

Week 1: Welcome to statistics and data

1. Intro to statistical thinking

Stat 140 - 04

Mount Holyoke College

Dr. Shan Shan

Slides posted at <http://sshanshans.github.io/stat140>

1. A huge thank you
2. Today: What is statistics?
3. Main ideas
 1. What are data?
 2. Three components of statistics
 3. I notice, I wonder
 4. About this course
4. Summary

What ultimately matters in this course is not where you end up relative to your classmates, but where you end up relative to yourself when you began.

1. A huge thank you
2. Today: What is statistics?
3. Main ideas
 1. What are data?
 2. Three components of statistics
 3. I notice, I wonder
 4. About this course
4. Summary



1. A huge thank you
2. Today: What is statistics?
3. Main ideas
 1. What are data?
 2. Three components of statistics
 3. I notice, I wonder
 4. About this course
4. Summary

1. A huge thank you
2. Today: What is statistics?
3. **Main ideas**
 1. What are data?
 2. Three components of statistics
 3. I notice, I wonder
 4. About this course
4. Summary

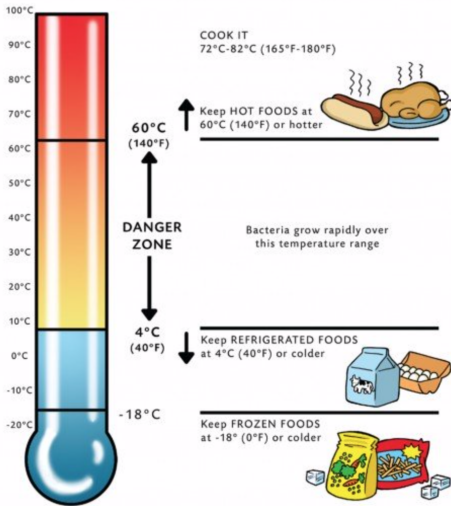
140

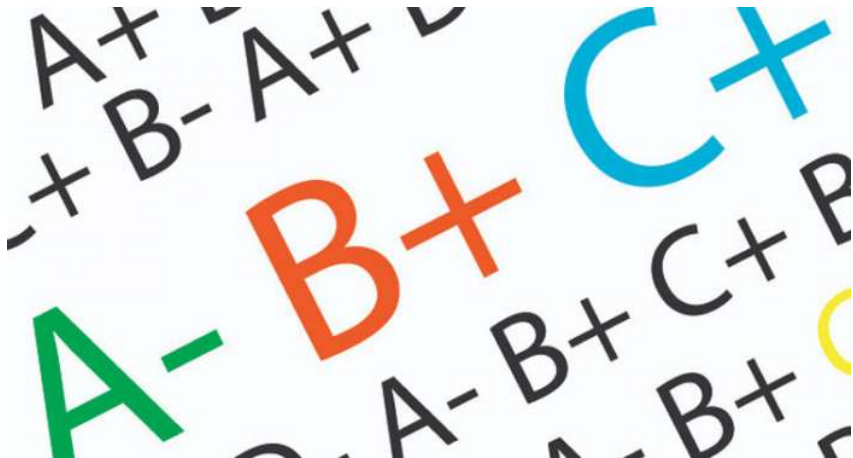


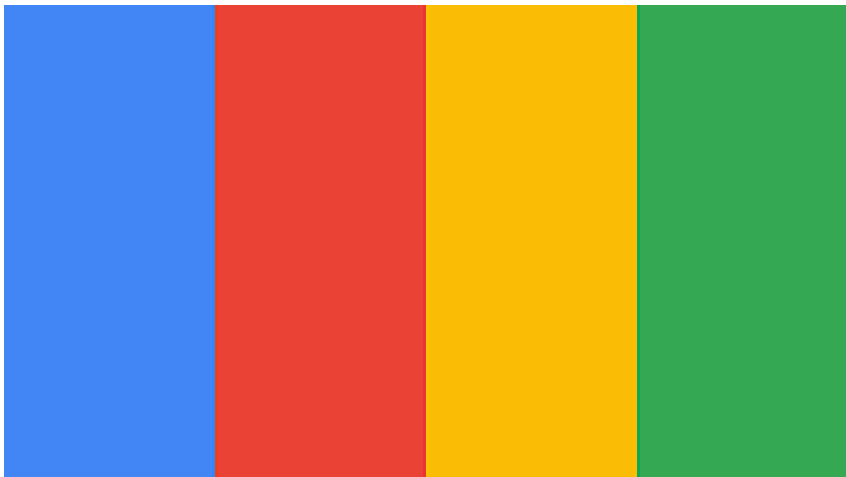
**STAT 140: Introduction to the
Ideas and Applications of
Statistics**



Data are values together with their context







Largamente e cantato





Data are recorded values, whether numbers or labels, together with their context.



1. A huge thank you
2. Today: What is statistics?
- 3. Main ideas**
 1. What are data?
 - 2. Three components of statistics**
 3. I notice, I wonder
 4. About this course
4. Summary

Data analysis

THE GATHERING, DISPLAY, AND SUMMARY OF DATA;

Probability

THE LAWS OF CHANCE, IN AND OUT OF THE CASINO;

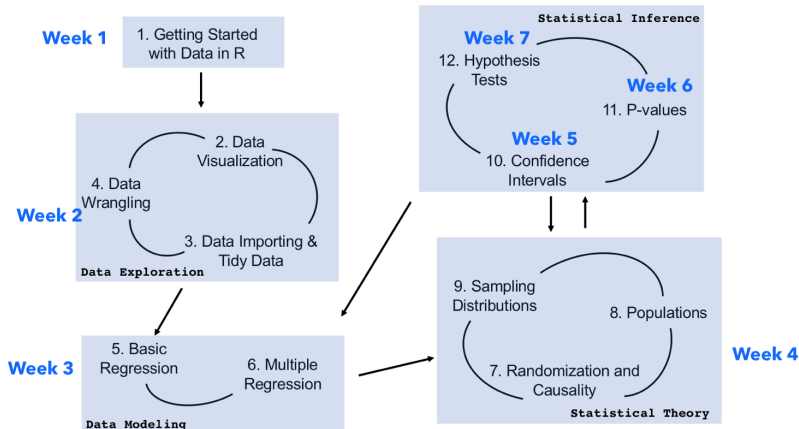
Statistical inference

THE SCIENCE OF DRAWING STATISTICAL CONCLUSIONS FROM SPECIFIC DATA, USING A KNOWLEDGE OF PROBABILITY.

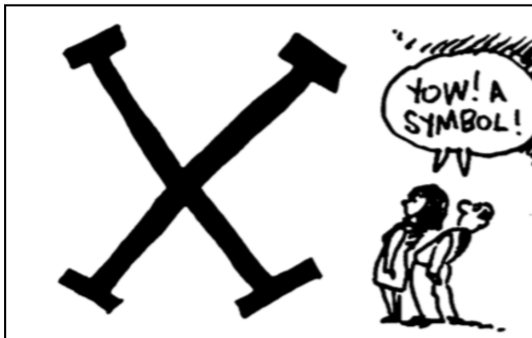


Picture from *The cartoon guide to statistics*

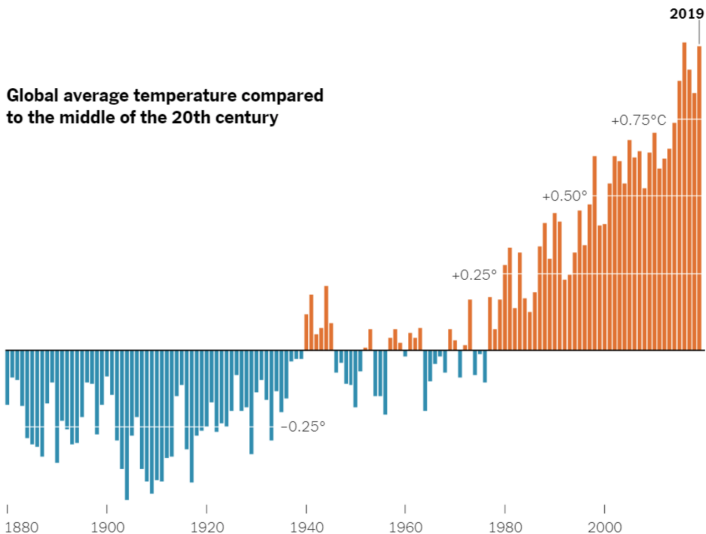
We will look at all three as applied to a wide variety of situations where statistics plays a crucial role in the modern world.



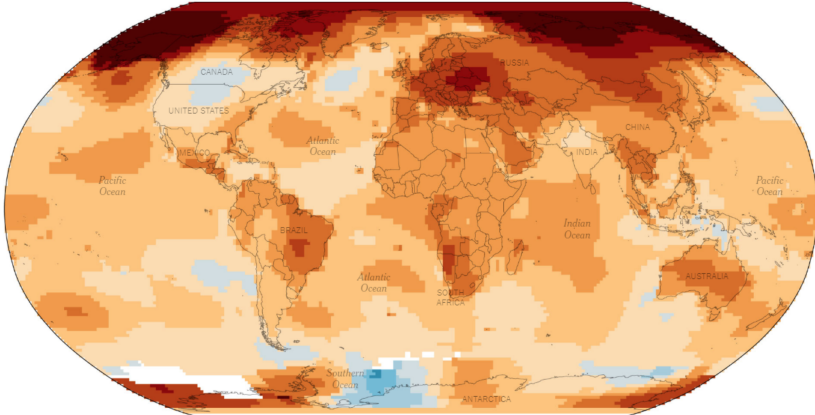
You will be asked to work on a dataset of your own choosing, and communicate your statistical findings (exploratory data analysis, probability model, inference) in a presentation and a final paper.



1. A huge thank you
2. Today: What is statistics?
- 3. Main ideas**
 1. What are data?
 2. Three components of statistics
 - 3. I notice, I wonder**
 4. About this course
4. Summary



Degrees cooler or warmer in 2019 compared to the middle of the 20th century



Tutorial exercise: 25 minutes

1. Introduce yourselves in your breakout group.
2. Identify in your group who will be
 - post writer
 - discussion leader
 - spokesperson
3. Work on the questions in the "I wonder, I notice" activity.
The post writer: post your answer in the follow up discussion. Each group only needs to submit once.

More info: <https://climate.nasa.gov>

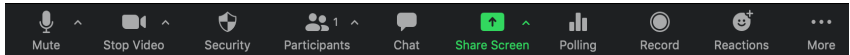
Rename yourself on Zoom with the following format:

[Time difference]_[Your preferred name] (Your pronoun)

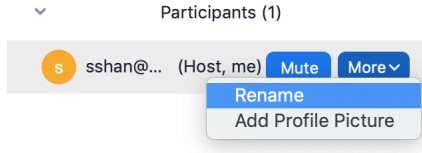
E.g.,

- ▶ 0_Shan Shan (she/her)
- ▶ -3_Shan Shan (she/her) (if I'm currently in California)
- ▶ +12_Shan Shan (she/her) (if I'm currently in Beijing)

Step 1: Find 'Participants' tab



Step 2: Find yourself → More → Rename



1. A huge thank you
2. Today: What is statistics?
- 3. Main ideas**
 1. What are data?
 2. Three components of statistics
 3. I notice, I wonder
 - 4. About this course**
4. Summary

- ▶ Before class
 - read texts or watch videos
- ▶ Class meetings
 - interactive
 - some lectures, lots of learn-by-doing
 - work on examples together (not collected)
- ▶ Every Wednesday 6PM
 - new problem set out, old one due
- ▶ Every Friday 6PM
 - new project out, old one due
 - first three (MP01-03) are mini-group projects
 - last four (FP01-04) is your final project

More information: <https://sshanshans.github.io>

All communication will take place via Piazza:

1. All official announcements and communication will happen over Piazza.
2. Any questions regarding course content and course organization should be posted on Piazza. You are strongly encouraged to answer other students' questions when you know the answer.
3. If there are private matters specific to you (e.g. special accommodations, requesting alternative arrangements etc.), please create a private post on Piazza. (Let me show you how.)
4. For longer discussions (e.g., help on R project), please come to office hours. (Let me show you where to find my office hours).

- ▶ I will respond to Piazza posts sent during the week within 24h. I will respond to Piazza posts sent during the weekend at my own discretion.
- ▶ Be precise with your descriptions:
 - Good: "I am getting the following error and I'm not sure how to resolve it - Error: could not find function "ggplot""
 - Bad: "R giving errors, help me! Aaaaarrrrrh!"
- ▶ You should come to office hours having done your own legwork and ruled out basic logical errors. Tell me what your goals are, what you think might be the problem, and what advice you need to move forward.
- ▶ Be prepared to send me your Rmarkdown document and any dataset you used.

1. I'm well aware that a huge volume of code is available on the web to solve any number of problems.
2. Unless I explicitly tell you not to use something, you can use any online resources (e.g. StackOverflow, RStudio Community) but you must explicitly cite where you obtained any code you directly use (or use as inspiration).
3. You are welcome to discuss the problems together and ask for advice, but you may not send or make use of code from another team.
4. Do not simply copy other people's project (online or your classmates). This is cheating, and violates the MHC honor code.

- ▶ Check the daily schedule for course materials and to-dos
- ▶ Check the office hour calendar for a quick overview of due dates.
- ▶ You can ask DataCamp to get a hint or show complete answer.

Instructions 100 XP

- Do you still remember what you have learned in the first chapter? Assign the value "60!" to the variable `vegas`. Remember: R is case sensitive!

Take Hint (-30 XP)

Instructions 70 XP

- Do you still remember what you have learned in the first chapter? Assign the value "60!" to the variable `vegas`. Remember: R is case sensitive!

Show Answer (-70 XP)

Hint

Just type the following line in the editor:

```
vegas <- "60!"
```

- ▶ Start "HW01" early and remember all assignments are due on Gradescope.

1. A huge thank you
2. Today: What is statistics?
3. Main ideas
 1. What are data?
 2. Three components of statistics
 3. I notice, I wonder
 4. About this course
4. Summary

1. What are data?
2. Three components of statistics
3. I notice, I wonder
4. About this course