```

SELECT ra, dec, parallax, pmra, pmdec, phot_g_mean_mag, phot_bp_mean_mag,
       phot_rp_mean_mag, bp_rp, bp_g, g_rp
FROM gaiadr2.gaia_source
WHERE 1=CONTAINS(POINT('ICRS', ra, dec), CIRCLE('ICRS', 10.684708, 41.268750, 2.2))

```

A tabela resultante (Fig. 37b) tem 217.335 linhas e 11 colunas.

Table Access Protocol (TAP) Query

Window TAP Registry Edit Interop Help

Select Service Use Service Resume Job Running Jobs

Metadata

Find:

Name	Descr	Or
gaiaedr2 (41)		
gaiaedr2.allwise_t		
gaiaedr2.allwise_r		
gaiaedr2.apassdr		
gaiaedr2.apassdr		
gaiaedr2.aux_allw		
gaiaedr2.aux_iers		
gaiaedr2.aux_sso		
gaiaedr2.aux_sso		
gaiaedr2.dr1_neic		
gaiaedr2.gaia_sou		
gaiaedr2.gsc23_b		
gaiaedr2.gsc23_n		

Table	Columns	FKeys	Hints
Service	Schema		
Name	Type	Unit	Index
phot_rp_mean_flux_error	DOUBLE	'electron'.s**1	
phot_rp_mean_flux_ove...	REAL		
phot_rp_mean_mag	REAL	mag	
phot_bp_rp_excess_fac...	REAL		
phot_proc_mode	SMALLINT		
bp_rp	REAL	mag	
bp_g	REAL	mag	
g_rp	REAL	mag	
radial_velocity	DOUBLE	km.s**1	
radial_velocity_error	DOUBLE	km.s**1	
rv_nb_transits	INTEGER		
rv_template_teff	REAL	K	
rv_template_logg	REAL	log(cm.s**2)	
rv_template_fe_h	REAL	'dex'	

Service Capabilities

Query Language: ADQL-2.0 Max Rows: 3000000 (default) Uploads: 100Mb

ADQL Text

Mode: Synchronous

```

1
SELECT ra, dec, parallax, para, pmdec, phot_g_mean_mag, phot_bp_mean_mag, phot_rp_m
FROM gaiadr2.gaia_source
WHERE 1=CONTAINS(POINT('ICRS', ra, dec), CIRCLE('ICRS', 10.684708, 41.268750, 2.2))

```

Run Query

(a)

TOPCAT(5): Table Browser

Window Rows Help

Table Browser for 5: TAP_6_gaiadr2.gaia_source

	ra	dec	parallax	pmra	pmdec
1	8.22626	42.16845	0.12878	-0.85839	-0.71855
2	8.26077	42.41713	1.59443	0.70349	-0.11473
3	8.3693	42.35024	-0.25761	-1.84796	-0.72388
4	8.32863	42.58995			
5	8.13732	42.21585	0.7113	5.53503	-1.05355
6	8.41237	42.11121	-0.6397	-0.81394	0.28934
7	8.31321	42.32176	0.36412	-2.15693	-2.57902
8	8.4245	42.27865			
9	8.31339	42.43497	0.57497	0.01988	3.60264
10	8.2037	42.05935	0.44277	-13.61177	-7.05625
11	8.48133	42.11049	1.12021	-1.46653	0.7663
12	8.45915	42.488	0.82386	-0.44617	-3.67718
13	8.39616	42.52934	0.97867	-2.10345	-0.1857
14	8.35234	42.41116	0.42938	-4.79239	-1.3897
15	8.37511	42.10871	0.40891	-1.18955	-0.54283
16	8.09868	42.07891	0.07943	-1.08255	-4.7466
17	8.40601	42.10934	0.01637	1.57017	-3.3455
18	8.35979	42.35017	1.25649	21.49988	-2.41387
19	8.42462	42.4752	0.11599	-0.86241	-3.4554
20	8.32529	41.94607	0.96411	-9.14793	-3.80815
21	8.46821	42.34917	0.7183	-1.38579	-3.14943
22	8.009	42.12951			
23	8.35817	42.4677	0.92425	-16.33898	-4.67236
24	8.20703	42.3247	1.29924	4.06212	-0.29991
25	8.44262	42.34917	1.32158	-2.23777	-2.81955
26	8.33918	42.46502	0.7919	2.12529	-1.01629
27	8.21021	42.42843	0.70193	7.26768	1.5657
28	8.44404	42.17457	0.78364	-6.7443	-0.46593
29	8.27007	42.38736	0.45678	-0.36946	-2.2579

Total: 217,335 Visible: 217,335 Selected: 0

(b)

Figura 37

A.3 Vizier

O Vizier é uma biblioteca de catálogos astronômicos que pode ser acessado pela *web* ou outras interfaces. O TOPCAT oferece uma forma direta de consultar os catálogos através do serviço Vizier Catalogue Service (🌀), acessível no menu VO. Ao buscar um catálogo, é possível que várias tabelas sejam carregadas no TOPCAT, diferente da busca de cone, por exemplo.

A primeira caixa da janela do Vizier (Fig. 38a), Row Selection, controla quais linhas desejamos selecionar no catálogo escolhido. A opção Cone Selection é bem similar à busca de cone padrão do TOPCAT – todas as linhas dentro do raio de seleção serão carregadas no programa. A opção All Rows nos permite baixar o catálogo completo sem restrições. Independente do caso, devemos definir um limite de linhas que podem ser carregadas no seletor Maximum Row Count.

A segunda caixa, Column Selection, controla quais colunas serão incluídas na tabela de saída. Cada opção faz algo um pouco diferente:

- standard: seleciona um conjunto de colunas consideradas mais interessantes pelo serviço;
- default: além das colunas standard, tem colunas de posição (`_RAJ2000` e `_DEJ2000`) e no caso de uma seleção de cone, contém também uma coluna `_r` com a distância até a posição central do cone;
- all: seleciona todas as colunas do catálogo.

A terceira caixa, Catalogue Selection, é onde selecionamos o catálogo de fato e há várias formas de fazer isso:

- By Category: nos permite selecionar uma coleção de catálogos a partir de palavras-chave para o comprimento de onda, missão/telescópio e palavras diversas (Fig. 38);
- By Keyword: procura catálogos que contenham as palavras listadas nos nomes (Fig. 39);

As duas opções anteriores tem duas caixas de seleção: Sub-Table Details e Include Obsolete Tables. A primeira controla se o Vizier vai exibir apenas o nível mais alto do catálogo – que pode conter diversas tabelas – ou se vai listar cada tabela do catálogo como um item diferente. A segunda controla se versões antigas dos catálogos serão exibidas na lista de opções.

- Surveys e Missions: fornecem uma lista curta de catálogos completos observados por telescópios, missões ou levantamentos específicos.

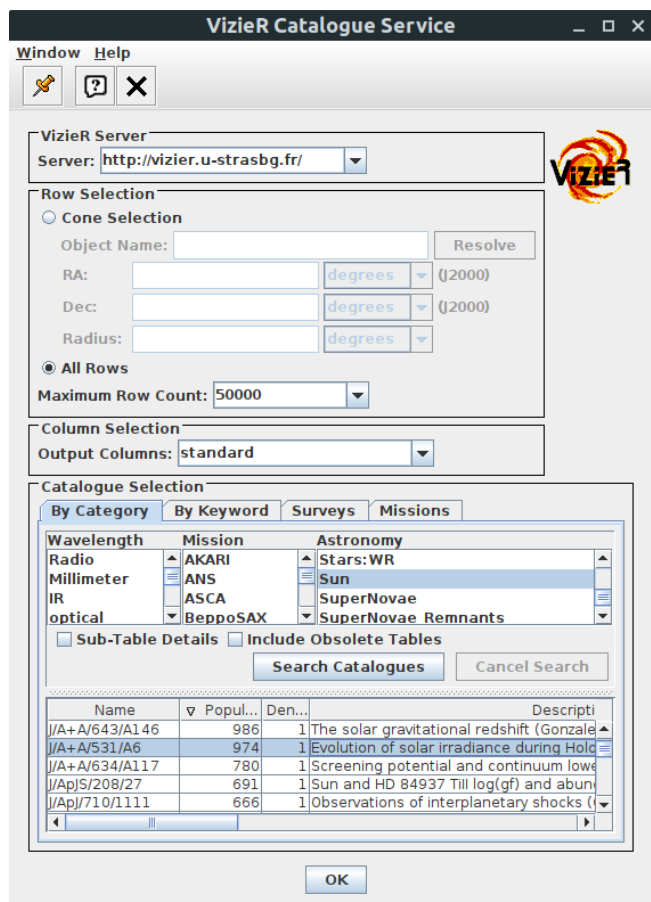


Figura 38

Após clicar em Search Catalogues, o VizierR exibirá uma tabela com as opções de catálogos com colunas como nome, descrição, número de linhas etc. Uma das colunas mais interessantes é a Popularity que indica o número de consultas feitas ao catálogo no VizierR. Por fim, após selecionar qual catálogo desejamos carregar, clicamos em OK e as tabelas ficarão disponíveis no TOPCAT.

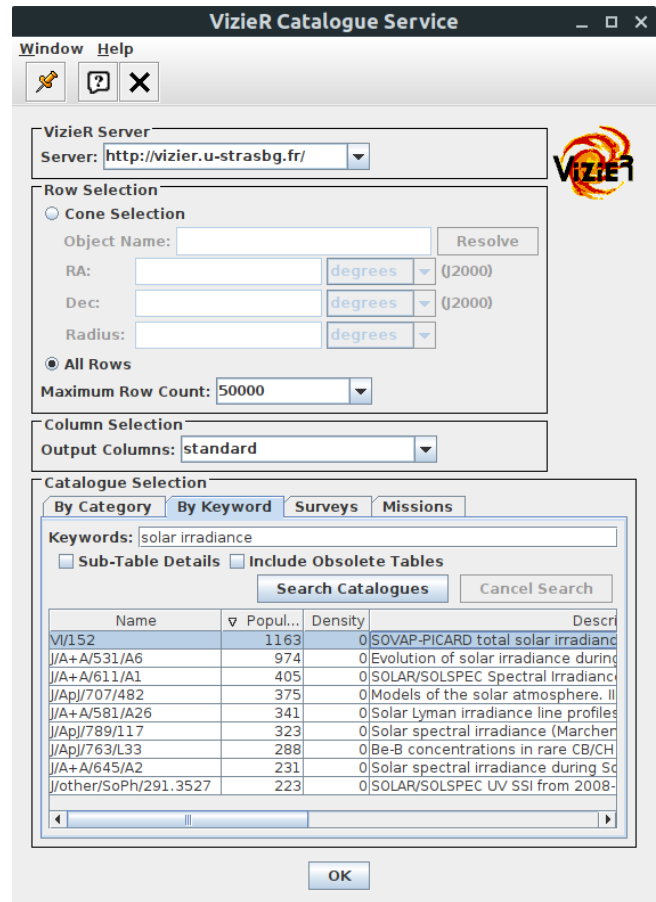



Figura 39

B STILTS

STILTS (Starlink Tables Infrastructure Library Tool Set) é uma interface de linha de comando das funcionalidades do TOPCAT. O principal uso é reproduzir gráficos criados no TOPCAT a partir de algumas linhas de código. Isso é feito através da opção STILTS  no painel de controle dos gráficos (ver seção 4). É possível instalar o STILTS como um programa separado, mas ele já vem incluído no TOPCAT e pode ser acessado via linha de comando.

Como exemplo de seu uso, vamos reproduzir o gráfico da Fig. 20b com um comando STILTS. Primeiro, precisamos do comando gerado pelo próprio TOPCAT e que está na aba Command (Fig. 40):

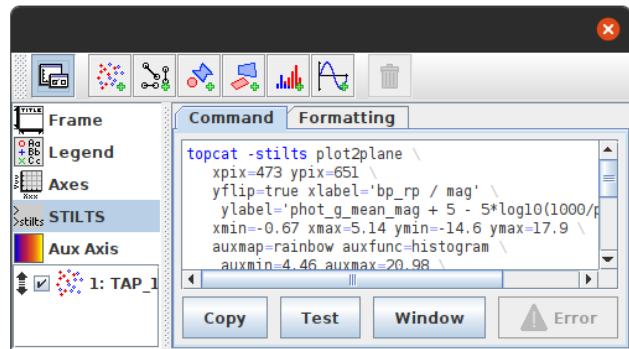


Figura 40

```
topcat -stilts plot2plane \  
xpix=471 \  
yflip=true xlabel='bp_rp / mag' \  
  ylabel='phot_g_mean_mag + 5 - 5*log10(1000/parallax)' \  
xmin=-0.67 xmax=5.14 ymin=-14.6 \  
  ymax=17.9 \  
auxmap=rainbow auxfunc=histogram \  
  auxmin=4.5 auxmax=21 \  
auxvisible=true \  
  auxlabel='phot_g_mean_mag_min / mag' \  
legend=false \  
layer=Mark \  
  in=TAP_1_gaiadr2.gaia_source \  
  x=bp_rp \  
  y='phot_g_mean_mag + 5 - 5*log10(1000/parallax)' \  
  weight=phot_g_mean_mag \  
  shading=weighted combine=min
```

Para usar esse comando, precisamos salvar a tabela `TAP_1_gaiadr2.gaia_source` e alterar a linha `in=TAP_1_gaiadr2.gaia_source` para apontar para o caminho correto da tabela (o STILTS também aceita URLs!). Salvei a tabela como `/home/rayssa/gaia.vot`. Finalmente, precisamos acessar o STILTS. Para acessar a versão que vem com o TOPCAT, precisamos das duas primeiras palavras do comando STILTS: o comando para chamar o topcat e a *flag* `-stilts`.

Chamando o TOPCAT a partir do arquivo Java, o comando fica

```
java -jar topcat-*.jar -stilts plot2plane \  
xpix=471 \  

```



```

yflip=true xlabel='bp_rp / mag' \
  ylabel='phot_g_mean_mag + 5 - 5*log10(1000/parallax)' \
xmin=-0.67 xmax=5.14 ymin=-14.6 \
  ymax=17.9 \
auxmap=rainbow auxfunc=histogram \
  auxmin=4.5 auxmax=21 \
auxvisible=true \
  auxlabel='phot_g_mean_mag_min / mag' \
legend=false \
layer=Mark \
  in=/home/rayssa/gaia.vot \
  x=bp_rp \
  y='phot_g_mean_mag + 5 - 5*log10(1000/parallax)' \
  weight=phot_g_mean_mag \
  shading=weighted combine=min

```

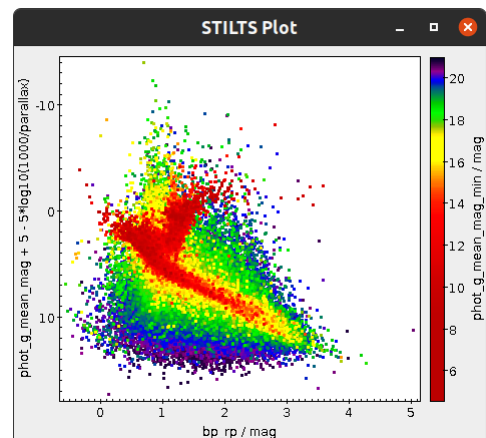
basta colar na linha de comando, como na Fig. 41a e uma nova janela com o gráfico irá aparecer (Fig. 41b). Dessa forma não precisamos lembrar todos os detalhes de como produzimos o gráfico e podemos usar o mesmo comando com diferentes dados. Para salvar os gráficos em um arquivo basta adicionar o parâmetro de saída `out=nomedoarquivo.extensão`.

```

Terminal
Topcat$java -jar topcat-*.jar -stilts plot2plane \
> xpix=471 \
> yflip=true xlabel='bp_rp / mag' \
> ylabel='phot_g_mean_mag + 5 - 5*log10(1000/parallax)' \
> xmin=-0.67 xmax=5.14 ymin=-14.6 \
> ymax=17.9 \
> auxmap=rainbow auxfunc=histogram \
> auxmin=4.5 auxmax=21 \
> auxvisible=true \
> auxlabel='phot_g_mean_mag_min / mag' \
> legend=false \
> layer=Mark \
> in=/home/rayssa/gaia.vot \
> x=bp_rp \
> y='phot_g_mean_mag + 5 - 5*log10(1000/parallax)' \
> weight=phot_g_mean_mag \
> shading=weighted combine=min

```

(a)



(b)

Figura 41

O STILTS oferece muitas outras funcionalidades do TOPCAT, dependendo das suas necessidades, talvez seja uma boa ideia dar uma olhada na [documentação](#).