

# Big Data in Astronomy: Introduction to Large Surveys

AST 597 B

[mjuric@astro.washington.edu](mailto:mjuric@astro.washington.edu)

# Today

- Administrivia
- Introduction to Large Surveys

# About Me

- Mario Juric (mar-ee-oh you-rich)
- What I do:
  - Science derived from large surveys: Galactic structure, properties of the solar system
  - Astronomical algorithms and software research
  - Optical astronomer (=> some bias towards optical surveys)
- Hats I wear
  - Professor @ Astronomy
  - Sr. Fellow @ UW's eScience Institute
  - Data Management Project Scientist @ LSST

# About Me

- Nominal Coordinates:
  - Office: C320
  - E-mail: [mjuric@astro.washington.edu](mailto:mjuric@astro.washington.edu)
  - Open door policy: stop by any time!

## How to read a Professor's door

Closed



Half-open



Wide Open



Slightly ajar



"I may or may  
not be on this  
continent."

"I'm probably  
in a meeting."

"I just walked in to  
get a few things  
before I have to  
run to my next  
meeting."

"Proceed with  
caution."

# About This Course

- The goal of this course is to prepare you for research with large survey data, teach you how to think about such data sets, and give you an overview of what is or soon will be available.
- We will:
  - Learn about how large observational data sets are changing astronomy
  - Introduce tools and techniques for working with large data sets (incl. access, analysis, and visualization)
  - Introduce and work hands-on with existing data sets (including SDSS, WISE, Kepler and others)
  - Learn about major upcoming surveys (including PanSTARRS, Gaia, and LSST)

# The Goal

- Late March 2015 (reading astro-ph):

*"Oh, this is an interesting paper! Looks like if I combined IR and optical observations I could find these kinds of objects... But wait, I know there's WISE and SDSS! I'll run a SQL query against those databases, use LSD to join the result, analyze the data in Python using scikits tools, and send the result to my collaborator as an IPython notebook via github.*

- You'll know what's out there, what it's good for, and how to use it.

# Learning about Learning

- In survey astronomy (and related computing disciplines), things change too quickly to be learned only once – it's a process that never ends.
- The major emphasis of this course will be on showing how to independently find information, evaluate, and use it.

# Mailing List

- This is a topics course, and many of you are auditing
- If you're not enrolled, e-mail me your e-mail address so I can add you to the class mailing list
  - [mjuric@astro.washington.edu](mailto:mjuric@astro.washington.edu)
- All announcements, etc., will go via that list

# Lectures

- MW, 3:30pm-4:50pm
  - Note: Next class – merged with ASTR 511, Thursday, 1:45pm
- Q: Does this work for everyone?
- Lecture structure:
  - At most 1hr for introducing new materials, etc.
  - At least 20 minutes for open-ended discussion
  - Interrupt and ask questions at any time!

# Course Materials

- I'll be adding most of what we need to the following repository on GitHub:

<https://github.com/mjuric/ast597b>

# Smile, You're On Camera!

- Actually, it's mostly me who's on camera...
- Actually, it's mostly my slides that will be on camera...
- Joining the class live via Google Hangouts: <http://goo.gl/9oTASp>
- Viewing the recording later, via YouTube
  - I'll e-mail the link after class
  - Note: recorded classes will be private, and won't be kept after this quarter ends.

# Syllabus

- <https://github.com/mjuric/ast597b>
- Three parts
  - Software / Data Analysis Tools (~ January)
  - Major Surveys (~ February)
  - Future Surveys and Data Challenges (~March)

# Grading, Homeworks, Etc.

- Two homeworks (50% of the grade):
  - All homeworks will be IPython notebooks, turned in via GitHub
  - HW #1, due Feb 11, focused on software tools
  - HW #2, due Feb 25, focused on survey data analysis
- Paper presentations (25% of the grade):
  - Present a paper on a survey / dataset of your choosing, making sure to cover:
    - The goals of the survey
    - The data it contains (and any caveats about it)
    - What it may be useful for
    - The best ways to obtain it
- Final project (25% of the grade):
  - More about this in a week or so ☺.

# Questions?

- Large Survey and Why They're Different

# Sur·vey

*verb*

/sər'vā/

1. (of a person or their eyes) look carefully and thoroughly at (someone or something), especially so as to appraise them.

"her green eyes surveyed him coolly"

*synonyms:* look at, look over, [observe](#), [view](#), [contemplate](#), [regard](#), gaze at, stare at, [eye](#); [More](#)

2. examine and record the area and features of (an area of land) so as to construct a map, plan, or description.

"he surveyed the coasts of New Zealand"

*noun*

/'sər,vā/

1. a general view, examination, or description of someone or something.

"the author provides a survey of the relevant literature"

*synonyms:* [study](#), [review](#), [consideration](#), [overview](#); [More](#)

2. an act of surveying an area of land.

"the flight involved a detailed aerial survey of military bases"



Translations, word origin, and more definitions

## Hipparchus of Rhodes (180-125 BC)

Discovered the precession of the equinoxes.

Measured the length of the year to ~6 minutes.

In 129 BC, constructed one of the first star catalogs\*, containing about 850 stars.



*n.b.: also the one to blame for the magnitude system ...*

## Galileo Galilei (1564-1642)

Researched a variety of topics in physics, but called out here for the introduction of the *Galilean telescope*.

Galileo's telescope allowed us for the first time to *zoom in* on the cosmos, and study the individual objects in great detail.





*Joseph von Fraunhofer (1787-1826)*

Mounted a prism in front of an objective of a small telescope, and pointed it to the Sun.

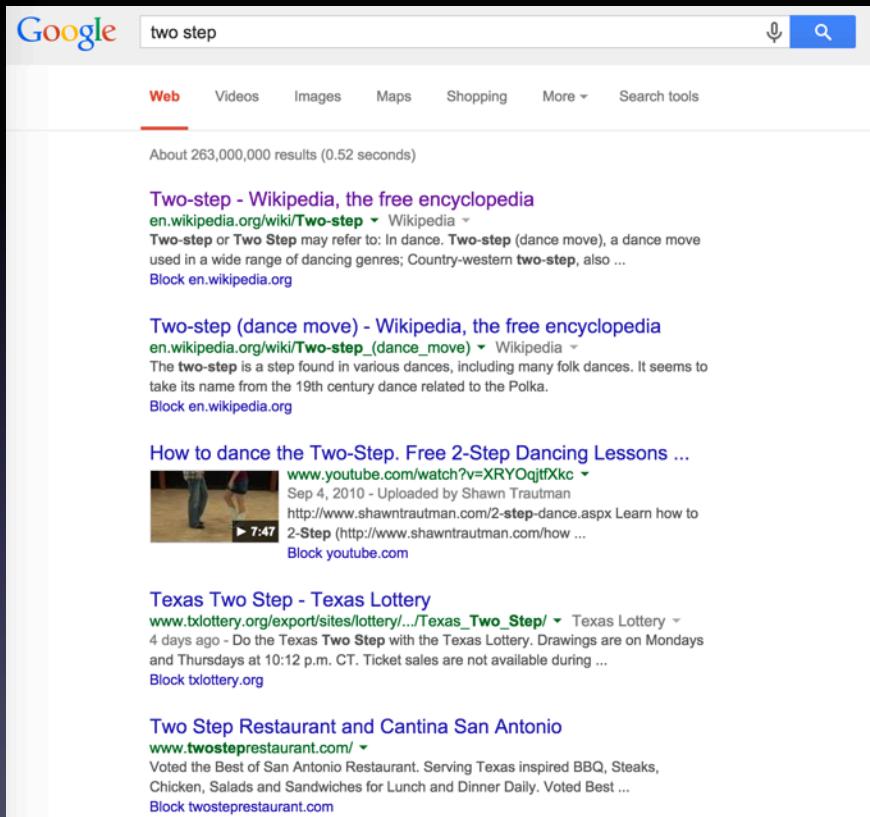
In 1859, Kirchhoff and Bunsen understood what Fraunhofer saw.

The birth of modern astrophysics!

# The Astrophysics Two-Step

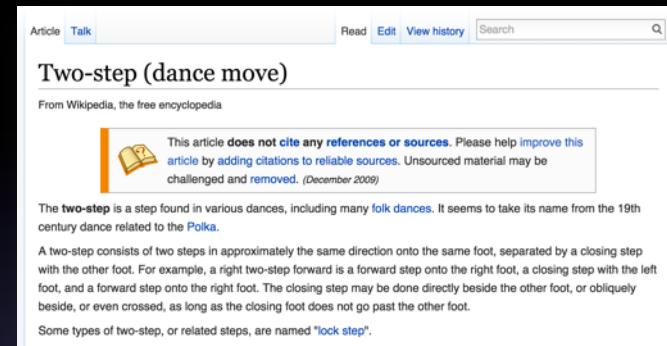
- Surveys
  - Construct catalogs and maps of objects in the sky. Focus on coarse classification and discovering targets for further follow-up.
- Large telescopes
  - Acquire detailed observations of a few representative objects. Understand the details of astrophysical processes that govern them, and extrapolate that understanding to the entire class.

# Analogy: Google Search



Google search results for "two step". The search bar shows "two step". Below it, there are tabs for Web, Videos, Images, Maps, Shopping, More, and Search tools. The search results include:

- Two-step - Wikipedia, the free encyclopedia**  
[en.wikipedia.org/wiki/Two-step](https://en.wikipedia.org/wiki/Two-step) ▾ Wikipedia ▾  
Two-step or Two Step may refer to: In dance. **Two-step (dance move)**, a dance move used in a wide range of dancing genres; Country-western **two-step**, also ...  
Block en.wikipedia.org
- Two-step (dance move) - Wikipedia, the free encyclopedia**  
[en.wikipedia.org/wiki/Two-step\\_\(dance\\_move\)](https://en.wikipedia.org/wiki/Two-step_(dance_move)) ▾ Wikipedia ▾  
The **two-step** is a step found in various dances, including many folk dances. It seems to take its name from the 19th century dance related to the Polka.  
Block en.wikipedia.org
- How to dance the Two-Step. Free 2-Step Dancing Lessons ...**  
 [www.youtube.com/watch?v=XRYOqjtXkc](https://www.youtube.com/watch?v=XRYOqjtXkc) ▾  
Sep 4, 2010 - Uploaded by Shawn Trautman  
<http://www.shawntrautman.com/2-step-dance.aspx> Learn how to 2-Step ([http://www.shawntrautman.com/how\\_...](http://www.shawntrautman.com/how_...))  
Block youtube.com
- Texas Two Step - Texas Lottery**  
[www.txdottery.org/export/sites/lottery/.../Texas\\_Two\\_Step/](https://www.txdottery.org/export/sites/lottery/.../Texas_Two_Step/) ▾ Texas Lottery ▾  
4 days ago - Do the Texas Two Step with the Texas Lottery. Drawings are on Mondays and Thursdays at 10:12 p.m. CT. Ticket sales are not available during ...  
Block txdottery.org
- Two Step Restaurant and Cantina San Antonio**  
[www.twosteprestaurant.com/](https://www.twosteprestaurant.com/) ▾  
Voted the Best of San Antonio Restaurant. Serving Texas inspired BBQ, Steaks, Chicken, Salads and Sandwiches for Lunch and Dinner Daily. Voted Best ...  
Block twosteprestaurant.com



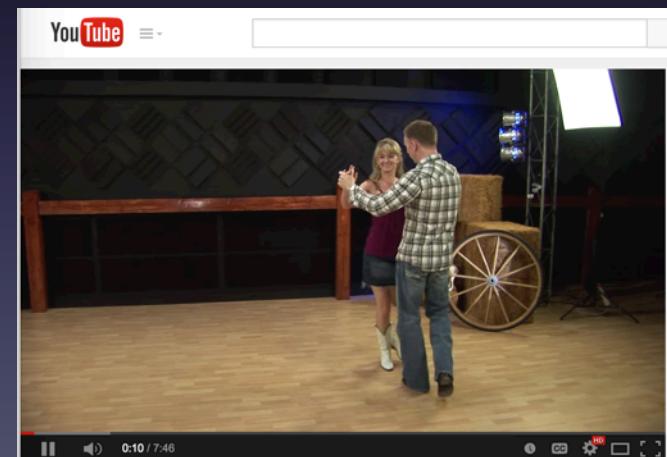
Wikipedia article on **Two-step (dance move)**. From Wikipedia, the free encyclopedia.

This article does not cite any references or sources. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (December 2009)

The **two-step** is a step found in various dances, including many **folk dances**. It seems to take its name from the 19th century dance related to the **Polka**.

A two-step consists of two steps in approximately the same direction onto the same foot, separated by a closing step with the other foot. For example, a right two-step forward is a forward step onto the right foot, a closing step with the left foot, and a forward step onto the right foot. The closing step may be done directly beside the other foot, or obliquely beside, or even crossed, as long as the closing foot does not go past the other foot.

Some types of two-step, or related steps, are named "**lock step**".



YouTube video titled "How to dance the Two-Step. Free 2-Step Dancing Lessons w/Shawn Trautman". The video shows a man and a woman dancing the two-step on a wooden floor. The video player shows the progress bar at 0:10 / 7:46.

Google's index is a catalog of the Web. We use it to "zoom in" on individual entries to find out more.

How to dance the Two-Step. Free 2-Step Dancing Lessons w/Shawn Trautman



Subscribe

852,553

+ Add to Share More

719 139

**Web**

News

Shopping

Maps

Apps

More ▾

Search tools



About 1,400,000,000 results (0.43 seconds)

## Seattle, WA

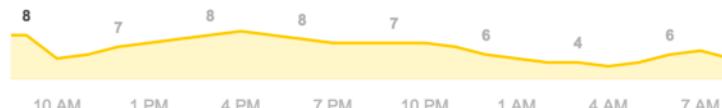
Monday 9:00 AM

Mostly Cloudy



8 °C | °F

Precipitation: 10%  
Humidity: 81%  
Wind: 0 km/h

**Temperature** **Precipitation** **Wind**

Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon

[More on weather.com](#)[Feedback](#)

## National and Local Weather Forecast, Hurricane, Radar and ...

[www.weather.com](#) ▾ The Weather Channel ▾

The Weather Channel and [weather.com](#) provide a national and local **weather** forecast for cities, as well as **weather** radar, report and hurricane coverage.

[Block weather.com](#)

## The Weather Channel

2,336,769 followers on Google+

[Follow](#)

### Recent posts

Stunning images from astrophotographer and night photographer Rogelio Bernal Andreo showing the darker beauty of Hawaii, from its captivating volcanoes to ...  
Jan 9, 2015

But, it's more than just a catalog of pointers – more and more, Google itself collects, processes, indexes, visualizes, and serves the actual information we need.

More and more often, our “research” begins and ends with Google!

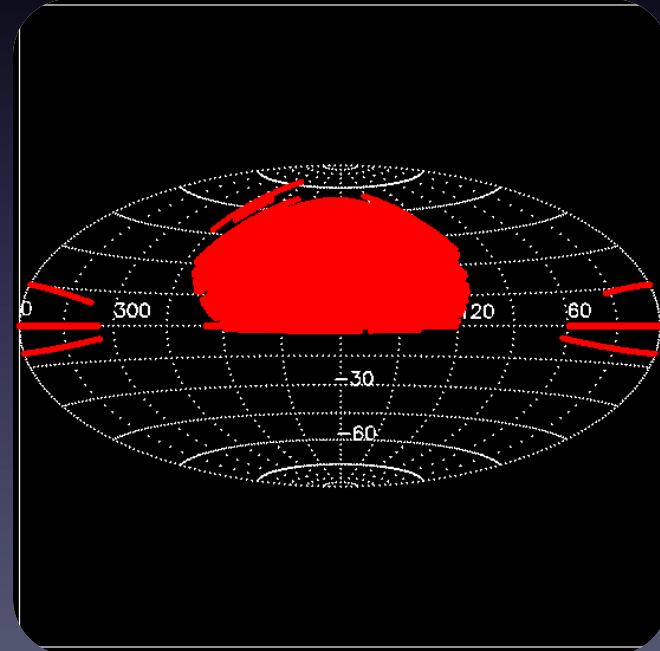
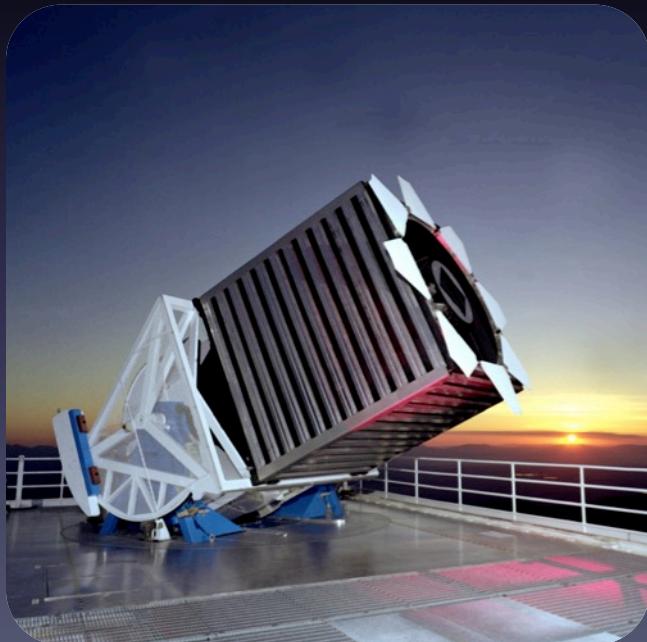
# Modern Large Astronomical Surveys are the Googles of the Sky

- There's a close parallel with large surveys in astronomy, in scale, quality, and richness of the collected information
  - Scale: We're entering the era when we can catalog the entire sky
  - Quality: Those catalogs will be as precise as the measurements taken with "pointed" observations (used to be ~5-10x worse)
  - Richness: Those catalogs contain not only positions and magnitudes, but also shapes, profiles, and temporal behavior of the objects.
- Quite often, the research begins and ends with the survey.
- This is what makes large surveys of today not just bigger, but different. They're more than just "finding charts".

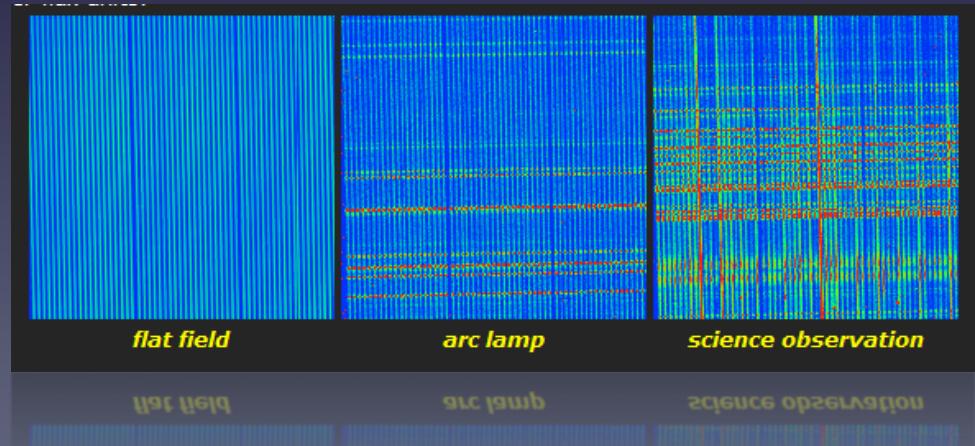
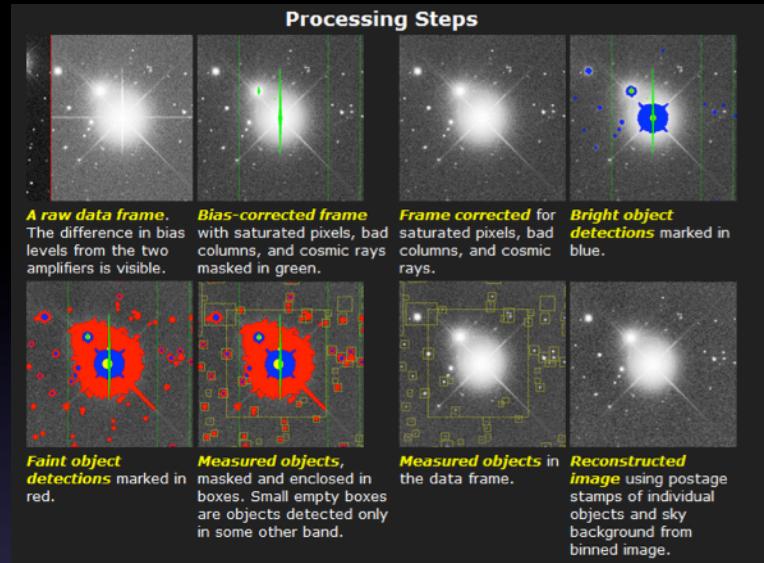
# Sloan Digital Sky Survey

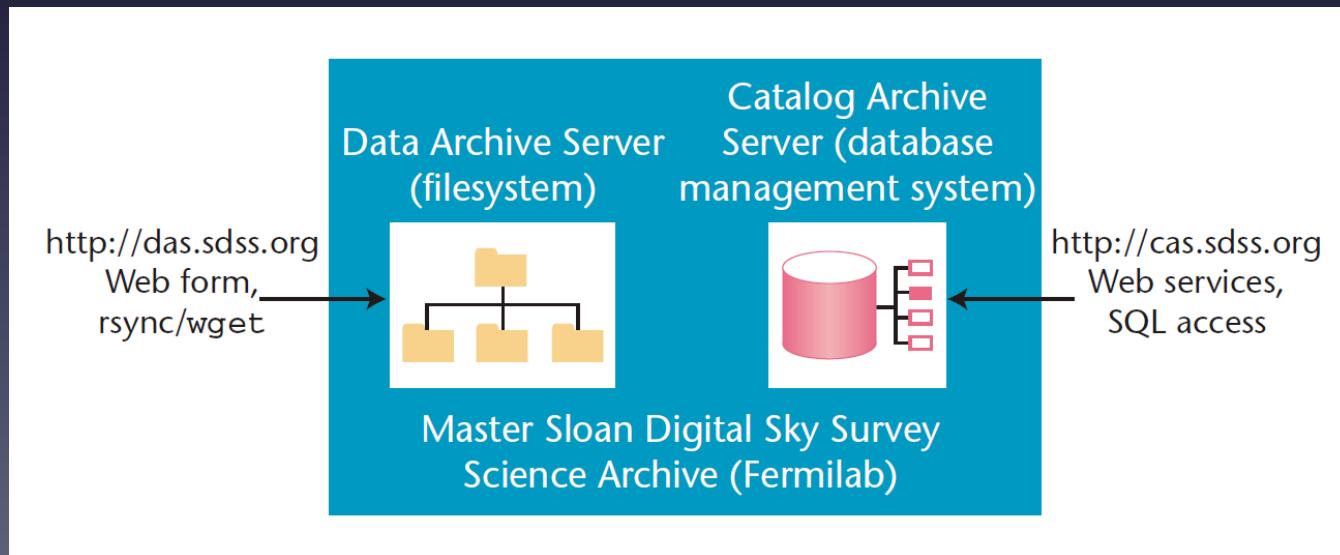
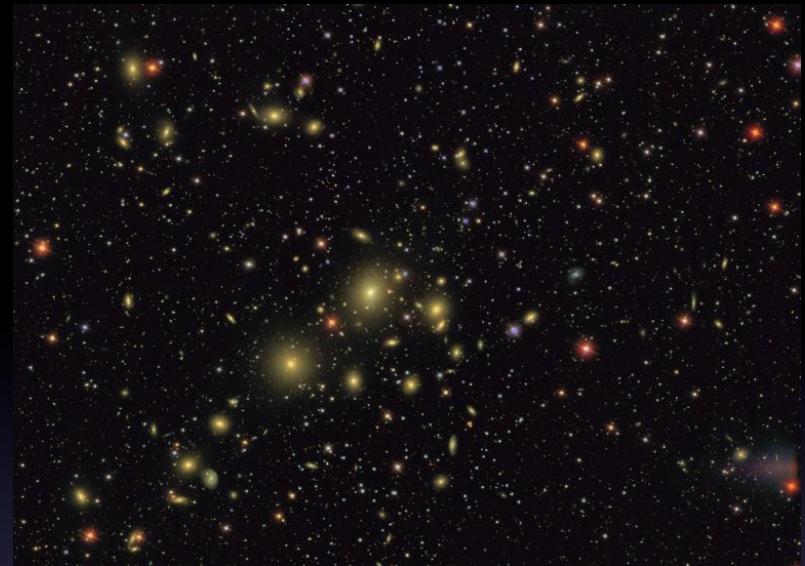
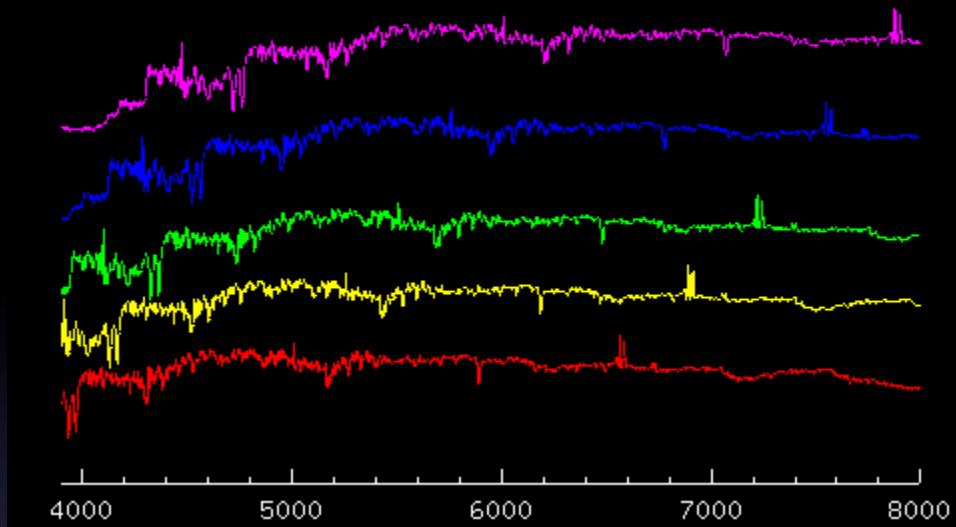
2.5m telescope    10000 deg<sup>2</sup>    0.1" astrometry    r<22.5 flux limit

5 band, 2%, photometry for >50M stars, >280k R=2000 stellar spectra



# Observing With SDSS





**Table 1. Sloan Digital Sky Survey public data releases.**

Release	Date	Size (catalogs)	Images (millions)	Spectra (thousands)	Distribution/mirrors
EDR <sup>2</sup>	June 2001	200 Gbytes	14	54	Johns Hopkins Univ. (JHU), San Diego Supercomputing Center (SDSC), UK, Japan
DR1 <sup>6</sup>	June 2003	1 Tbyte	53	186	JHU, SDSC, Canadian Astronomical Data Centre (CADC), Univ. of Pittsburgh, UK, Germany, Japan, India
DR2 <sup>5</sup>	Mar. 2004	2 Tbytes	88	330	JHU, Univ. of Pittsburgh, SDSC, Germany
DR3 <sup>3</sup>	Sept. 2004	3 Tbytes	141	478	JHU, UK, India
DR4 <sup>4</sup>	June 2005	4 Tbytes	180	608	JHU, Germany, Hungary, Brazil
DR5 <sup>6</sup>	June 2006	5 Tbytes	215	738	JHU, India, Russia, Hungary, Australia
DR6 <sup>7</sup>	June 2007	6 Tbytes	287 M	1.27 M	JHU, India, Hungary
DR7 <sub>a</sub>	June 2007	6 Tbytes	582 M	1.75 M	JHU, Canada, UK, Australia
DR7 <sub>b</sub>	June 2008	6 Tbytes	612	1.38 M	JHU, Canada, UK, Australia
DR7 <sub>c</sub>	June 2008	4 Tbytes	180	908	JHU, Germany, UK, Brazil

SDSS SkyServer DR7

http://cas.sdss.org/dr7/en/

# Sloan Digital Sky Survey / SkyServer

SDSS

Home Tools Schema Projects Astronomy SDSS Contact Us Download Site Search Help

Welcome to the DR7 site!!!

This website presents data from the Sloan Digital Sky Survey, a project to make a map of a large part of the universe. We would like to show you the beauty of the universe, and share with you our excitement as we build the largest map in the history of the world.

News

The site hosts data from Data Release 7 (DR7). What's new in DR7, what's new on this site, and known problems. More...

For Astronomers

A separate branch of this website for professional astronomers (English) More...

SDSS is supported by

National Science Foundation (NSF)

NASA

Ministry of Education, Culture, Sports, Science and Technology (MEXT)

Powered by Microsoft

**SkyServer Tools**

- Famous places
- Get images
- Visual Tools
- Explore
- Search
- Object Cross-ID
- CasJobs

**Science Projects**

- Basic
- Advanced
- Challenges
- For Kids
- Games and Contests
- Teachers
- Links to other projects

**Info Links**

- About Astronomy
- About the SDSS
- About the SkyServer
- SDSS Data Release 7
- SDSS Project Website
- Open SkyQuery
- Images of RC3 Galaxies

**Help**

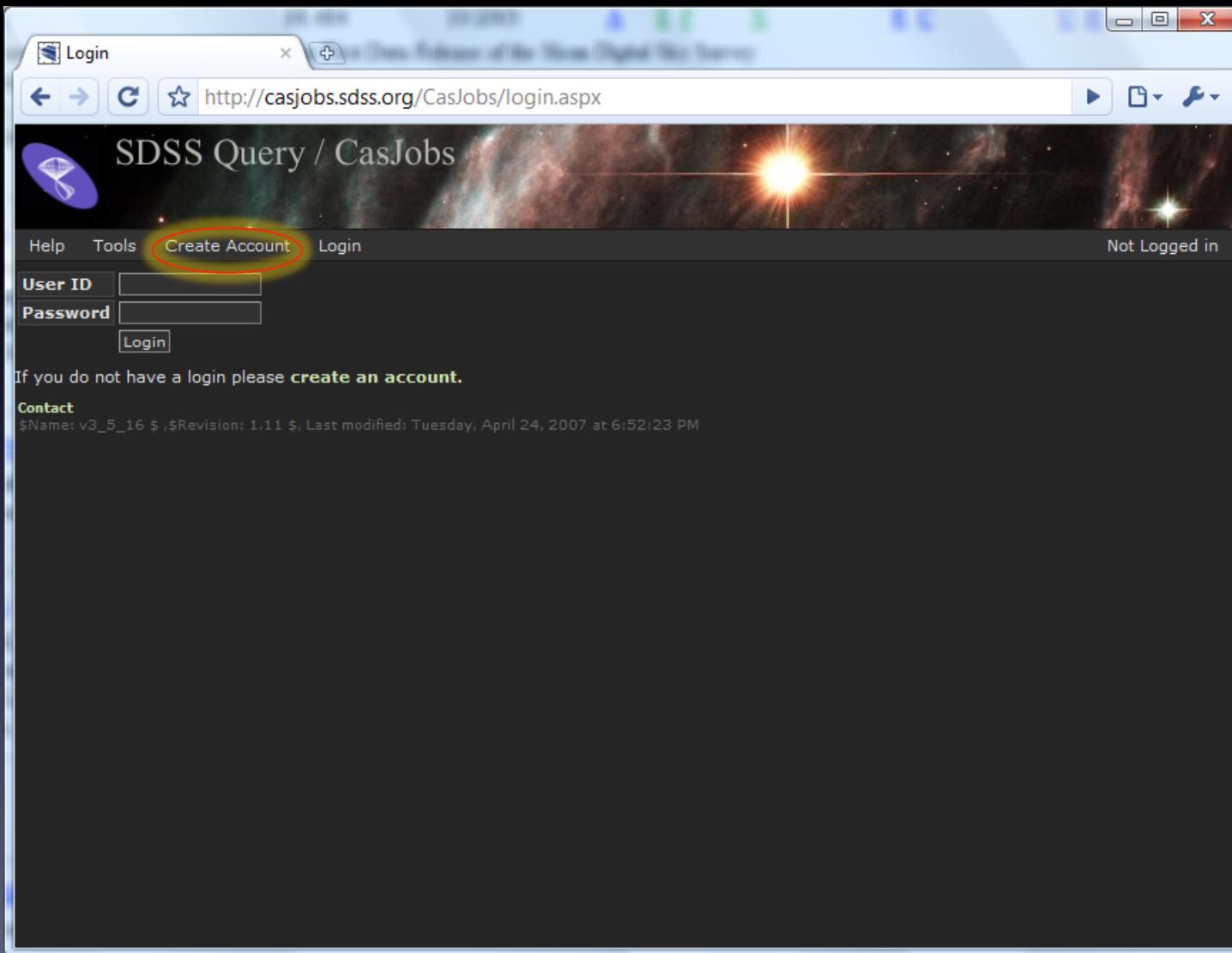
- Getting Started
- FAQ
- How To
- Glossary
- Schema Browser
- Sample SQL Queries
- Details of SDSS Data

**Contact Us**

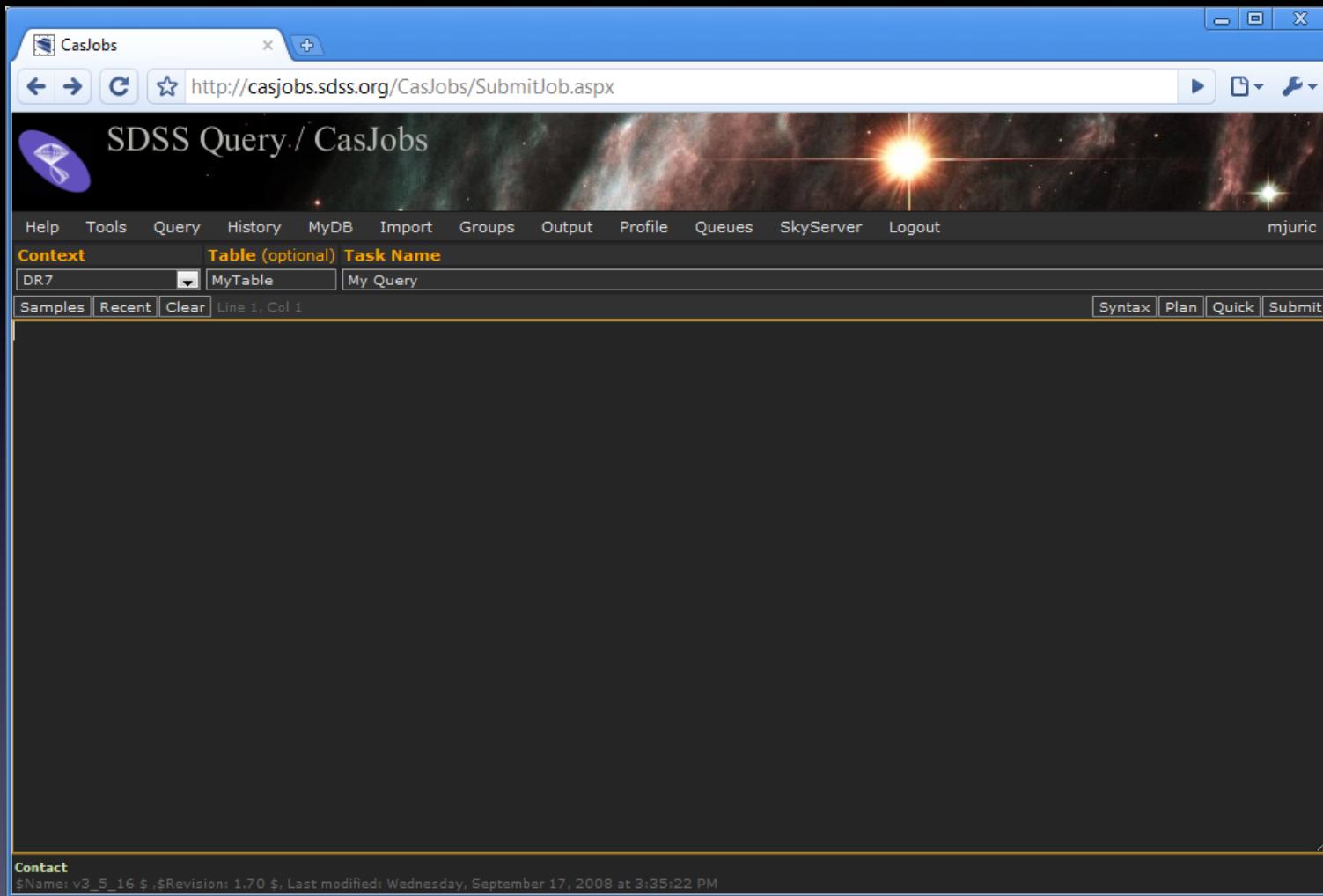
The contents for boundaries of the different reports

Report No.	Report Name	Description
1	SDSS DR7	SDSS DR7
2	SDSS DR7	SDSS DR7
3	SDSS DR7	SDSS DR7
4	SDSS DR7	SDSS DR7
5	SDSS DR7	SDSS DR7
6	SDSS DR7	SDSS DR7
7	SDSS DR7	SDSS DR7
8	SDSS DR7	SDSS DR7
9	SDSS DR7	SDSS DR7
10	SDSS DR7	SDSS DR7
11	SDSS DR7	SDSS DR7
12	SDSS DR7	SDSS DR7
13	SDSS DR7	SDSS DR7
14	SDSS DR7	SDSS DR7
15	SDSS DR7	SDSS DR7
16	SDSS DR7	SDSS DR7
17	SDSS DR7	SDSS DR7
18	SDSS DR7	SDSS DR7
19	SDSS DR7	SDSS DR7
20	SDSS DR7	SDSS DR7
21	SDSS DR7	SDSS DR7
22	SDSS DR7	SDSS DR7
23	SDSS DR7	SDSS DR7
24	SDSS DR7	SDSS DR7
25	SDSS DR7	SDSS DR7
26	SDSS DR7	SDSS DR7
27	SDSS DR7	SDSS DR7
28	SDSS DR7	SDSS DR7
29	SDSS DR7	SDSS DR7
30	SDSS DR7	SDSS DR7
31	SDSS DR7	SDSS DR7
32	SDSS DR7	SDSS DR7
33	SDSS DR7	SDSS DR7
34	SDSS DR7	SDSS DR7
35	SDSS DR7	SDSS DR7
36	SDSS DR7	SDSS DR7
37	SDSS DR7	SDSS DR7
38	SDSS DR7	SDSS DR7
39	SDSS DR7	SDSS DR7
40	SDSS DR7	SDSS DR7
41	SDSS DR7	SDSS DR7
42	SDSS DR7	SDSS DR7
43	SDSS DR7	SDSS DR7
44	SDSS DR7	SDSS DR7
45	SDSS DR7	SDSS DR7
46	SDSS DR7	SDSS DR7
47	SDSS DR7	SDSS DR7
48	SDSS DR7	SDSS DR7
49	SDSS DR7	SDSS DR7
50	SDSS DR7	SDSS DR7
51	SDSS DR7	SDSS DR7
52	SDSS DR7	SDSS DR7
53	SDSS DR7	SDSS DR7
54	SDSS DR7	SDSS DR7
55	SDSS DR7	SDSS DR7
56	SDSS DR7	SDSS DR7
57	SDSS DR7	SDSS DR7
58	SDSS DR7	SDSS DR7
59	SDSS DR7	SDSS DR7
60	SDSS DR7	SDSS DR7
61	SDSS DR7	SDSS DR7
62	SDSS DR7	SDSS DR7
63	SDSS DR7	SDSS DR7
64	SDSS DR7	SDSS DR7
65	SDSS DR7	SDSS DR7
66	SDSS DR7	SDSS DR7
67	SDSS DR7	SDSS DR7
68	SDSS DR7	SDSS DR7
69	SDSS DR7	SDSS DR7
70	SDSS DR7	SDSS DR7
71	SDSS DR7	SDSS DR7
72	SDSS DR7	SDSS DR7
73	SDSS DR7	SDSS DR7
74	SDSS DR7	SDSS DR7
75	SDSS DR7	SDSS DR7
76	SDSS DR7	SDSS DR7
77	SDSS DR7	SDSS DR7
78	SDSS DR7	SDSS DR7
79	SDSS DR7	SDSS DR7
80	SDSS DR7	SDSS DR7
81	SDSS DR7	SDSS DR7
82	SDSS DR7	SDSS DR7
83	SDSS DR7	SDSS DR7
84	SDSS DR7	SDSS DR7
85	SDSS DR7	SDSS DR7
86	SDSS DR7	SDSS DR7
87	SDSS DR7	SDSS DR7
88	SDSS DR7	SDSS DR7
89	SDSS DR7	SDSS DR7
90	SDSS DR7	SDSS DR7
91	SDSS DR7	SDSS DR7
92	SDSS DR7	SDSS DR7
93	SDSS DR7	SDSS DR7
94	SDSS DR7	SDSS DR7
95	SDSS DR7	SDSS DR7
96	SDSS DR7	SDSS DR7
97	SDSS DR7	SDSS DR7
98	SDSS DR7	SDSS DR7
99	SDSS DR7	SDSS DR7
100	SDSS DR7	SDSS DR7
101	SDSS DR7	SDSS DR7
102	SDSS DR7	SDSS DR7
103	SDSS DR7	SDSS DR7
104	SDSS DR7	SDSS DR7
105	SDSS DR7	SDSS DR7
106	SDSS DR7	SDSS DR7
107	SDSS DR7	SDSS DR7
108	SDSS DR7	SDSS DR7
109	SDSS DR7	SDSS DR7
110	SDSS DR7	SDSS DR7
111	SDSS DR7	SDSS DR7
112	SDSS DR7	SDSS DR7
113	SDSS DR7	SDSS DR7
114	SDSS DR7	SDSS DR7
115	SDSS DR7	SDSS DR7
116	SDSS DR7	SDSS DR7
117	SDSS DR7	SDSS DR7
118	SDSS DR7	SDSS DR7
119	SDSS DR7	SDSS DR7
120	SDSS DR7	SDSS DR7
121	SDSS DR7	SDSS DR7
122	SDSS DR7	SDSS DR7
123	SDSS DR7	SDSS DR7
124	SDSS DR7	SDSS DR7
125	SDSS DR7	SDSS DR7
126	SDSS DR7	SDSS DR7
127	SDSS DR7	SDSS DR7
128	SDSS DR7	SDSS DR7
129	SDSS DR7	SDSS DR7
130	SDSS DR7	SDSS DR7
131	SDSS DR7	SDSS DR7
132	SDSS DR7	SDSS DR7
133	SDSS DR7	SDSS DR7
134	SDSS DR7	SDSS DR7
135	SDSS DR7	SDSS DR7
136	SDSS DR7	SDSS DR7
137	SDSS DR7	SDSS DR7
138	SDSS DR7	SDSS DR7
139	SDSS DR7	SDSS DR7
140	SDSS DR7	SDSS DR7
141	SDSS DR7	SDSS DR7
142	SDSS DR7	SDSS DR7
143	SDSS DR7	SDSS DR7
144	SDSS DR7	SDSS DR7
145	SDSS DR7	SDSS DR7
146	SDSS DR7	SDSS DR7
147	SDSS DR7	SDSS DR7
148	SDSS DR7	SDSS DR7
149	SDSS DR7	SDSS DR7
150	SDSS DR7	SDSS DR7
151	SDSS DR7	SDSS DR7
152	SDSS DR7	SDSS DR7
153	SDSS DR7	SDSS DR7
154	SDSS DR7	SDSS DR7
155	SDSS DR7	SDSS DR7
156	SDSS DR7	SDSS DR7
157	SDSS DR7	SDSS DR7
158	SDSS DR7	SDSS DR7
159	SDSS DR7	SDSS DR7
160	SDSS DR7	SDSS DR7
161	SDSS DR7	SDSS DR7
162	SDSS DR7	SDSS DR7
163	SDSS DR7	SDSS DR7
164	SDSS DR7	SDSS DR7
165	SDSS DR7	SDSS DR7
166	SDSS DR7	SDSS DR7
167	SDSS DR7	SDSS DR7
168	SDSS DR7	SDSS DR7
169	SDSS DR7	SDSS DR7
170	SDSS DR7	SDSS DR7
171	SDSS DR7	SDSS DR7
172	SDSS DR7	SDSS DR7
173	SDSS DR7	SDSS DR7
174	SDSS DR7	SDSS DR7
175	SDSS DR7	SDSS DR7
176	SDSS DR7	SDSS DR7
177	SDSS DR7	SDSS DR7
178	SDSS DR7	SDSS DR7
179	SDSS DR7	SDSS DR7
180	SDSS DR7	SDSS DR7
181	SDSS DR7	SDSS DR7
182	SDSS DR7	SDSS DR7
183	SDSS DR7	SDSS DR7
184	SDSS DR7	SDSS DR7
185	SDSS DR7	SDSS DR7
186	SDSS DR7	SDSS DR7
187	SDSS DR7	SDSS DR7
188	SDSS DR7	SDSS DR7
189	SDSS DR7	SDSS DR7
190	SDSS DR7	SDSS DR7
191	SDSS DR7	SDSS DR7
192	SDSS DR7	SDSS DR7
193	SDSS DR7	SDSS DR7
194	SDSS DR7	SDSS DR7
195	SDSS DR7	SDSS DR7
196	SDSS DR7	SDSS DR7
197	SDSS DR7	SDSS DR7
198	SDSS DR7	SDSS DR7
199	SDSS DR7	SDSS DR7
200	SDSS DR7	SDSS DR7
201	SDSS DR7	SDSS DR7
202	SDSS DR7	SDSS DR7
203	SDSS DR7	SDSS DR7
204	SDSS DR7	SDSS DR7
205	SDSS DR7	SDSS DR7
206	SDSS DR7	SDSS DR7
207	SDSS DR7	SDSS DR7
208	SDSS DR7	SDSS DR7
209	SDSS DR7	SDSS DR7
210	SDSS DR7	SDSS DR7
211	SDSS DR7	SDSS DR7
212	SDSS DR7	SDSS DR7
213	SDSS DR7	SDSS DR7
214	SDSS DR7	SDSS DR7
215	SDSS DR7	SDSS DR7
216	SDSS DR7	SDSS DR7
217	SDSS DR7	SDSS DR7
218	SDSS DR7	SDSS DR7
219	SDSS DR7	SDSS DR7
220	SDSS DR7	SDSS DR7
221	SDSS DR7	SDSS DR7
222	SDSS DR7	SDSS DR7
223	SDSS DR7	SDSS DR7
224	SDSS DR7	SDSS DR7
225	SDSS DR7	SDSS DR7
226	SDSS DR7	SDSS DR7
227	SDSS DR7	SDSS DR7
228	SDSS DR7	SDSS DR7
229	SDSS DR7	SDSS DR7
230	SDSS DR7	SDSS DR7
231	SDSS DR7	SDSS DR7
232	SDSS DR7	SDSS DR7
233	SDSS DR7	SDSS DR7
234	SDSS DR7	SDSS DR7
235	SDSS DR7	SDSS DR7
236	SDSS DR7	SDSS DR7
237	SDSS DR7	SDSS DR7
238	SDSS DR7	SDSS DR7
239	SDSS DR7	SDSS DR7
240	SDSS DR7	SDSS DR7
241	SDSS DR7	SDSS DR7
242	SDSS DR7	SDSS DR7
243	SDSS DR7	SDSS DR7
244	SDSS DR7	SDSS DR7
245	SDSS DR7	SDSS DR7
246	SDSS DR7	SDSS DR7
247	SDSS DR7	SDSS DR7
248	SDSS DR7	SDSS DR7
249	SDSS DR7	SDSS DR7
250	SDSS DR7	SDSS DR7
251	SDSS DR7	SDSS DR7
252	SDSS DR7	SDSS DR7
253	SDSS DR7	SDSS DR7
254	SDSS DR7	SDSS DR7
255	SDSS DR7	SDSS DR7
256	SDSS DR7	SDSS DR7
257	SDSS DR7	SDSS DR7
258	SDSS DR7	SDSS DR7
259	SDSS DR7	SDSS DR7
260	SDSS DR7	SDSS DR7
261	SDSS DR7	SDSS DR7
262	SDSS DR7	SDSS DR7
263	SDSS DR7	SDSS DR7
264	SDSS DR7	SDSS DR7
265	SDSS DR7	SDSS DR7
266	SDSS DR7	SDSS DR7
267	SDSS DR7	SDSS DR7
268	SDSS DR7	SDSS DR7
269	SDSS DR7	SDSS DR7
270	SDSS DR7	SDSS DR7
271	SDSS DR7	SDSS DR7
272	SDSS DR7	SDSS DR7
273	SDSS DR7	SDSS DR7
274	SDSS DR7	SDSS DR7
275	SDSS DR7	SDSS DR7
276	SDSS DR7	SDSS DR7
277	SDSS DR7	SDSS DR7
278	SDSS DR7	SDSS DR7
279	SDSS DR7	SDSS DR7
280	SDSS DR7	SDSS DR7
281	SDSS DR7	SDSS DR7
282	SDSS DR7	SDSS DR7
283	SDSS DR7	SDSS DR7
284	SDSS DR7	SDSS DR7
285	SDSS DR7	SDSS DR7
286	SDSS DR7	SDSS DR7
287	SDSS DR7	SDSS DR7
288	SDSS DR7	SDSS DR7
289	SDSS DR7	SDSS DR7
290	SDSS DR7	SDSS DR7
291	SDSS DR7	SDSS DR7
292	SDSS DR7	SDSS DR7
293	SDSS DR7	SDSS DR7
294	SDSS DR7	SDSS DR7
295	SDSS DR7	SDSS DR7
296	SDSS DR7	SDSS DR7
297	SDSS DR7	SDSS DR7
298	SDSS DR7	SDSS DR7
299	SDSS DR7	SDSS DR7
300	SDSS DR7	SDSS DR7
301	SDSS DR7	SDSS DR7
302	SDSS DR7	SDSS DR7
303	SDSS DR7	SDSS DR7
304	SDSS DR7	SDSS DR7
305	SDSS DR7	SDSS DR7
306	SDSS DR7	SDSS DR7
307	SDSS DR7	SDSS DR7
308	SDSS DR7	SDSS DR7
309	SDSS DR7	SDSS DR7
310	SDSS DR7	SDSS DR7
311	SDSS DR7	SDSS DR7
312	SDSS DR7	SDSS DR7
313	SDSS DR7	SDSS DR7
314	SDSS DR7	SDSS DR7
315	SDSS DR7	SDSS DR7
316	SDSS DR7	SDSS DR7
317	SDSS DR7	SDSS DR7
318	SDSS DR7	SDSS DR7
319	SDSS DR7	SDSS DR7
320	SDSS DR7	SDSS DR7
321	SDSS DR7	SDSS DR7
322	SDSS DR7	SDSS DR7
323	SDSS DR7	SDSS DR7
324	SDSS DR7	SDSS DR7
325	SDSS DR7	SDSS DR7
326	SDSS DR7	SDSS DR7
327	SDSS DR7	SDSS DR7
328	SDSS DR7	SDSS DR7
329	SDSS DR7	SDSS DR7
330	SDSS DR7	SDSS DR7
331	SDSS DR7	SDSS DR7
332	SDSS DR7	SDSS DR7
333	SDSS DR7	SDSS DR7
334	SDSS DR7	SDSS DR7
335	SDSS DR7	SDSS DR7
336	SDSS DR7	SDSS DR7
337	SDSS DR7	SDSS DR7
338	SDSS DR7	SDSS DR7
339	SDSS DR7	SDSS DR7
340	SDSS DR7	SDSS DR7
341	SDSS DR7	SDSS DR7
342	SDSS DR7	SDSS DR7
343	SDSS DR7	SDSS DR7
344	SDSS DR7	SDSS DR7
345	SDSS DR7	SDSS DR7
346	SDSS DR7	SDSS DR7
347	SDSS DR7	SDSS DR7
348	SDSS DR7	SDSS DR7
349	SDSS DR7	SDSS DR7
350	SDSS DR7	SDSS DR7
351	SDSS DR7	SDSS DR7
352	SDSS DR7	SDSS DR7
353	SDSS DR7	SDSS DR7
354	SDSS DR7	SDSS DR7
355	SDSS DR7	SDSS DR7
356	SDSS DR7	SDSS DR7
357	SDSS DR7	SDSS DR7
358	SDSS DR7	SDSS DR7
359	SDSS DR7	SDSS DR7
360	SDSS DR7	SDSS DR7
361	SDSS DR7	SDSS DR7
362	SDSS DR7	SDSS DR7
363	SDSS DR7	SDSS DR7
364	SDSS DR7	SDSS DR7
365	SDSS DR7	SDSS DR7
366	SDSS DR7	SDSS DR7
367	SDSS DR7	SDSS DR7
368	SDSS DR7	SDSS DR7
369	SDSS DR7	SDSS DR7
370	SDSS DR7	SDSS DR7
371	SDSS DR7	SDSS DR7
372	SDSS DR7	SDSS DR7
373	SDSS DR7	SDSS DR7
374	SDSS DR7	SDSS DR7
375	SDSS DR7	SDSS DR7
376	SDSS DR7	SDSS DR7
377	SDSS DR7	SDSS DR7
378	SDSS DR7	SDSS DR7
379	SDSS DR7	SDSS DR7
380	SDSS DR7	SDSS DR7
381	SDSS DR7	SDSS DR7
382	SDSS DR7	SDSS DR7
383	SDSS DR7	SDSS DR7
384	SDSS DR7	SDSS DR7
385	SDSS DR7	SDSS DR7
386	SDSS DR7	SDSS DR7
387	SDSS DR7	SDSS DR7
388	SDSS DR7	SDSS DR7
389	SDSS DR7	SDSS DR7
390	SDSS DR7	SDSS DR7
391	SDSS DR7</td	

## CasJobs: Running long and complex queries



# SDSS CasJobs Service



CasJobs

http://casjobs.sdss.org/CasJobs/SubmitJob.aspx

SDSS Query./ CasJobs

Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout mjuric

Context Table (optional) Task Name

DR7 MyTable My Query

Samples Recent Clear Line 1, Col 1 Syntax Plan Quick Submit

Example:

Get the positions (in Galactic coordinates) and dereddened magnitudes of all stars within 1 degree of  $\alpha = 341.6$ ,  $\delta = 31.7$

Contact  
\$Name: v3\_5\_16 \$,\$Revision: 1.70 \$, Last modified: Wednesday, September 17, 2008 at 3:35:22 PM

The screenshot shows a web browser window titled 'CasJobs' with the URL 'http://casjobs.sdss.org/CasJobs/SubmitJob.aspx'. The page header includes the SDSS logo and the title 'SDSS Query./ CasJobs'. A navigation bar at the top has links for Help, Tools, Query, History, MyDB, Import, Groups, Output, Profile, Queues, SkyServer, and Logout. A user name 'mjuric' is visible on the right. Below the header is a search form with fields for 'Context' (set to DR7), 'Table (optional)' (set to MyTable), and 'Task Name' (set to My Query). There are buttons for Samples, Recent, Clear, Line 1, Col 1, Syntax, Plan, Quick, and Submit. The main area contains the text 'Example:' and 'Get the positions (in Galactic coordinates) and dereddened magnitudes of all stars within 1 degree of  $\alpha = 341.6$ ,  $\delta = 31.7$ '. At the bottom, there is a 'Contact' section with revision information: '\$Name: v3\_5\_16 \$,\$Revision: 1.70 \$, Last modified: Wednesday, September 17, 2008 at 3:35:22 PM'.

# An SQL Query

The screenshot shows a web-based SQL query interface titled "SDSS Query / CasJobs". The interface has a blue header bar with the title "CasJobs" and a back/forward, refresh, and search toolbar. Below the header is a URL bar showing "http://casjobs.sdss.org/CasJobs/SubmitJob.aspx". The main area is titled "SDSS Query / CasJobs" and features a logo of a telescope. A sidebar on the right contains a brief introduction to SQL.

**SQL – Structured Query Language**  
A language for specifying and extracting subsets of data from large tabular databases.  
Allows you to select only the data that satisfy a certain set of (possibly very complicated) restrictions that you specify

**Context**      **Table (optional)** **Task Name**  
DR7      MyTable      My Query  
Samples Recent Clear Line 18, Col 1

```
select

s.objid,
s.l, s.b,
s.extinction_r as Ar,
s.dered_u as u, s.dered_g as g, s.dered_r as r, s.dered_i as i, s.dered_z as z,
s.err_u, s.err_g, s.err_r, s.err_i, s.err_z,
s.flags

from
fGetNearbyObjEq(341.6, 31.7, 60) n, Star s

into
mydb.field1

where
n.objID=s.objID
```

# Anatomy of an SQL Query

The screenshot shows the SDSS Query / CasJobs interface. At the top, there's a toolbar with icons for back, forward, search, and a plus sign. Below it is a navigation bar with links for Help, Tools, Query, History, MyDB, Import, Groups, Output, Context, Table (optional), and Task Name. The Context dropdown is set to DR7, the Table dropdown is set to MyTable, and the Task Name dropdown is set to My Query. Below the toolbar, there are tabs for Samples, Recent, and Clear, with 'Line 18, Col 1' selected.

The main area contains an SQL query:

```
select s.objid,  
       s.l, s.b,  
       s.extinction_r as Ar,  
       s.dered_u as u, s.dered_g as g, s.dered_r as r, s.dered_i as i, s.dered_z as z,  
       s.err_u, s.err_g, s.err_r, s.err_i, s.err_z,  
       s.flags  
  
from fGetNearbyObjEq(341.6, 31.7, 60) n, Star s  
  
into mydb.field1  
  
where n.objID=s.objID
```

Annotations explain the parts of the query:

- SELECT** – select a subset of entries from the catalog, based on the criteria we will specify. (points to the 'select' keyword)
- Columns** – the columns we are interested in. Usually we don't want all (> 50!) available columns, but just some (e.g., id, position, extinction, magnitudes, errors, and flags). (points to the column list)
- FROM clause** – the tables from which to fetch the columns above. Can be "virtual". (points to the 'from' clause)
- INTO clause** – where to store the result. The results go to tables in your personal database. (points to the 'into' clause)
- WHERE clause** – the filter with which to filter the result. (points to the 'where' clause)

# Submitted Job

CasJobs 

http://casjobs.sdss.org/CasJobs/jobdetails.aspx?id=4928887

## SDSS Query / CasJobs

Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout mjuric

### 'My Query' Details

This page will automatically refresh every 30 seconds

There are 0 jobs ahead of this one in the DR7 queue.

JobID	TaskName	Context	Queue	Submitted	Started	Finished	Status
4928887	My Query	DR7	600	7/28/2010 4:33:04 AM			Ready

Executed on    
0 No Message

**Query**

```
select
s.objid,
s.l, s.b,
s.extinction_r as Ar,
s.dered_u as u, s.dered_g as g, s.dered_r as r, s.dered_i as i, s.dered_z as z,
s.err_u, s.err_g, s.err_r, s.err_i, s.err_z,
s.flags

from
fGetNearbyObjEq(341.6, 31.7, 60) n, Star s

into
mydb.field2

where
n.objID=s.objID
```

Contact  
\$Name: v3\_5\_16 \$,\$Revision: 1.23 \$, Last modified: Wednesday, May 14, 2008 at 1:52:25 AM

# Completed Job

The screenshot shows a web browser window titled "CasJobs" displaying the "SDSS Query / CasJobs" interface. The URL in the address bar is <http://casjobs.sdss.org/CasJobs/jobdetails.aspx?id=4928831&message=Details%20of%204928831>. The top navigation bar includes links for Help, Tools, Query, History, MyDB (which is highlighted with a yellow oval), Import, Groups, Output, Profile, Queues, SkyServer, and Logout. The user name "mjuric" is visible on the right.

The main content area is titled "'My Query' Details". It shows a table of job details:

JobID	TaskName	Context	Queue	Submitted	Started	Finished	Status
4928831	My Query	DR7	600	7/28/2010 4:03:05 AM	7/28/2010 4:03:13 AM	7/28/2010 4:03:43 AM	Finished

Below the table, it says "Executed on Rows DR7Best long 70454" and "Message: Query Complete".

The "Query" section contains the following SQL code:

```
select  
s.objid,  
s.l, s.b,  
s.extinction_r as Ar,  
s.dered_u as u, s.dered_g as g, s.dered_r as r, s.dered_i as i, s.dered_z as z,  
s.err_u, s.err_g, s.err_r, s.err_i, s.err_z,  
s.flags  
  
from  
fGetNearbyObjEq(341.6, 31.7, 60) n, Star s  
  
into  
mydb.field1  
  
where  
n.objID=s.objID
```

At the bottom left, there is a "Contact" link and a note: "\$Name: v3\_5\_16 \$,\$Revision: 1.23 \$, Last modified: Wednesday, May 14, 2008 at 1:52:25 AM\$". A white arrow points from the text "Query Complete" in the "Message" box towards the bottom-left corner of the query results table.

MyDB    miner - Google Search

http://casjobs.sdss.org/CasJobs/MyDB.aspx

# SDSS Query / CasJobs

Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout mjuric

MyDB Local Only

Views

Tables

Functions

Procedures

Sort by... All selected...

Rows	kB	Name
70,454	9,160	field1

Mario Juric 's MyDB

9,624 kB of 2,000,000 kB used

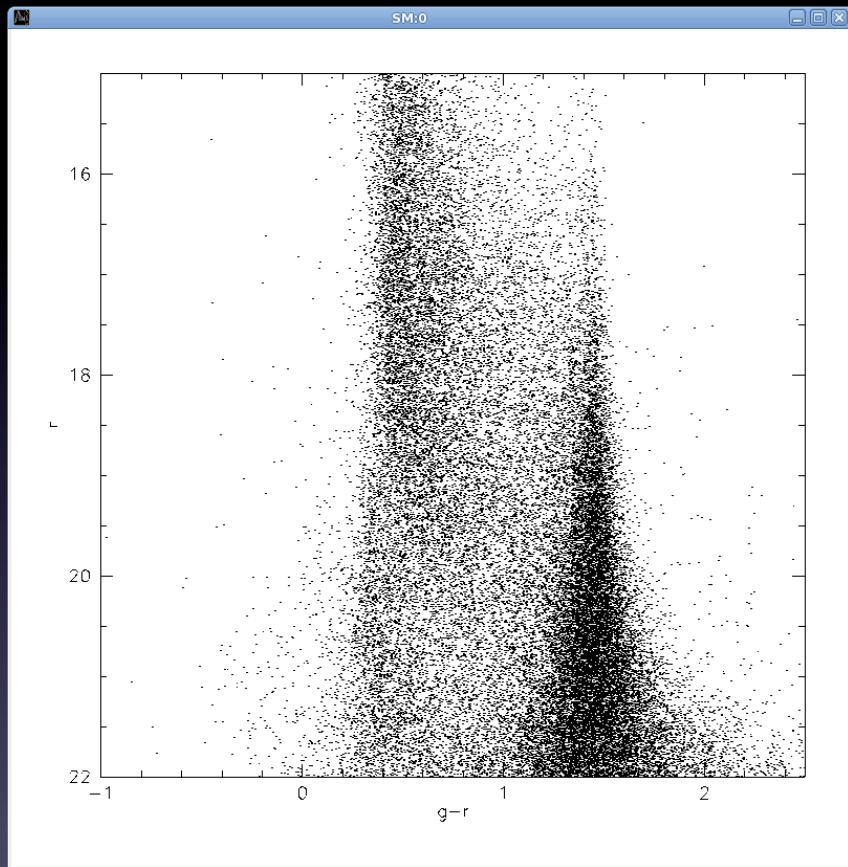
From this page you can get various information about the contents of both your MyDB and shared tables within your groups. Click the left table links to get information about a specific table, such as rows, columns or size. From the table pages you can also perform various table-specific tasks, such as:

- Download a table
- Manage your group tables
- Rename a table
- Drop a table

Sizes are approximations only.  
Row counts are approximations only. For exact value run a count.  
There's always some overhead, even empty MyDB's take up space.  
Group tables do not count towards your MyDB size limit.

Contact  
Name: v3\_5\_16 \$ , Revision: 1.64 \$, Last modified: Tuesday, January 27, 2009 at 3:19:32 PM

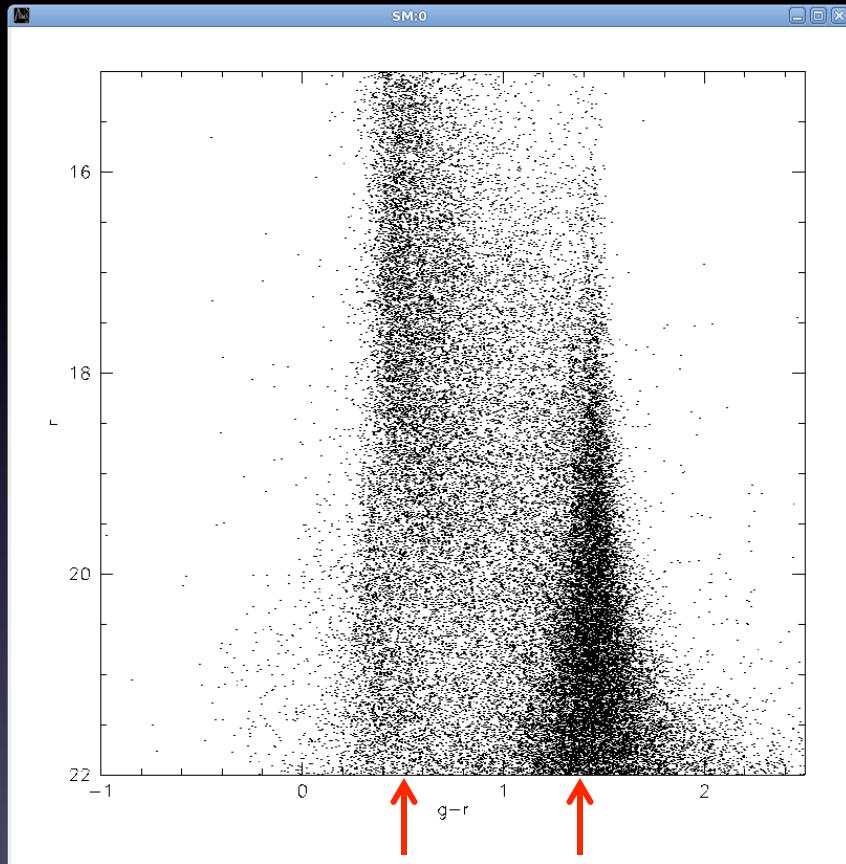
# Basic Data Analysis: A Color-Magnitude Diagram (CMD)



```
Untitled - Notepad
File Edit Format View Help
data "field1.csv"
read <1 2 b 3 Ar 4 u 5 g 6 r 7 i 8 z 9>
set gr = g-r
lweight 2
expand 1.001
limits -1 2.5 22 15
ptype 0 0
erase
box
points gr r
xlabel g-r
ylabel r
```

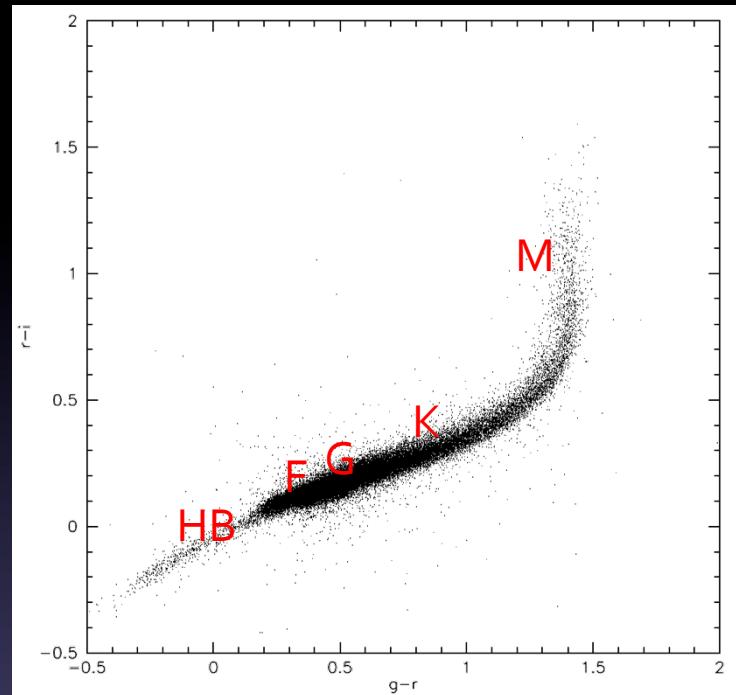
The screenshot shows a Windows Notepad window titled "Untitled - Notepad". The window contains Starmax command language (SM:0) code. The code reads data from a CSV file named "field1.csv" and performs several operations: it sets the color axis to "g-r", uses a line weight of 2, expands the data by a factor of 1.001, sets the limits for color (g-r) from -1 to 2.5 and magnitude (r) from 22 to 15, sets the point type to 0, erases the previous plot, draws a box around the data, and finally plots the data points with the color axis labeled "g-r" and the magnitude axis labeled "r".

# Interpreting Stellar CMDs

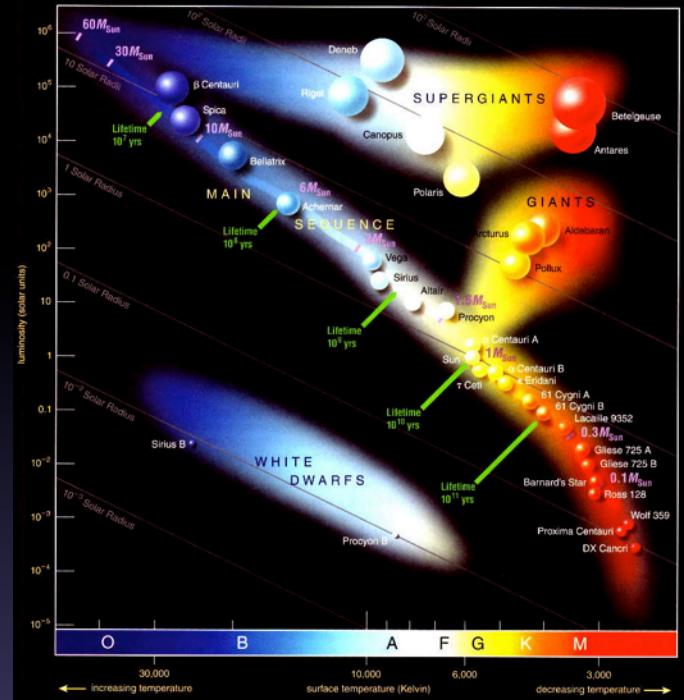
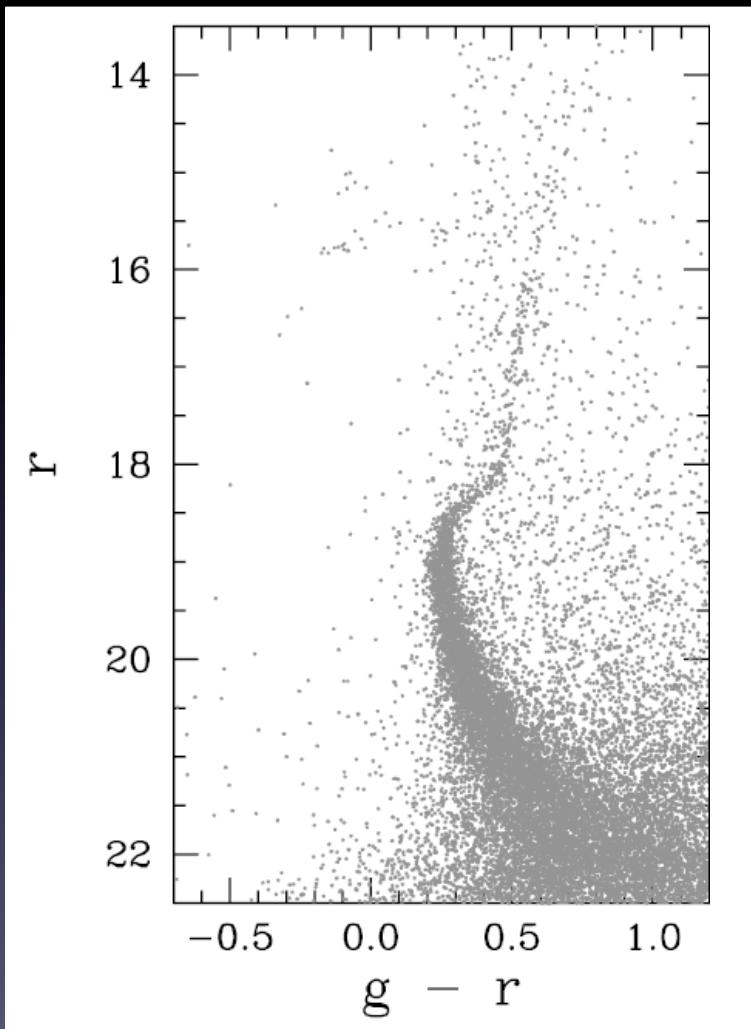


← More  
Luminous

Less →  
Luminous



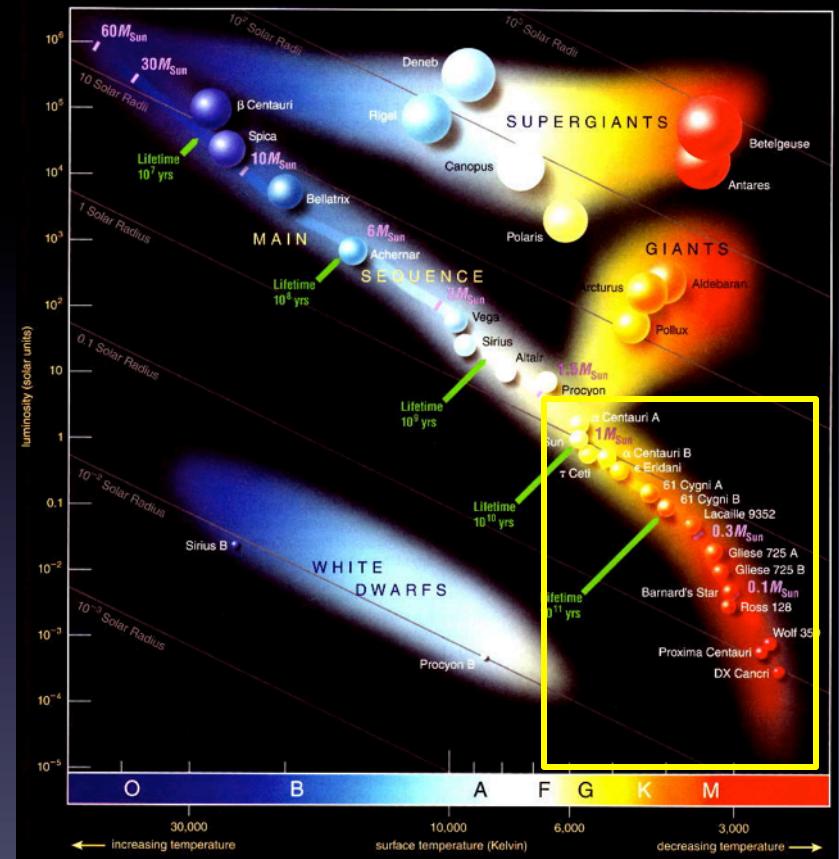
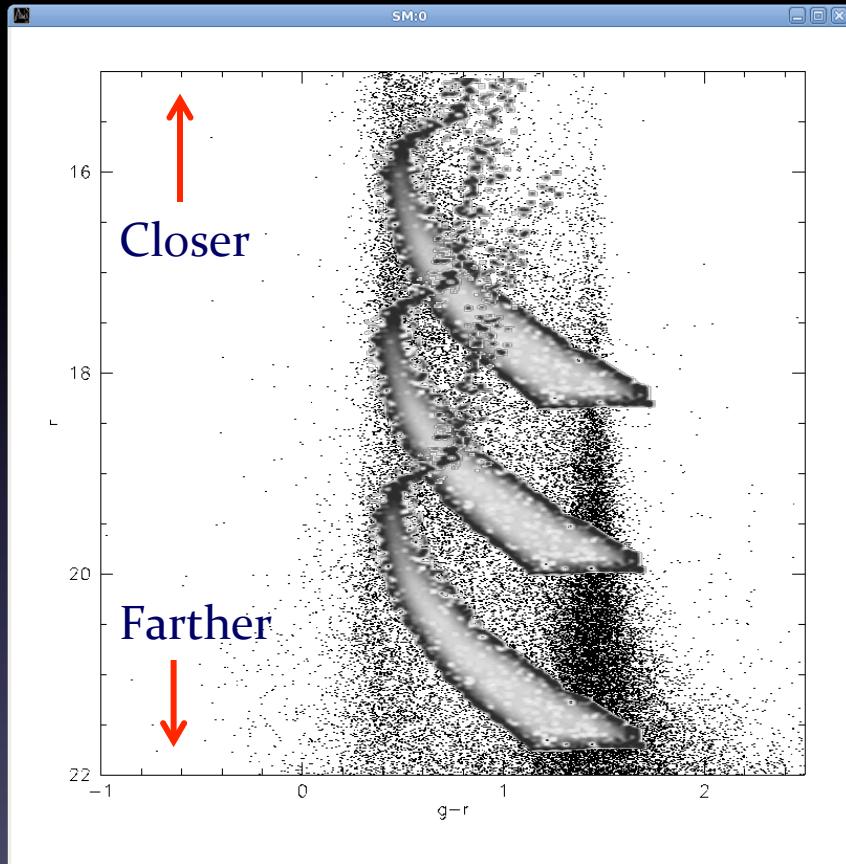
# A CMD of a Globular Cluster



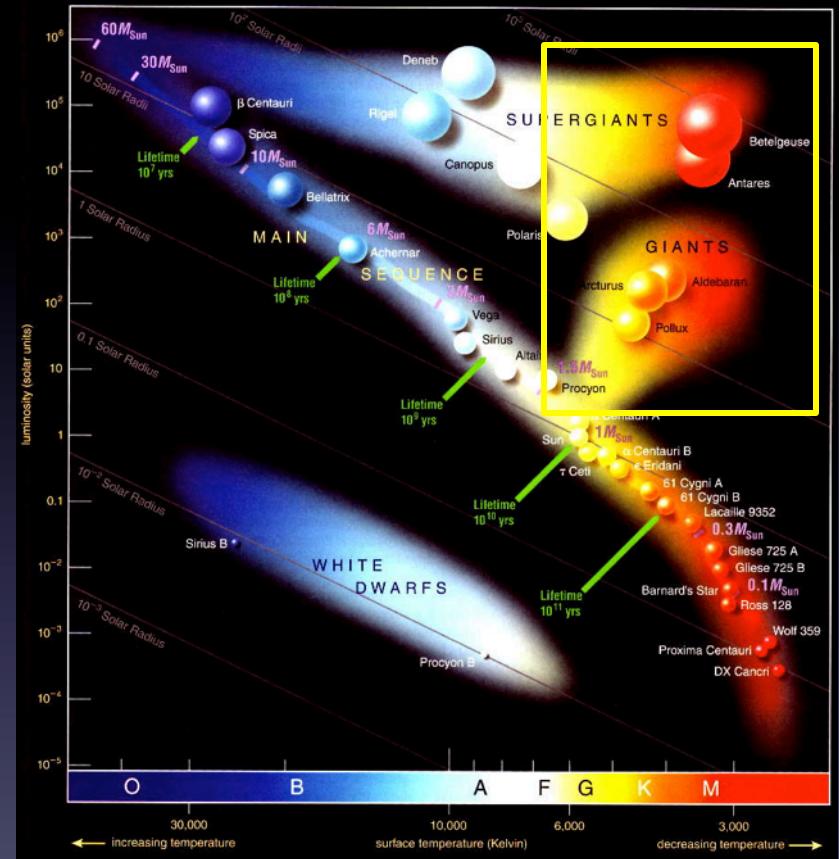
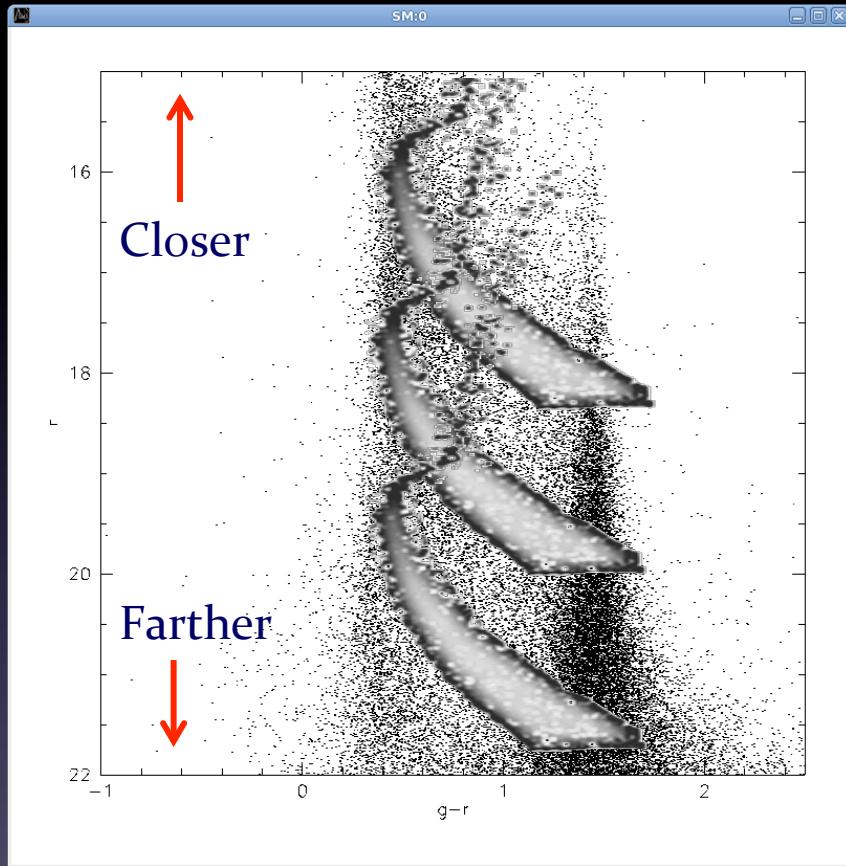
A globular cluster is a simple population (same age, same metallicity), at the same distance.

$$m = M + 5 \log_{10}(d/10\text{kpc})$$

# Interpreting “Field” CMDs



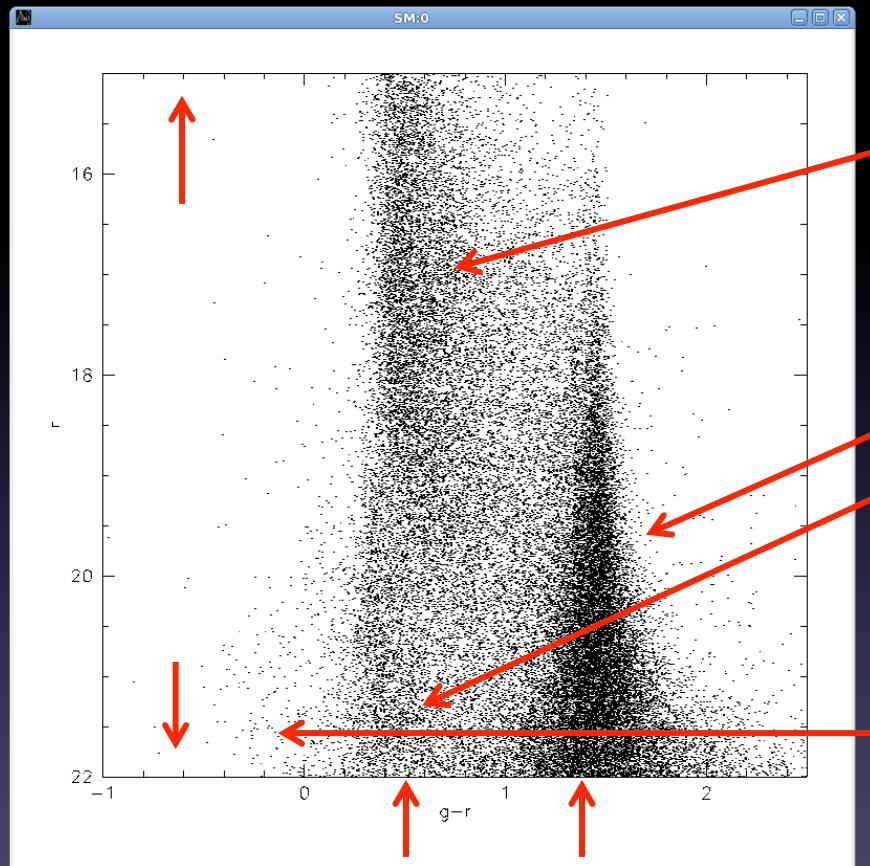
# Interpreting “Field” CMDs



What about the giants? Yes, they are there, but significantly less numerous (<5%) than the main sequence stars.

The same applies for early type stars, white dwarfs, etc.

# Basic Data Analysis: A Color-Magnitude Diagram



F stars M stars

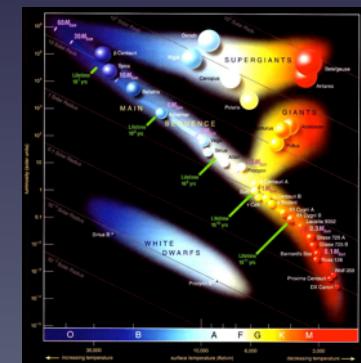
← More Luminous

Less →  
Luminous

Disk G/F stars

Nearby Disk K/M stars  
Halo stars

QSOs, RRLyrae,  
WDs, errors...



CasJobs

http://casjobs.sdss.org/CasJobs/SubmitJob.aspx

## SDSS Query./ CasJobs

Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout mjuric

Context Table (optional) Task Name

DR7 MyTable My Query

Samples Recent Clear Line 1, Col 1 Syntax Plan Quick Submit

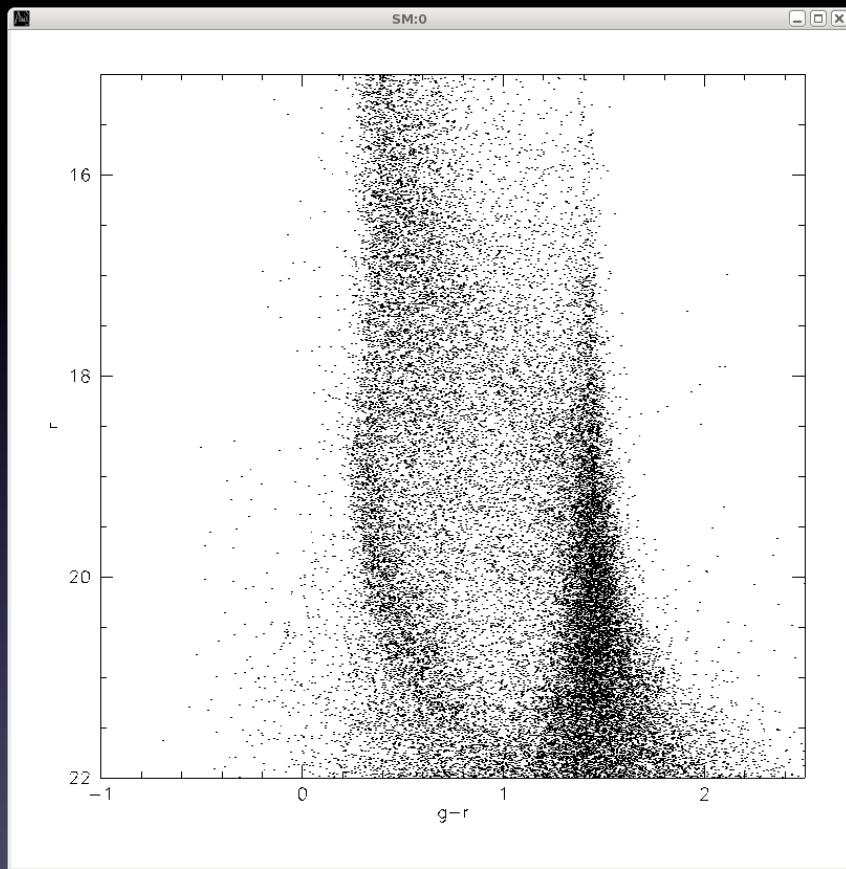
Get the positions (in Galactic coordinates) and dereddened magnitudes of all stars within 1 degree of  $\alpha = 21^{\circ}16', \delta = 31^{\circ}7'$

$\alpha = 115.2, \delta = 32.7$

Contact  
\$Name: v3\_5\_16 \$,\$Revision: 1.70 \$, Last modified: Wednesday, September 17, 2008 at 3:35:22 PM

The screenshot shows a web-based query interface for the SDSS. At the top, there's a header bar with the title 'CasJobs' and a URL 'http://casjobs.sdss.org/CasJobs/SubmitJob.aspx'. Below the header is a banner featuring a starry background with a prominent central star. The main menu includes 'Help', 'Tools', 'Query', 'History', 'MyDB', 'Import', 'Groups', 'Output', 'Profile', 'Queues', 'SkyServer', and 'Logout'. A user name 'mjuric' is visible on the right. The 'Query' section has a 'Context' dropdown set to 'DR7', a 'Table (optional)' dropdown set to 'MyTable', and a 'Task Name' input field set to 'My Query'. Below these are buttons for 'Samples', 'Recent', 'Clear', and 'Line 1, Col 1'. To the right of these buttons are four more buttons: 'Syntax', 'Plan', 'Quick', and 'Submit'. The main body of the page contains a large text area with the query instructions. At the bottom, there's a 'Contact' section with revision information.

# Color-Magnitude Diagram

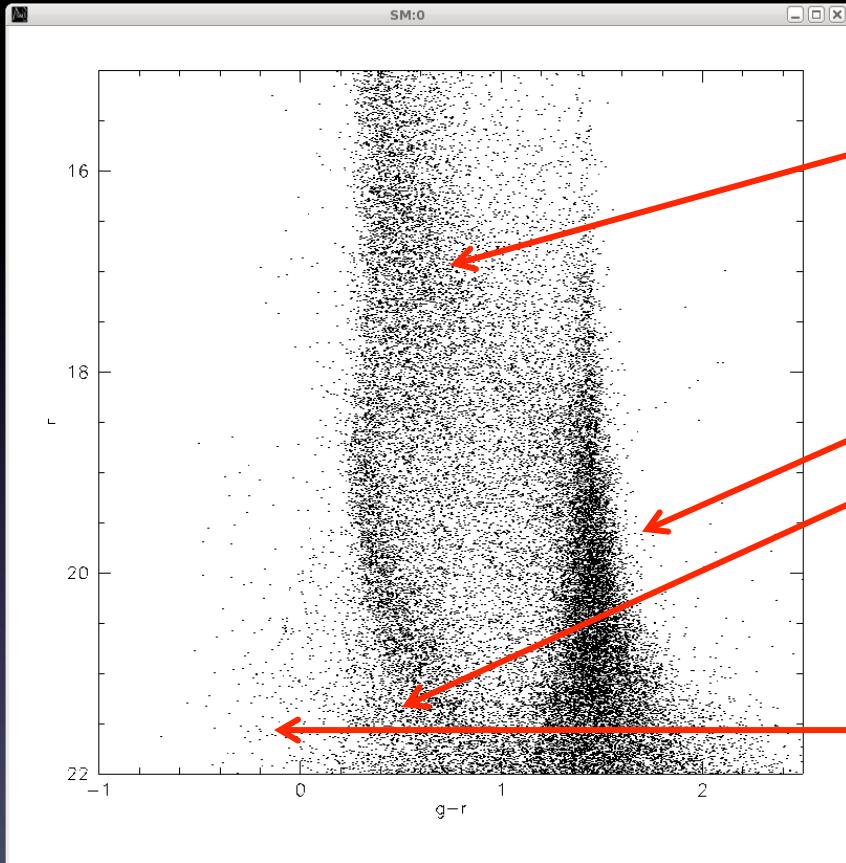


Untitled - Notepad

```
File Edit Format View Help
data "field2.csv"
read <1 2 b 3 Ar 4 u 5 g 6 r 7 i 8 z 9>

set gr = g-r
lweight 2
expand 1.001
llimits -1 2.5 22 15
ptype 0 0
erase
box
points gr r
xlabel g-r
ylabel r
```

# Color-Magnitude Diagram



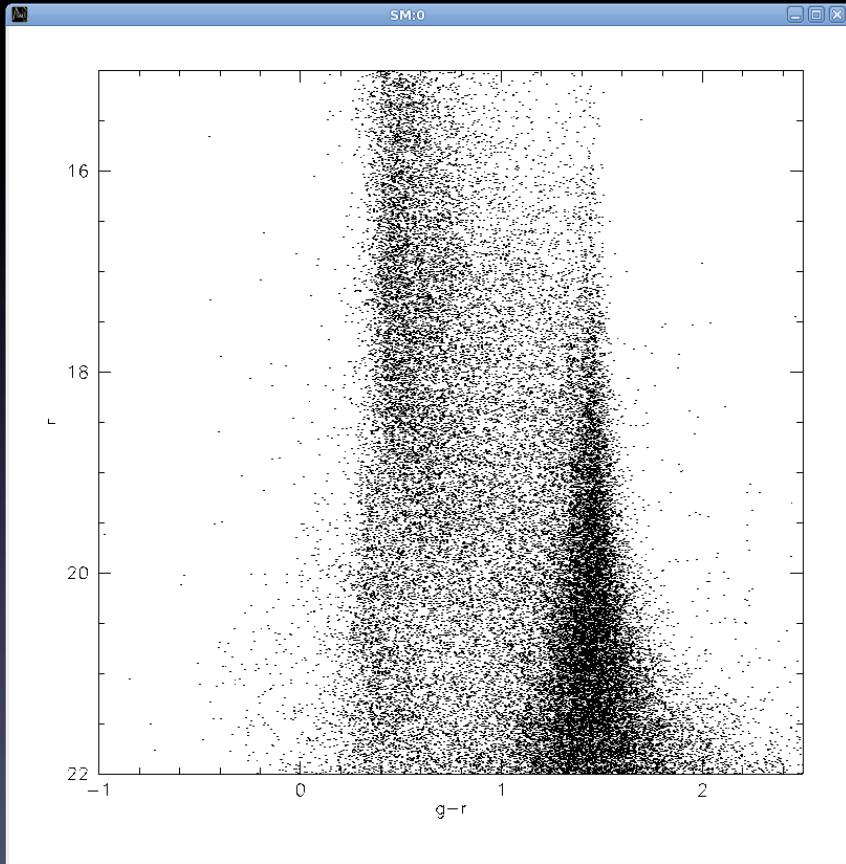
Disk G/F stars

Nearby Disk K/M stars  
Halo stars

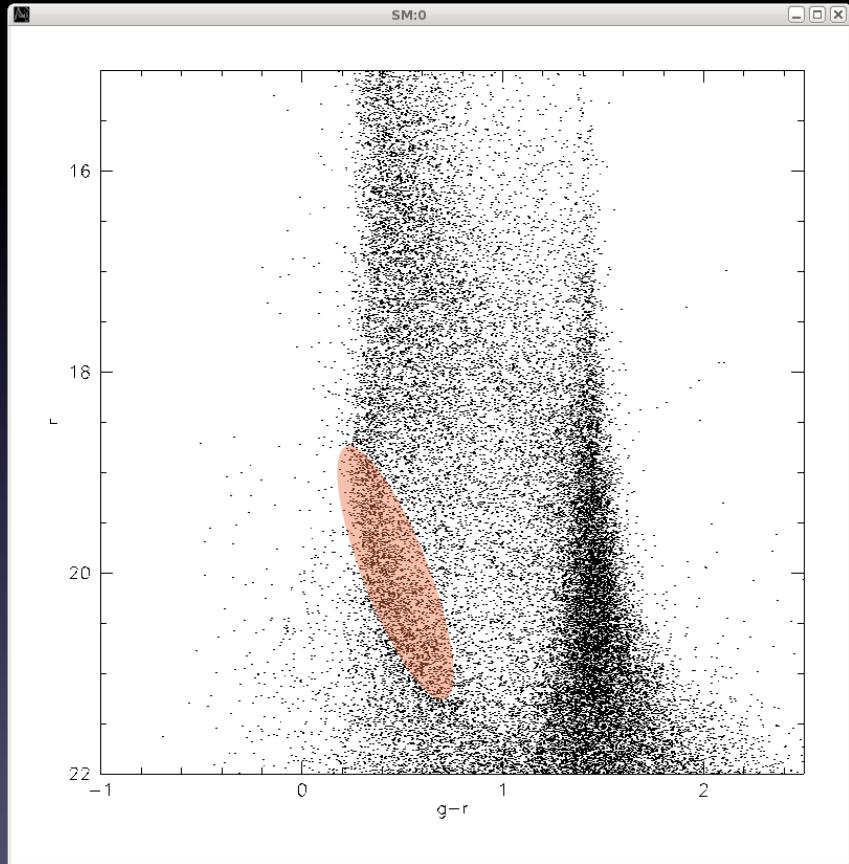
QSOs, RR Lyrae,  
WDs, errors...

# Notice Anything Different...?

Field #1



Field #2



# Monoceros Stream (Newberg et al. 2002)

## THE GHOST OF SAGITTARIUS AND LUMPS IN THE HALO OF THE MILKY WAY

HEIDI JO NEWBERG,<sup>1,2</sup> BRIAN YANNY,<sup>1,3</sup> CONNIE ROCKosi,<sup>4</sup> EVA K. GREBEL,<sup>5</sup> HANS-WALTER RIX,<sup>5</sup> JON BRINKMANN,<sup>6</sup> ISTVAN CSABAI,<sup>7</sup> GREG HENNESSY,<sup>8</sup> ROBERT B. HINDSLEY,<sup>8</sup> RODRIGO IBATA,<sup>9</sup> ZELJKO IVEZIĆ,<sup>10</sup> DON LAMB,<sup>4</sup> E. THOMAS NASH,<sup>3</sup> MICHAEL ODENKIRCHEN,<sup>5</sup> HEATHER A. RAVE,<sup>2</sup> D. P. SCHNEIDER,<sup>11</sup> J. ALLYN SMITH,<sup>12</sup> ANDREA STOLTE,<sup>5</sup> AND DONALD G. YORK<sup>4</sup>

*Received 2001 June 18; accepted 2001 December 5*

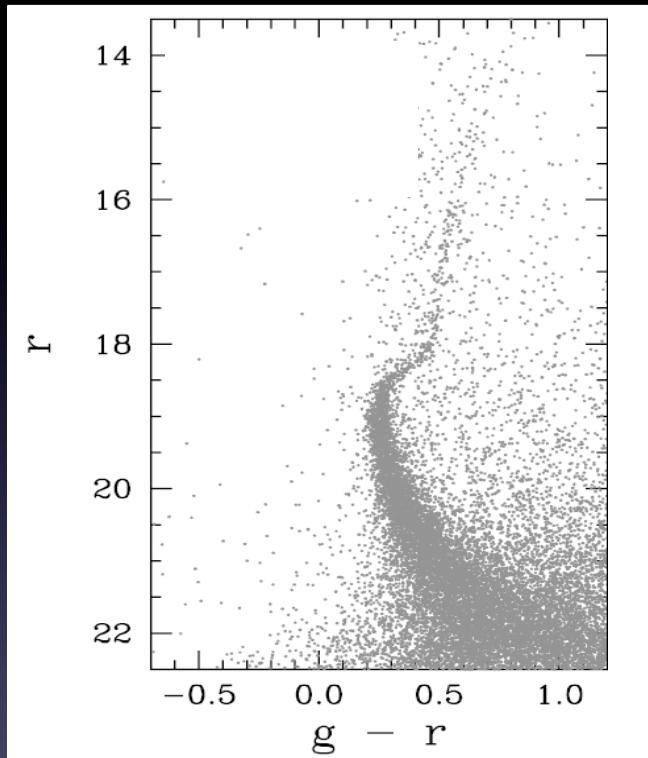
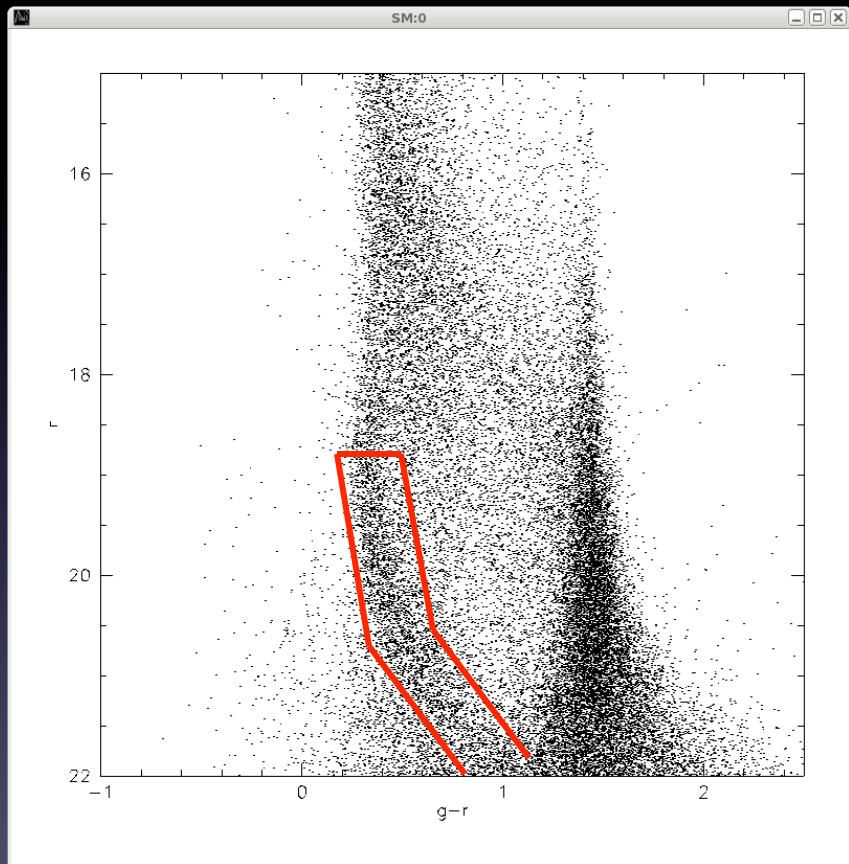
## ABSTRACT

We identify new structures in the halo of the Milky Way from positions, colors, and magnitudes of five million stars detected in the Sloan Digital Sky Survey. Most of these stars are within  $1^{\circ}26$  of the celestial equator. We present color-magnitude diagrams (CMDs) for stars in two previously discovered, tidally disrupted structures. The CMDs and turnoff colors are consistent with those of the Sagittarius dwarf galaxy, as had been predicted. In one direction, we are even able to detect a clump of red stars, similar to that of the Sagittarius dwarf, from stars spread across  $110 \text{ deg}^2$  of sky. Focusing on stars with the colors of F turnoff objects, we identify at least five additional overdensities of stars. Four of these may be pieces of the same halo structure, which would cover a region of the sky at least  $40^\circ$  in diameter, at a distance of 11 kpc from the Sun (18 kpc from the center of the Galaxy). The turnoff is significantly bluer than that of thick-disk stars, yet the stars lie closer to the Galactic plane than a power-law spheroid predicts. We suggest two models to explain this new structure. One possibility is that this new structure could be a new dwarf satellite of the Milky Way, hidden in the Galactic plane and in the process of being tidally disrupted. The other possibility is that it could be part of a disklike distribution of stars which is metal-poor, with a scale height of approximately 2 kpc and a scale length of approximately 10 kpc. The fifth overdensity, which is 20 kpc away, is some distance from the Sagittarius dwarf streamer orbit and is not associated with any known Galactic structure. We have tentatively identified a sixth overdensity in the halo. If this sixth structure is instead part of a smooth distribution of halo stars (the spheroid), then the spheroid must be very flattened, with axial ratio  $q = 0.5$ . It is likely that there are many smaller streams of stars in the Galactic halo.

*Subject headings:* Galaxy: halo — Galaxy: structure

# Learning about structures from CMDs

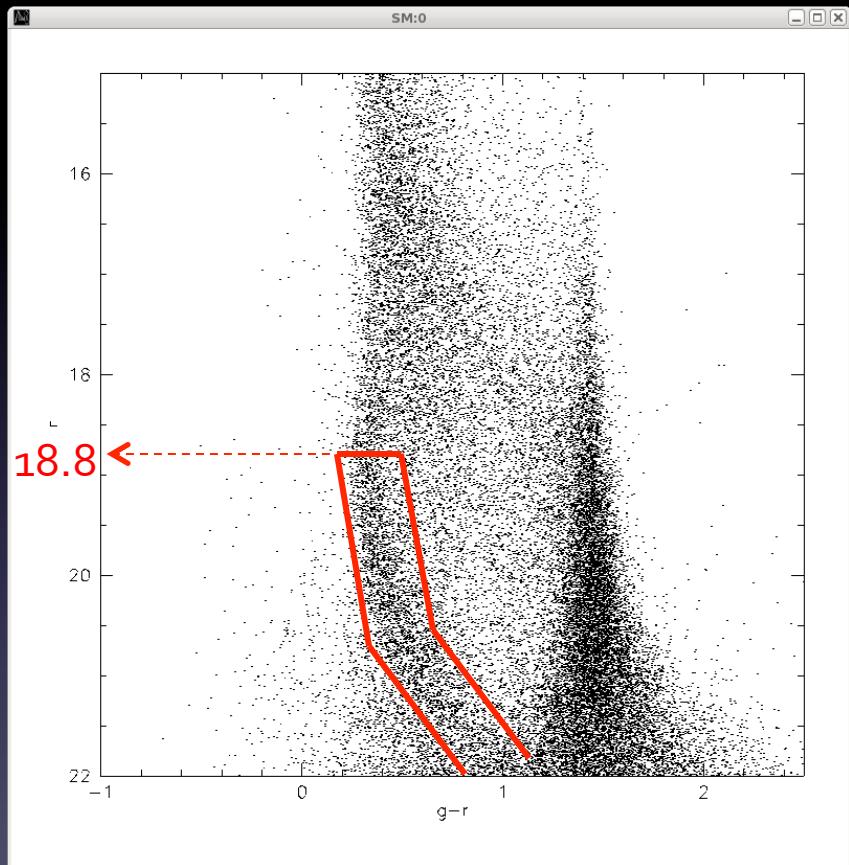
Monoceros stream CMD



⇒ Monoceros stream is a  
localized structure

# Learning about structures from CMDs

Monoceros stream CMD

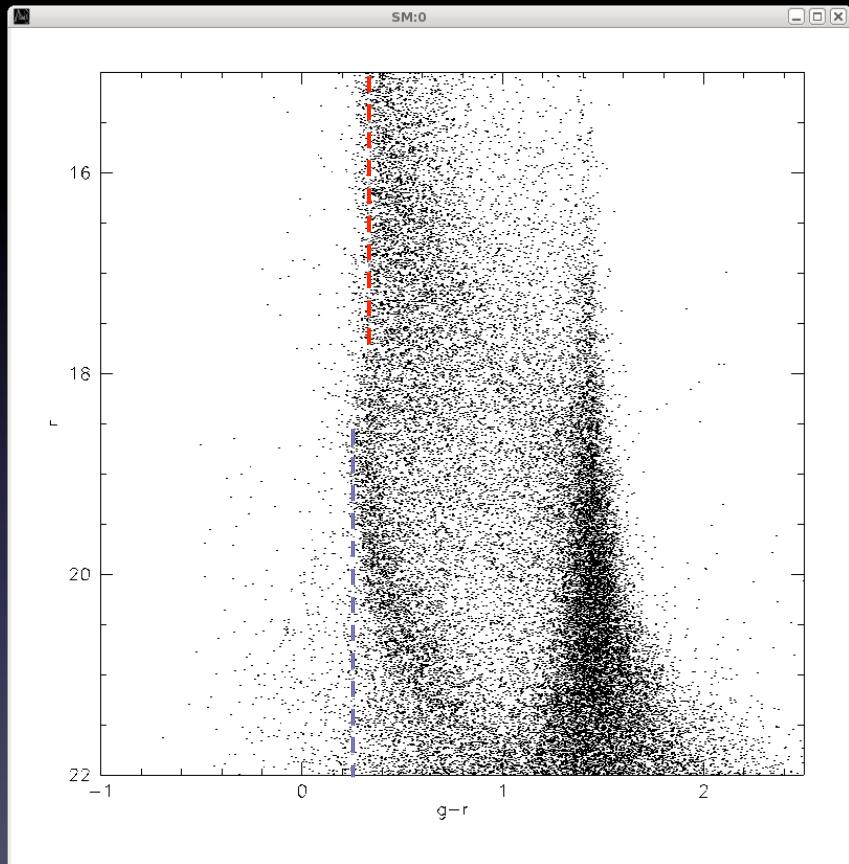


$$M_{MSTO} \approx 3.5$$
$$d = 10^{\frac{m-M}{5}+1}$$

⇒ *The distance to the Monoceros stream has to be around ~11.5kpc*

# Learning about structures from CMDs

Monoceros stream CMD



Turn-off color of the Monoceros stream is bluer than that of the disk  
⇒ Mon. stars are either more metal poor, or younger than disk stars

# Monoceros Stream (Newberg et al. 2002)

## THE GHOST OF SAGITTARIUS AND LUMPS IN THE HALO OF THE MILKY WAY

HEIDI JO NEWBERG,<sup>1,2</sup> BRIAN YANNY,<sup>1,3</sup> CONNIE ROCKosi,<sup>4</sup> EVA K. GREBEL,<sup>5</sup> HANS-WALTER RIX,<sup>5</sup> JON BRINKMANN,<sup>6</sup> ISTVAN CSABAI,<sup>7</sup> GREG HENNESSY,<sup>8</sup> ROBERT B. HINDSLEY,<sup>8</sup> RODRIGO IBATA,<sup>9</sup> ZELJKO IVEZIĆ,<sup>10</sup> DON LAMB,<sup>4</sup> E. THOMAS NASH,<sup>3</sup> MICHAEL ODENKIRCHEN,<sup>5</sup> HEATHER A. RAVE,<sup>2</sup> D. P. SCHNEIDER,<sup>11</sup> J. ALLYN SMITH,<sup>12</sup> ANDREA STOLTE,<sup>5</sup> AND DONALD G. YORK<sup>4</sup>

*Received 2001 June 18; accepted 2001 December 5*

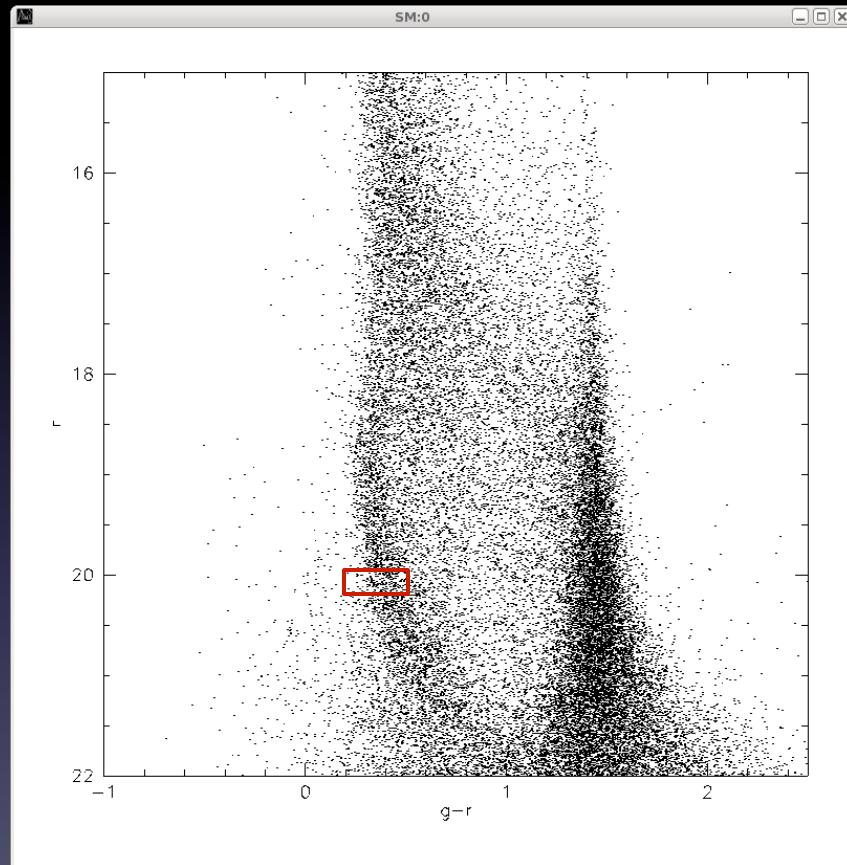
## ABSTRACT

We identify new structures in the halo of the Milky Way from positions, colors, and magnitudes of five million stars detected in the Sloan Digital Sky Survey. Most of these stars are within  $1^{\circ}26$  of the celestial equator. We present color-magnitude diagrams (CMDs) for stars in two previously discovered, tidally disrupted structures. The CMDs and turnoff colors are consistent with those of the Sagittarius dwarf galaxy, as had been predicted. In one direction, we are even able to detect a clump of red stars, similar to that of the Sagittarius dwarf, from stars spread across  $110 \text{ deg}^2$  of sky. Focusing on stars with the colors of F turnoff objects, we identify at least five additional overdensities of stars. Four of these may be pieces of the same halo structure, which would cover a region of the sky at least  $40^\circ$  in diameter, at a distance of 11 kpc from the Sun (18 kpc from the center of the Galaxy). The turnoff is significantly bluer than that of thick-disk stars, yet the stars lie closer to the Galactic plane than a power-law spheroid predicts. We suggest two models to explain this new structure. One possibility is that this new structure could be a new dwarf satellite of the Milky Way, hidden in the Galactic plane and in the process of being tidally disrupted. The other possibility is that it could be part of a disklike distribution of stars which is metal-poor, with a scale height of approximately 2 kpc and a scale length of approximately 10 kpc. The fifth overdensity, which is 20 kpc away, is some distance from the Sagittarius dwarf streamer orbit and is not associated with any known Galactic structure. We have tentatively identified a sixth overdensity in the halo. If this sixth structure is instead part of a smooth distribution of halo stars (the spheroid), then the spheroid must be very flattened, with axial ratio  $q = 0.5$ . It is likely that there are many smaller streams of stars in the Galactic halo.

*Subject headings:* Galaxy: halo — Galaxy: structure

Hmmm... How do I do this for the whole sky?

Monoceros stream CMD



CasJobs X

http://casjobs.sdss.org/CasJobs/jobdetails.aspx?id=4930248

## SDSS Query / CasJobs

Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout mjuric

### 'My Query' Details

[Resubmit Job](#)

JobID	TaskName	Context	Queue	Submitted	Started	Finished	Status
4930248	My Query	DR7	600	7/28/2010 8:04:02 AM	7/28/2010 8:04:10 AM	7/28/2010 8:09:21 AM	Finished

Executed on Rows Message

DR7Best long 94144      Query Complete

**Query**

```
select
    s.objid,
    s.l, s.b,
    s.ra, s.dec,
    s.extinction_r as Ar,
    s.dered_u as u, s.dered_g as g, s.dered_r as r, s.dered_i as i, s.dered_z as z,
    s.err_u, s.err_g, s.err_r, s.err_i, s.err_z,
    s.flags

from
    Star s

into
    mydb.northslice3

where
    b > 40
    and 0.2 < g - r and g - r < 0.4
    and 21 < r and r < 21.1
```

**Contact**  
\$Name: v3\_5\_16 \$,\$Revision: 1.23 \$, Last modified: Wednesday, May 14, 2008 at 1:52:25 AM

MyDB

http://casjobs.sdss.org/CasJobs/MyDB.aspx

# SDSS Query / CasJobs

Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout mjuric

Views Local Only

Tables

Functions

Procedures

Sort by... All selected...

Rows	kB	Name
94,144	13,768	northslice3

## Mario Juric 's MyDB

**14,232 kB of 2,000,000 kB used**

From this page you can get various information about the contents of both your and shared tables within your groups. Click the left table links to get information about a specific table, such as rows, columns or size. From the table pages you can perform various table-specific tasks, such as:

- Download a table
- Manage your group tables
- Rename a table
- Drop a table

*Sizes are approximations only.*  
*Row counts are approximations only. For exact value run a count.*  
*There's always some overhead, even empty MyDB's take up space.*  
*Group tables do not count towards your MyDB size limit.*

Contact  
\$Name: v3\_5\_16 \$,\$Revision: 1.64 \$, Last modified: Tuesday, January 27, 2009 at 3:15:10 PM

MyDB

http://casjobs.sdss.org/CasJobs/MyDB.aspx

# SDSS Query / CasJobs

mjuric

Views

Tables

Functions

Procedures

Sort by... All selected... Rows kB Name

94,144	13,768	northslice3
--------	--------	-------------

## northslice3

Contains ~94,144 rows (~13,768 kB)

Notes Sample Job Plot BPlot Download Publish Neighbors Rename Drop

### Table Schema type [size]

objid	I	b	ra	dec	Ar	u	g	r	g	i	z	Up
bigint [8]	float [8]	float [8]	float [8]	float [8]	float [8]	float [8]	float [8]	float [8]	float [8]	float [8]	float [8]	Up
Upper Limit	Up											
Lower Limit	Low											

X     X     X     X     X     X     X     X  
 Y     Y     Y     Y     Y     Y     Y     Y

### Plot Table

This produces non-interactive plots and is suitable for plotting a great numbers of objects. If you are only interested in plotting a few objects, you may want to try out the 'Quick' plot functionality, which can be reached by clicking 'Plot' on the query page after any 'Quick' query has completed.

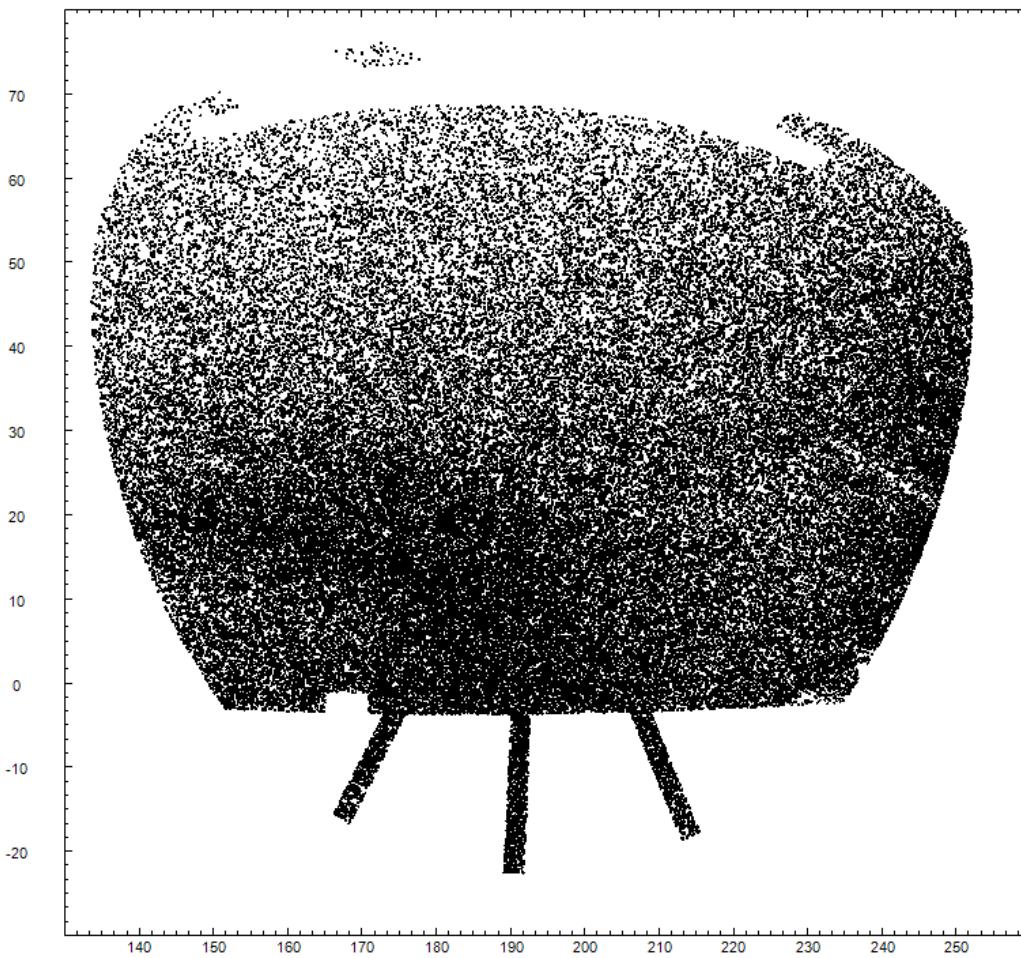
To plot a table, pick columns from above to represent values on the X and Y axis, then click 'Plot'. Only one 'X' and one 'Y' may be selected. A plotting output job with the parameters you've specified will be created and placed in queue.

You can also specify constraints on your plot by entering values into the limit fields. Any blank fields are considered unbounded. Constraints entered for columns that are not actively being plotted will still affect the plot.

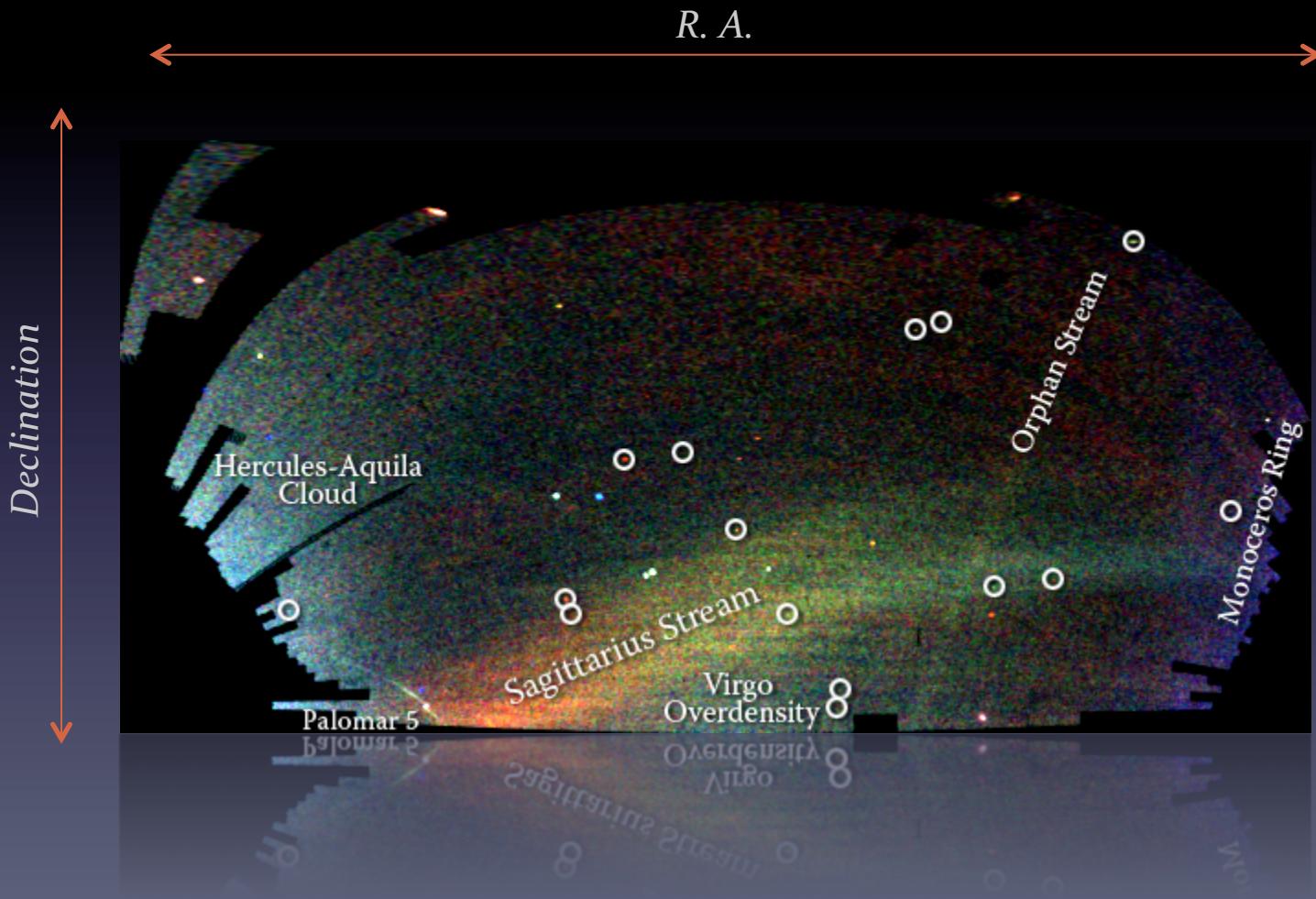
You may also manually specify your constraints by clicking 'Manually Enter Plot Constraints'.

Plot Manually Enter Plot Constraints

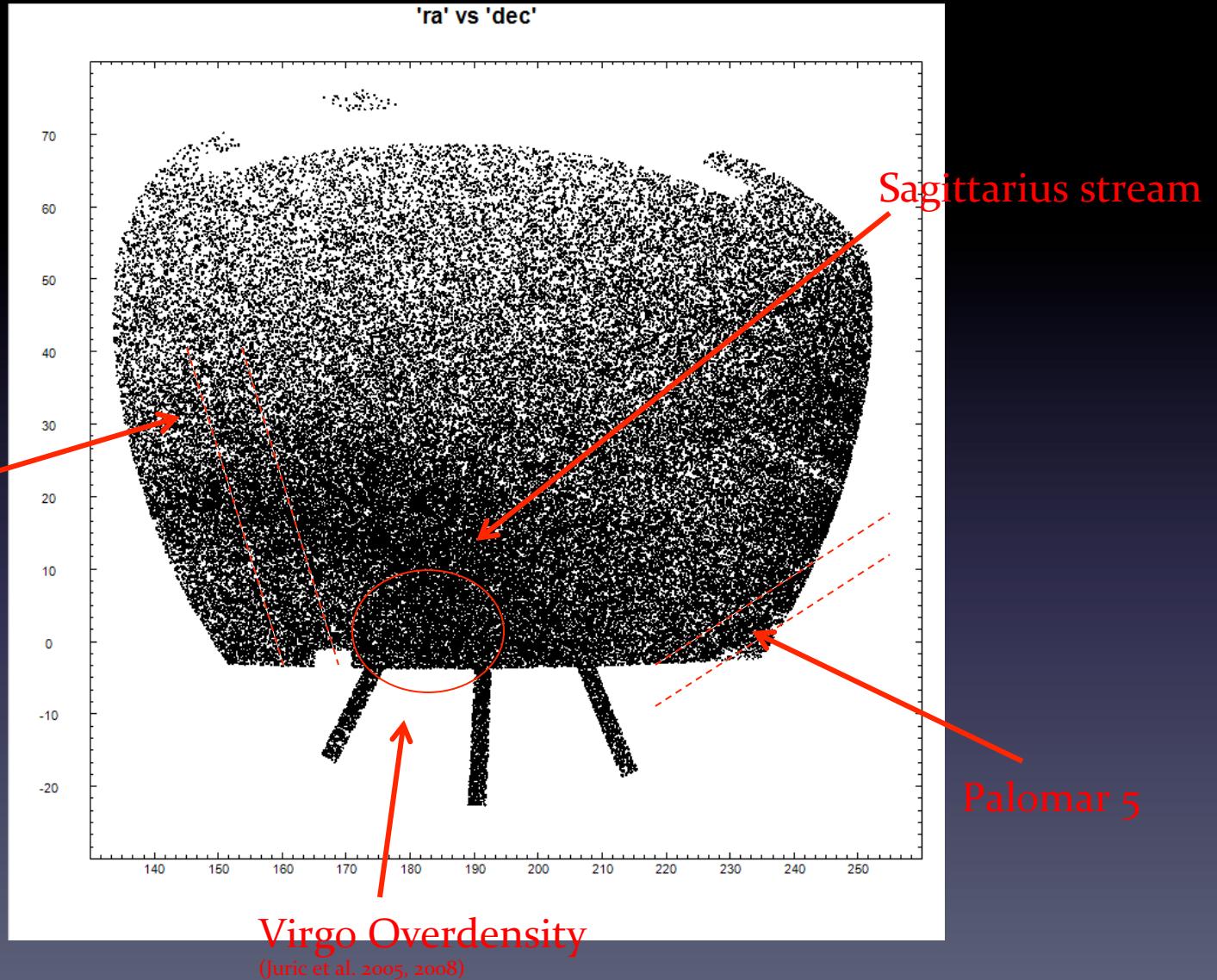
'ra' vs 'dec'



# Field of Streams



Orphan Stream  
(Belokurov et al. 2006)



Virgo Overdensity  
(Juric et al. 2005, 2008)

- Let's summarize what we just did:
  - In less than 10 minutes, we “observed” two fields 120deg apart (~4 months), to 22<sup>nd</sup> magnitude, in 5 bands
  - We “reduced” and extinction-corrected the data (~70,000 stars/field) and analyzed it using color-magnitude plots
  - We discovered a main sequence feature in one of the streams at distance around 11.5kpc, likely metal poor
  - We then proceeded to discover three more streams over the whole sky
- 10 years ago, this would have been a multi-night observing proposal for top telescopes in the world. The challenge is obtaining the data.
- Today it’s an SQL query – *the data is already there, the challenge is to ask the right question!*

# Whoa!



# Astronomy in the Age of Large Surveys

- Traditionally, astronomy was a data-starved science. Our approach to research and our analysis methods were shaped by this environment. Surveys are altering it; data is becoming abundant and of unprecedented quality.

# Sloan Digital Sky Survey

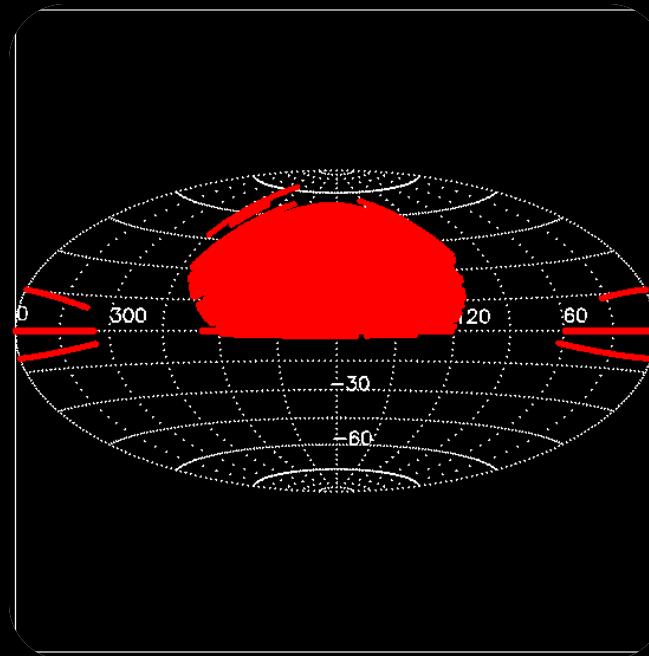
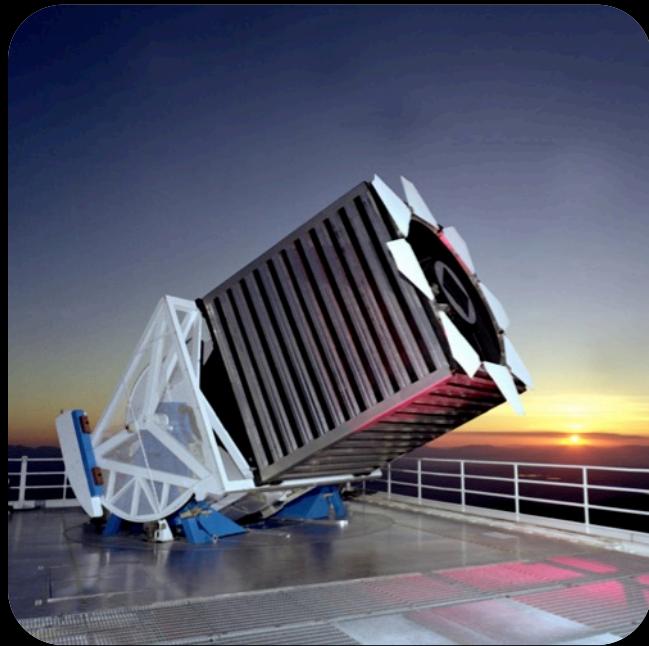
2.5m telescope

>10000 deg<sup>2</sup>

0.1" astrometry

r<22.5 flux limit

5 band, 2%, photometry for >50M stars  
>300k R=2000 stellar spectra



*10 years of ops: ~10 TB of imaging*

# Panoramic Survey Telescope and Rapid Response System

1.8m telescope

30000 deg<sup>2</sup>

50mas astrometry

r<23 flux limit

5 band, better than 1% photometry (goal)



~700 GB/night

# LSST: A Deep, Wide, Fast, Optical Sky Survey



8.4m telescope

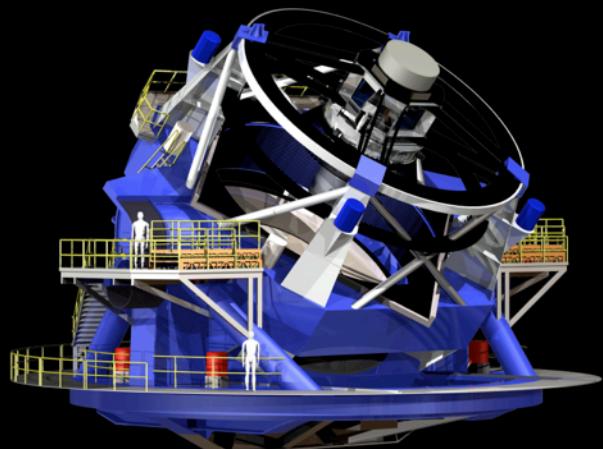
18000+ deg<sup>2</sup>

10mas astrom.

r<24.5 (<27.5@10yr)

ugrizy

0.5-1% photometry

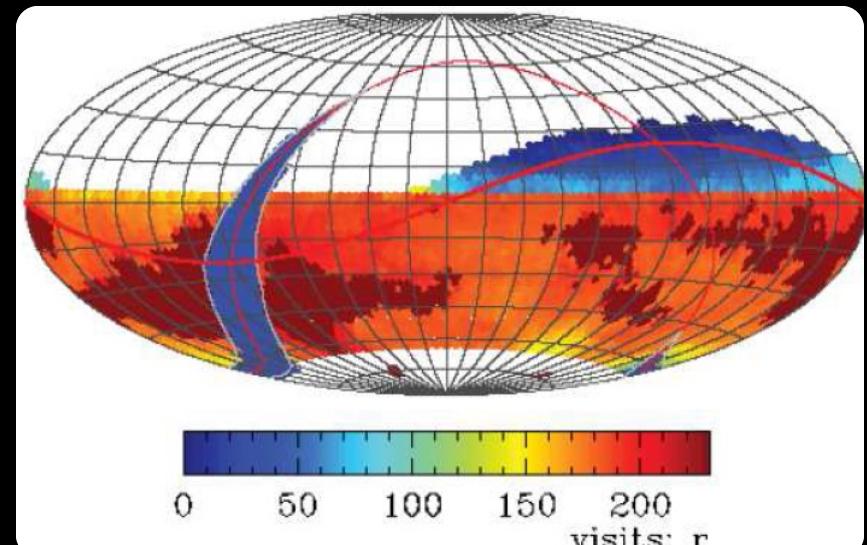


3.2Gpix camera

30sec exp/4sec rd

15TB/night

37 B objects



Imaging the visible sky, once every 3 days, for 10 years (825 revisits)

# Low-Frequency Array for Radio Astronomy (LOFAR)



Estimated Data Volume

Data Type	Data Rate [MB/s]	Data Rate [GB/h]
Burst Mode	100	360
Monitoring Mode	0.03	0.1
Spectrometer Mode	0.34	1.2

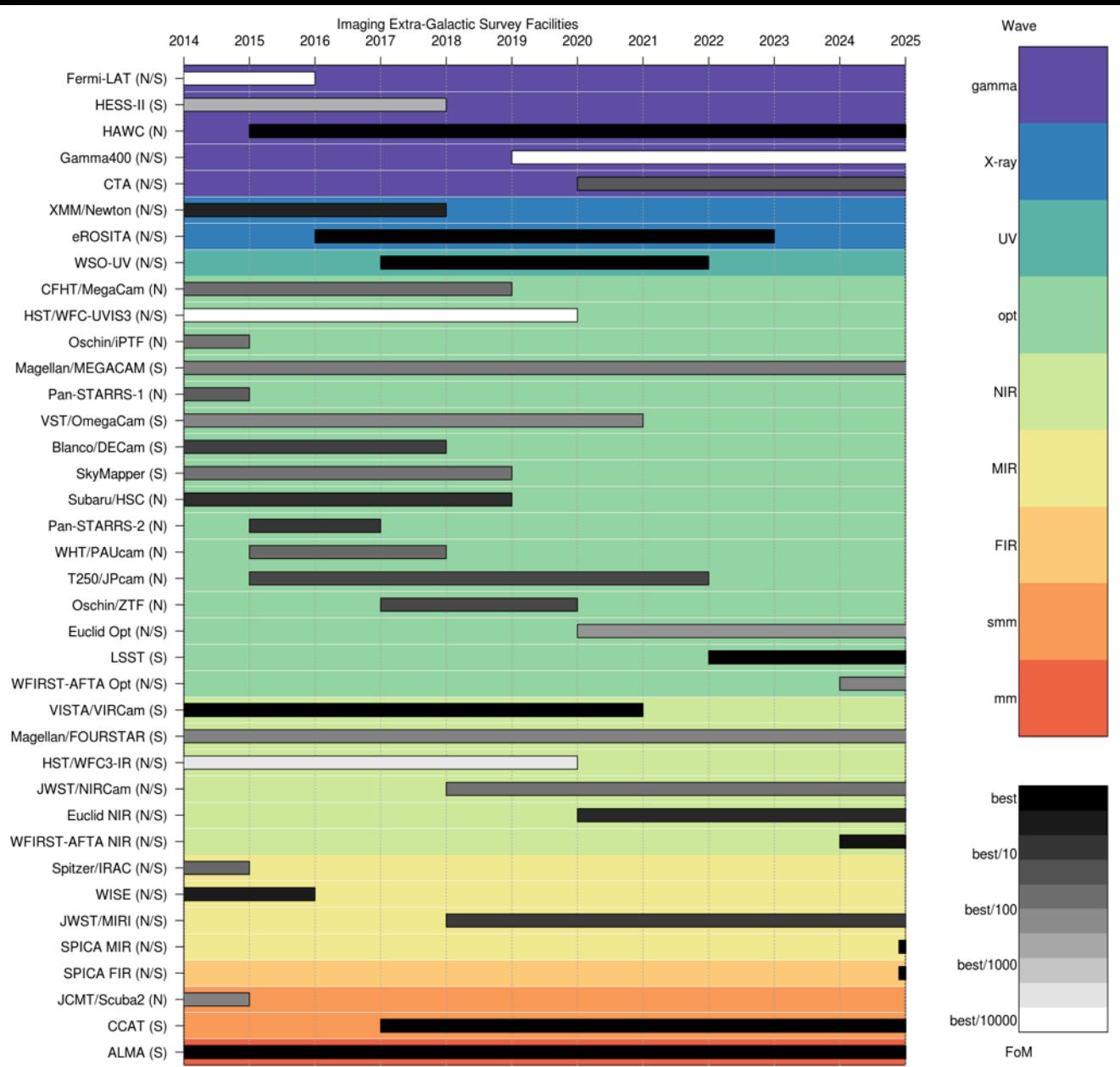


# Square Kilometer Array



SPDO / Swinburne Astronomy Productions

10 PB / hour



# Astronomy in the Age of Large Surveys

- Traditionally, astronomy was a data-starved science. Our approach to research and our analysis methods were shaped by this environment. Surveys are altering it; data is becoming abundant and of unprecedented quality.
- For optical, LSST will cap this transformation: it will deliver the positions, magnitudes and variability information for virtually *everything* in the southern sky to 24<sup>th</sup>-27<sup>th</sup> magnitude, with an order of magnitude better controlled systematics than current surveys.

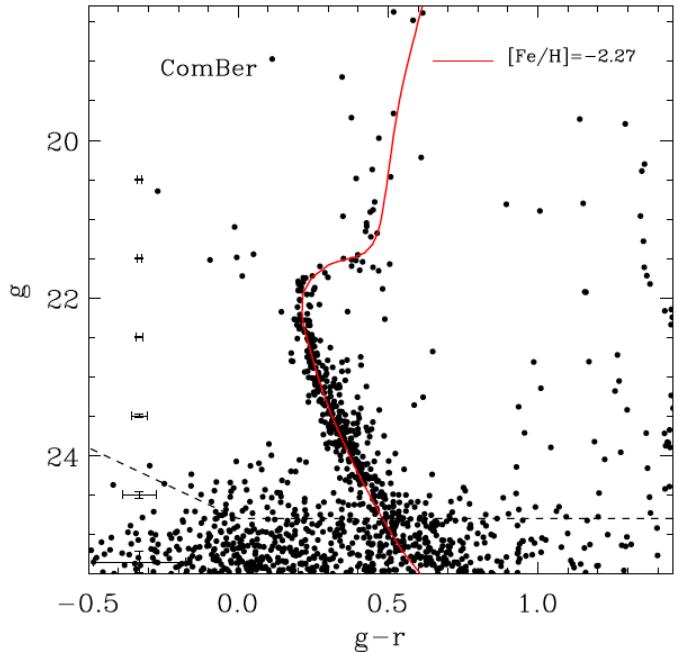
# A Quantitative Difference

Final Image Collection – All DRs	515 PB	All Data Releases Includes Virtual Data (475 PB)
Final Image Collection – DR11	114 PB	Data Release 11 (Year 10) Includes DR11 Virtual Data (88 PB) and all raw images from all years
Final Database	16 PB	Data Release 11 (Year 10) Includes Data, Indexes, and DB Swap
Final Disk Storage	375 PB	Archive and Base Sites
Final Tape Storage	121 PB	Single Site, Single Copy Only
Peak Number of Nodes	1750	Archive and Base Sites Compute and Database Nodes
Number of Alerts Generated	28 billion	Life of survey

- **Virtual Data is data that is dynamically recreated on-demand from provenance information**

# A Qualitative Difference

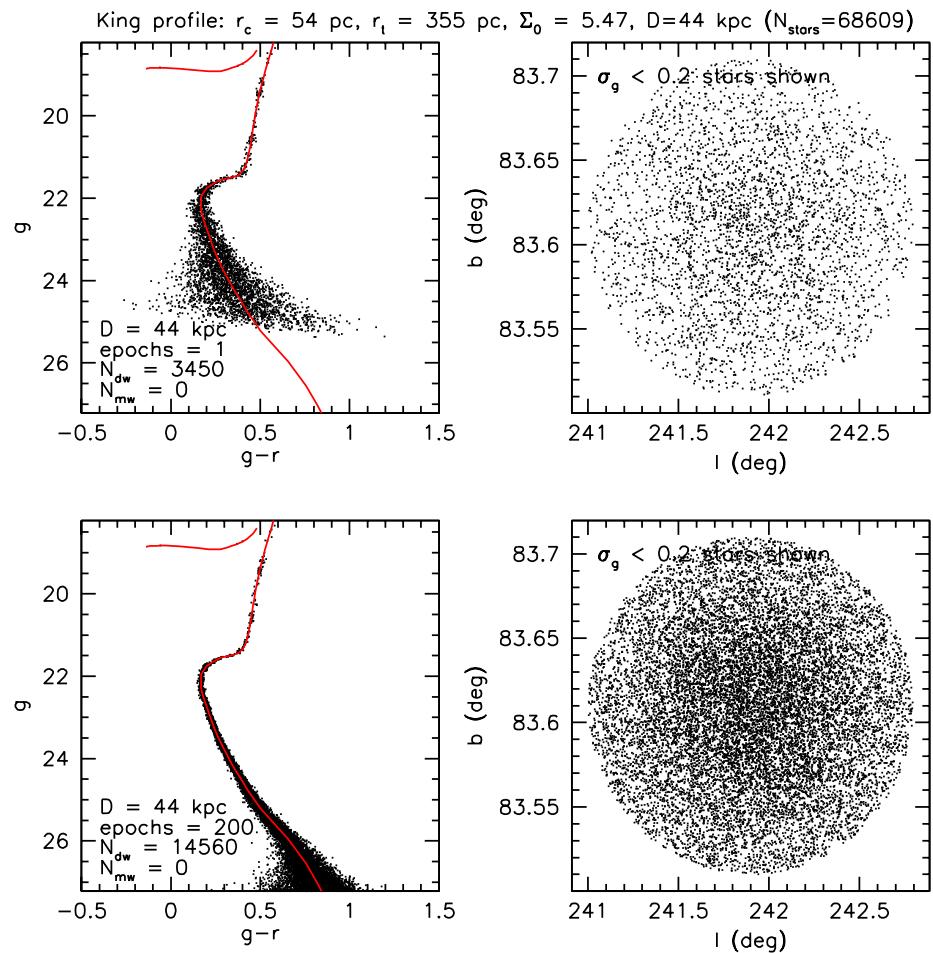
9 hours on CFHT/MegaCam (Munoz et al. 2010)



**Figure 3.** CMD for the inner region ( $r < 6'$ ) of ComBer. The dashed lines mark the 90% completeness level after  $\chi$  and sharp cuts have been applied to remove non-stellar objects. As it can be seen, our CFHT photometry reaches at least 3 mag below the main sequence turn-off of ComBer. We have complemented our photometry with SDSS data for  $g > 20$ . The error bars to the left were determined from the artificial star tests and represent the standard deviation of a Gaussian function fitted to the error distribution as a function of magnitude. A theoretical isochrone for a 13 Gyr old,  $[Fe/H] = -2.27$  population is shown with a solid red line (from Girardi et al. 2004).

(A color version of this figure is available in the online journal.)

LSST, single-epoch and full-survey (simulation)



# Astronomy in the Age of Large Surveys

- Traditionally, astronomy was a data-starved science. Our approach to research and our analysis methods were shaped by this environment. Surveys are altering it; data is becoming abundant and of unprecedented quality.
- For optical, LSST will cap this transformation: it will deliver the positions, magnitudes and variability information for virtually *everything* in the southern sky to 24<sup>th</sup>-27<sup>th</sup> magnitude, with an order of magnitude better controlled systematics than current surveys.
- **We're entering the age of abundance of high quality data. Success in research will depend on the ability to analyze and mine knowledge from that data.**

# Next time ...

- Software Tools
  - Python
  - IPython
  - Git and github
  - Basic numerical and analysis libraries