Announcements

- Welcome to Intro to Applied Statistics!
- Things to do:
  - Make sure you have a copy of the book
  - Ensure you can sign into Campuswire and Gradescope
  - Read over my full syllabus on our website here
  - Read over Ch 1.1 – 1.4 by Monday
- Copies of slides will be available on the webpage after each class
- We don't have class on Friday!

What even is this class?

Grab a few neighbors and discuss and list out some of your expectations for this class. We'll jointly make a list on the board in a few minutes.

- What topics are you expecting to cover?
- What skills are you hoping to learn?
- What activities do you see being central to this course?

Statistics is all about drawing meaningful conclusions from collections of observations.

My View

Statistics is all about drawing meaningful conclusions from collections of observations.

- Sometimes those conclusions might be easy: *People averaged a B on this test. They are probably mostly learning what they should be.*

My View

Statistics is all about drawing meaningful conclusions from collections of observations.

- Sometimes those conclusions might be easy: *People averaged a B on this test. They are probably mostly learning what they should be.*
- Sometimes they might not be so easy: *Can someone really tell the difference if milk is added to a cup before or after it contains tea?*

A Lady's Taste

- In the 1930's Muriel Bristol claimed to be able to tell, by taste, whether the tea or the milk was added to the cup first.
- How would you test this?
    - Was given 8 randomly ordered cups, 4 prepared tea first and 4 prepared milk first.
    - Muriel had to choose 4 prepared by a single method (all the tea-firsts or all the milk-firsts)
- How many would she have to get correct before you'd be convinced she could tell the difference?
    - A) 1
   - B) 2
    - C) 3
    - D) 4

A Lady's Taste

- In the 1930's Muriel Bristol claimed to be able to tell, by taste, whether the tea or the milk was added to the cup first.
- How would you test this?
    - Was given 8 randomly ordered cups, 4 prepared tea first and 4 prepared milk first.
    - Muriel had to choose 4 prepared by a single method (all the tea-firsts or all the milk-firsts)
- How many would she have to get correct before <u>you'd</u> be convinced she could tell the difference?
    - A) 1
    - B) 2
    - C) 3
    - D) 4
- Determining (and justifying) that number will be a large part of this class

My Goals

What I envision this class revolving around:

- How to plan observations to provide for as robust a statistical analysis as possible
  - Choosing samples
  - Minimizing external factors
- Interpret summary statistics and confidence intervals to evaluate how robust a conclusion you can really draw from the data
- Matching the right type of statistical test with the right questions and observations
- How to visualize and communicate statistical results
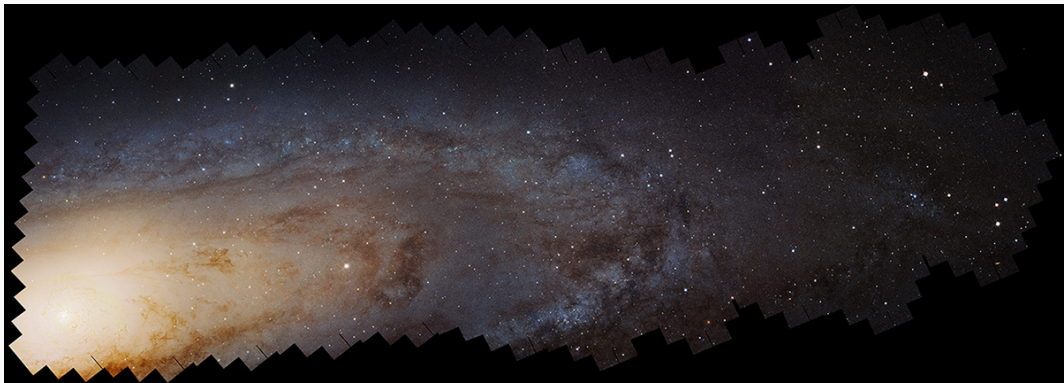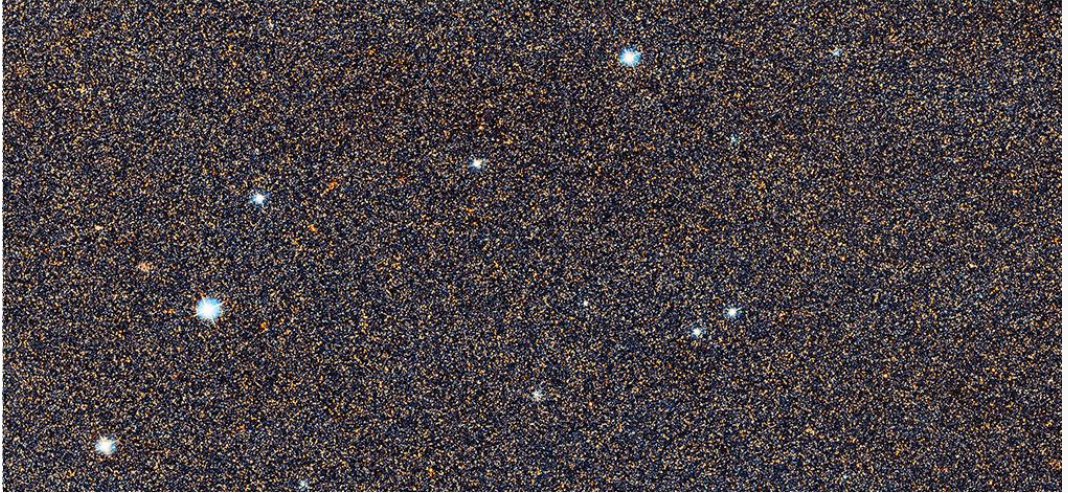
This is going to take a while. . .

- There are about 110 million distinct stars in this image
- Summary statistics and HR diagrams can tell us a wealth about exactly how the Andromeda Galaxy formed, but require working with 100 million row spreadsheets
- Characteristic as much of modern statistics deals with massive datasets
- Need new methods and tools to do so

My Goals (continued)

This class will also revolve around:

- Gaining comfort, familiarity, and some proficiency in the modern statistical software languange R.
    - How to import and view data
    - How to calculate the desired test statistics you wish to know
    - How to visualize large datasets and statistics

How we get there: Syllabus

My Vitals

Name: Jed Rembold

Office: Collins 311 (it is shared)

Hours: MWTh 3-5 (but I'm here a lot)

Email: jjrembold@willamette.edu

Phone: 503-370-6860

Grading

Participation	6%
Homework	15%
In-class Labs	15%
Midterms (2)	24%
Final Project	20%
Final Exam	20%

Participation

- Graded through participation in class polling questions
- Usually will be around 1-3 polls per lecture
- Answering at all (right or wrong) gets you full participation points for the day
- Answering correctly gets you a bit of extra credit
- Polling site is: http://rembold-class.ddns.net
    - If you want your credit, make sure the name you enter in is close enough to your roster name for my scripts to figure it out!

Homework

- Homework will be due weekly on Monday nights
- Each assignment will be a collection of a few practice problems and then two problems that will be graded.
    - Practice problems will have their solutions available in the back of the book or provided.
    - Graded problems should be scanned/imaged to pdf and submitted through Gradescope.
- All work should be shown and conclusions clearly explained
- You have 10 cumulative late days to use throughout the semester before any late work is only worth 50%

In-class Labs

- You'll be working through a guided lab dealing extensively with doing things in R.
- Bring your laptops!
- Should be able to mostly finish in the class hour
- Exercises and "On Your Own" portions should be typed up in Rmarkdown and exported to pdf for uploading to Gradescope
- Most labs on Mondays or Wednesdays, but pay attention to the schedule!

Tests

- You have two midterms:
  - Friday, March 8
  - Friday, April 19
- Final
  - May 10
  - Comprehensive
- Tests during class hours and thus limited in length
- You'll get an index card to write whatever you might find useful and bring with you to the test

- Should have already gotten the invitation
- Classroom forum to enable better communication and asking of questions
    - <u>Please</u> post here instead of emailing me questions
    - Everyone can benefit from your question and answer that way
- I will also use for making general announcements or occassional polling, so don't be a stranger!

Understanding Data

A Data Table

Data collected on students in a statistics class on a variety of variables:

variable
↓

Stu.	gender	intro_extra	⋯	dread	
1	male	extravert	⋯	3	
2	female	extravert	⋯	2	
3	female	introvert	⋯	4	←
4	female	extravert	⋯	2	observation
⋮	⋮	⋮	⋮	⋮	
27	male	extravert	⋯	3	
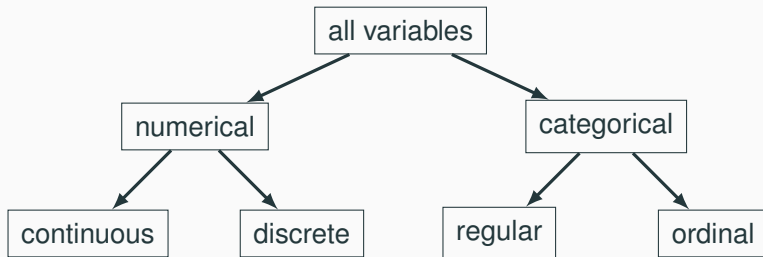
Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender:

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>
- sleep:

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>
- sleep: <u>numerical, continuous</u>

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- `gender`: <u>categorical</u>
- `sleep`: <u>numerical, continuous</u>
- `bedtime`:

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>
- sleep: <u>numerical, continuous</u>
- bedtime: <u>categorical, ordinal</u>

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>
- sleep: <u>numerical, continuous</u>
- bedtime: <u>categorical, ordinal</u>
- countries:

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>
- sleep: <u>numerical, continuous</u>
- bedtime: <u>categorical, ordinal</u>
- countries: <u>numerical, discrete</u>

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- `gender`: <u>categorical</u>
- `sleep`: <u>numerical, continuous</u>
- `bedtime`: <u>categorical, ordinal</u>
- `countries`: <u>numerical, discrete</u>
- `dread`:

Example Variable Types

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender: <u>categorical</u>
- sleep: <u>numerical, continuous</u>
- bedtime: <u>categorical, ordinal</u>
- countries: <u>numerical, discrete</u>
- dread: <u>categorical, ordinal - could also be used as numerical</u>

Going Forward

- Remember no class on Friday!
- Have read over Chapter 1.1 – 1.4 by Monday
- Make sure to have something to respond to polling questions
- See about installing R and Rstudio installed on your personal laptop
  - A guide will be posted on the web-page
  - Come see me if you are having any issues
  - Definitely want it up and running by next Wednesday
  - Let me know if you can't bring a laptop, as we can work around that to a limited degree