IST772
Introductory Class (Week 1)

Pre-class activity: Do you have R and R-Studio installed on your computer? If not, please install them now. You will need them today!

Accommodations for COVID-19

- Stay Safe Pledge
  - “In this extraordinary time of a pandemic, we are responsible for ourselves and each other. We will work together. We will rise to the occasion. We will take necessary steps to keep each other safe. And we will not put ourselves or others at risk.”
- Wear your mask over your mouth and nose at all times (Alert level Blue or Red)
  - No eating or drinking in class

Introductions

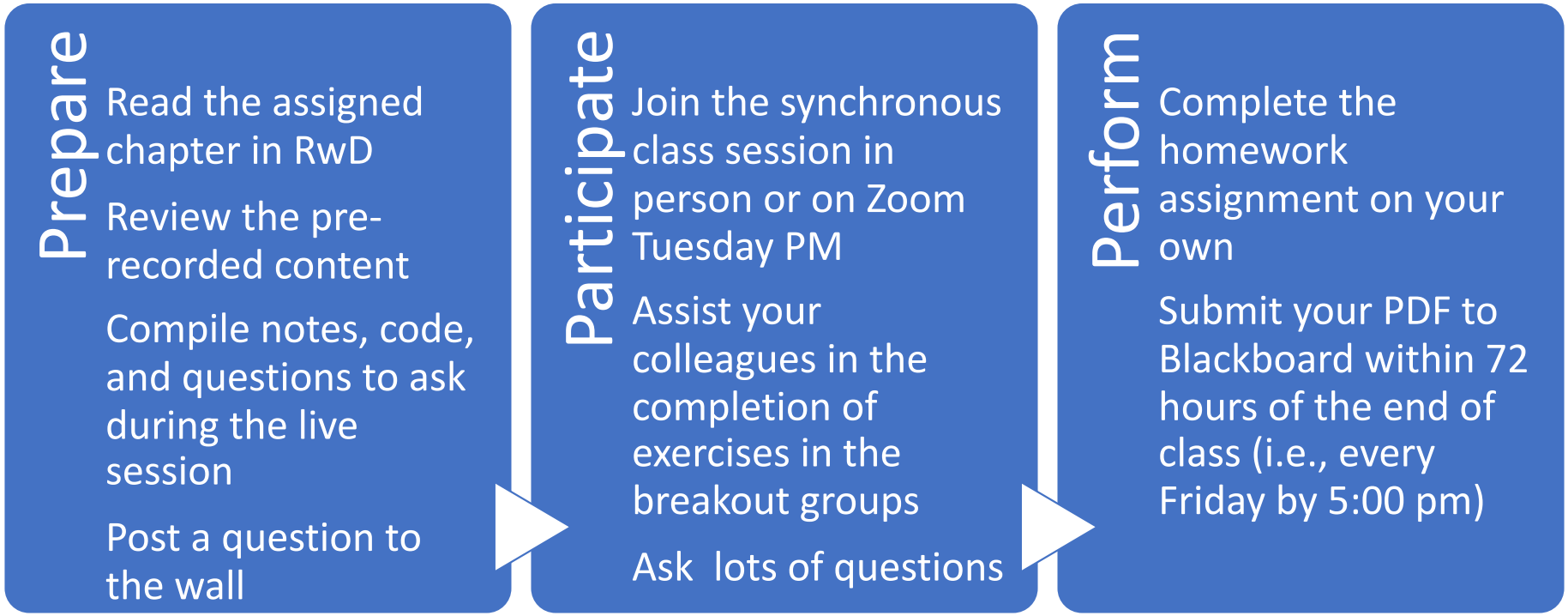
- Me (Kevin Crowston, https://crowston.syr.edu/): iSchool professor for 26 years!!! And currently Associate Dean for Research

- You: Say your name, where is home, and one other thing about yourself, for example:
  - If you have a pet, like a dog or a cat
  - If you have an interesting summer activity or place you like to visit
  - If you are a musician or artist, what you play or do

Syllabus Overview: Course Learning Goals

- Demonstrate knowledge of contemporary inferential statistical concepts from the perspective of two contemporary philosophies (frequentist and Bayesian).
- Practice effective data science analytics, including data preparation, screening, cleaning, and assumption testing. Communicate statistical results to others using language that accurately describes uncertainty
- Demonstrate competence and/or mastery of the skills needed for use of a popular statistics and data management platform to conduct sound and reproducible analyses .

Syllabus Overview:
Weekly workflow of a “flipped” class



Grading

- 11 homeworks, 5 points each, dropping lowest = 50 points
  - Homeworks are given as Rstudio notebooks and are submitted as PDFs
- 5 practice tests, 2 points each = 10 points (weeks 4, 5, 8, 9, and 10)
- 1 midterm based on the first 5 weeks of the course = 15 points
  - In class 7
- 1 final exam comprising an analysis of real data sets = 25 points
  - Done during final period 13–17 December (so don't plan to leave Syracuse before the end of the semester)

Standard HW header is in each problem set notebook

Standard Header for Homework

---

title: "IST772 Problem Set 1"

author: "Your name here"

Attribution statement: (choose only one and complete as necessary)

1. I did this homework by myself, with help from the book and the professor.

2. I did this homework with help from the book and the professor and these Internet sources:

3. I did this homework with help from <Name of another student> but did not cut and paste any code.

Think and share

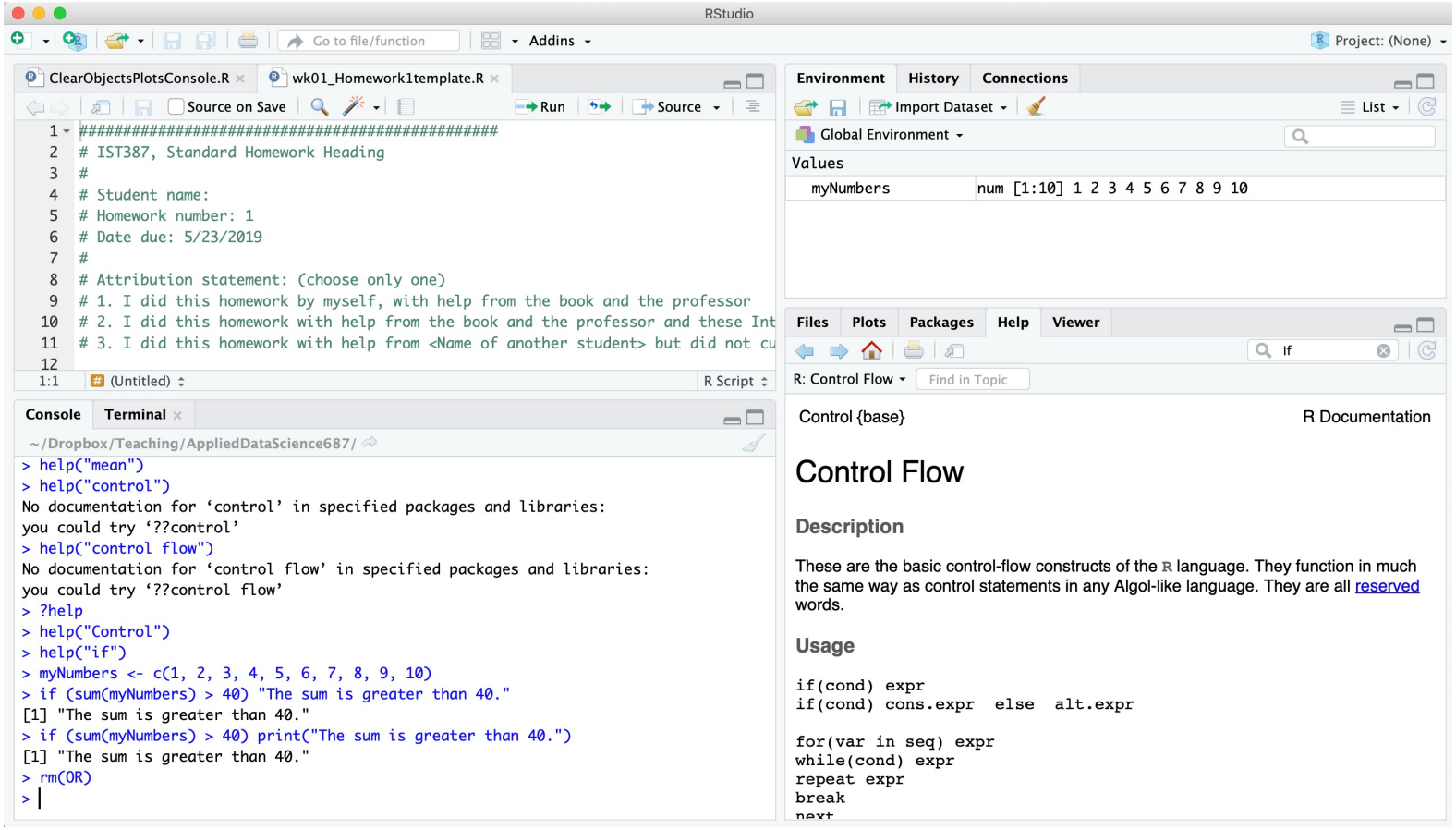
- In most breakout groups, I will randomly pair you with others in the class.
- Get to know those people a bit. Share your story with them.
- All handouts are on Blackboard
- Open `Week 1/1. Week1thinkPairShare.pdf`
- You'll be assigned a question to discuss for 10 minutes…
- …then share for 2 minutes
- Share any questions you have at https://codeshare.io/aJDyRX

R-Studio Clinic #1

Breakout 1 – R-Studio Skills Checklist

- This is not a class in how to program in R, but everything about the homework and tests will be easier if you have solid R skills.
- This breakout practices your use of each of the four major windows in R-Studio.
- You will have an opportunity to begin the important habit of adding comments to your code files.
- You will share your group's code at the end for the benefit of your group members and other people in the class at https://codeshare.io/aJDyRX

R-Studio Main Window Panes



Doing Good R Pair Work in Breakout Groups

- One person should be the driver and should share their R-Studio screen.
- The other people should be the navigator. The navigator dictates code and comments.
- Most classes have multiple breakouts. Try to switch roles.
- Take your time and make sure that both people understand what is going on.

Navigator dictates the code. Can/should consult with driver for ideas and suggestions.

Credit: SolutionsIQ

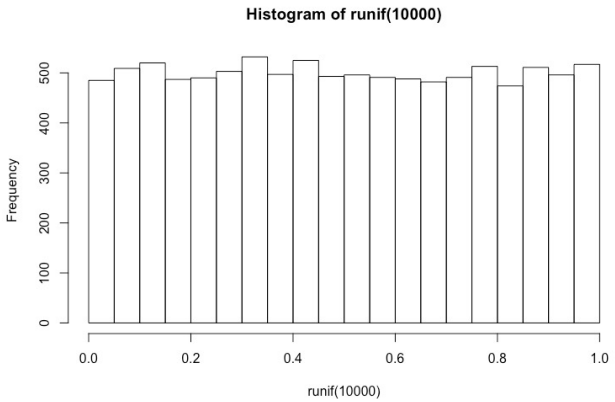
Driver runs the keyboard and the mouse. Listens to the navigator. Can comment and suggest.

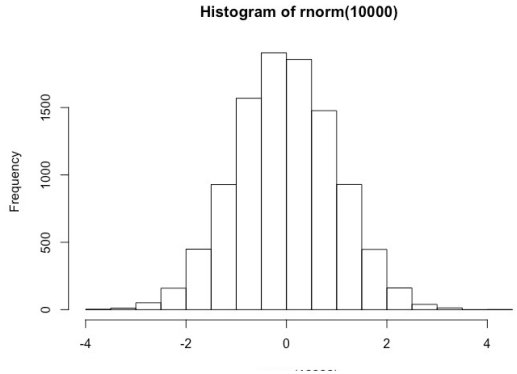
A key goal of each breakout is to learn from each more: both mistakes and successes.

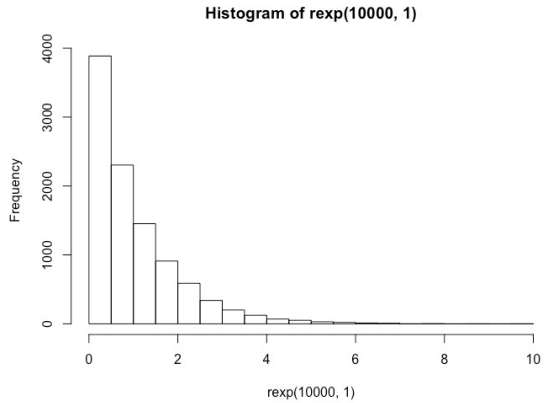
The exercise (2. R Studio Clinics.Rmd)

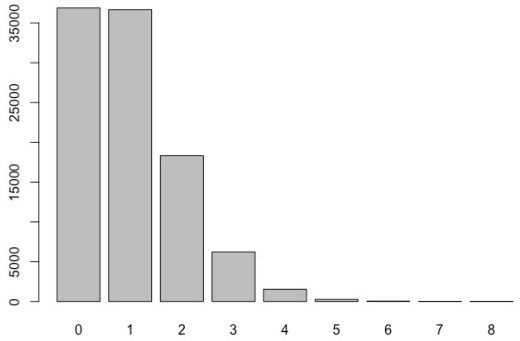
```
# Navigate to the appropriate folder using the Files tab of the Content pane;
# click to load tinyData.Rdata (from the Google Drive).
# Or use the open dialog in the Environment tab of the Environment pane.
# No Rstudio? Use https://www.tutorialspoint.com/execute_r_online.php

load("tinydata.Rdata")      # Don't need if success with the dialog
td <- c(0,0,0,2,3,4,5)      # If all else fails!

mean(td)
sd(td)

# Share the code you wrote at https://codeshare.io/aJDyRX
```

Distributions









R-Studio Clinic #2 – Descriptive Statistics and Distributions

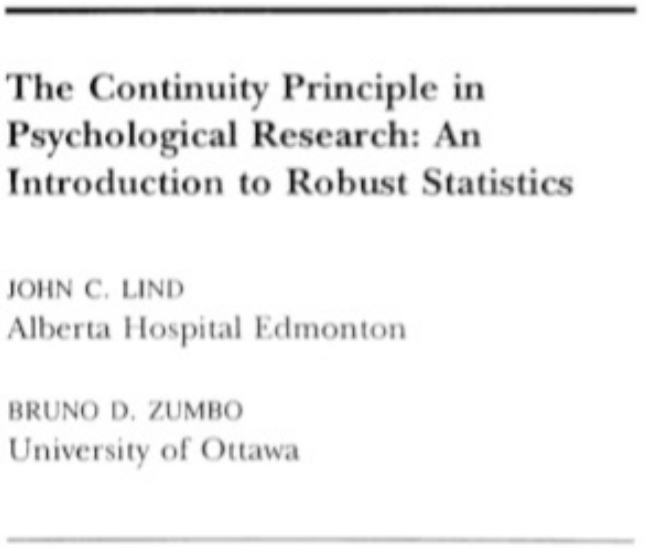
- Calculate mean, var, sd on the variables in the built-in rock dataset
- Examine the histograms and describe the distributions
- Share your group's code and comments at https://codeshare.io/aJDyRX for the benefit of your group members and other people in the class.

Share your code

- Share the code you wrote at https://codeshare.io/aJDyRX

Paper of the Week – Lind & Zumbo, 2012

- "Robust" statistics are less sensitive to outliers and distortions of normality
- Estimators such Huber's M-Estimator calculate means and standard deviations which, when replicated over many samples are more likely to be close to the population values
- Try the huber( ) function in the MASS package: huber(iris$Sepal.Length)



The Continuity Principle in Psychological Research: An Introduction to Robust Statistics

JOHN C. LIND
Alberta Hospital Edmonton

BRUNO D. ZUMBO
University of Ottawa

Homework 1

- Based on problems 1, 3, and 4 on page 20
- Complete the `ps week 1.Rmd` notebook, knit to PDF and submit the PDF