

Agenda (Week 6 Day 1)

1. Announcement:

- You will be given 3 minutes each to present homework to a partner next class
- You will be randomly selected to present your work to the class!!!

2. Midterm Discussion

3. Time series in R

Midterm: Format

- DAY 1

- in-class hard copy
- written exam (bring pen)
- closed notes
- no internet
- all material is fair game

- DAY 2

- in-class
- open notes
- exploratory data assignment in R
- will submit in a new midterm repo

Midterm: DATA

Data structures: long versus wide

- Data types: numeric, character, factors
 - nominal, ordinal, interval, ratio
- Data generating process
- Organizing data using data.table
- How to index a data.frame or matrix
- What functions do we use to initially 'check' the data?

Midterm: VARIABLES

- Measures of summarization
- Distributions, what are they?
 - Differences between histogram and density
 - Assumptions of a normal distribution
- Correlations
- How best to explore univariate versus bivariate data
- When to use which plots: scatterplots, histograms, boxplots, violin plots, line plots, etc. (And what these plots illustrate).
- Rankings and flows (alluvial diagrams)
- Time series data and visualization

Midterm: Principals of Visualization

- Critiquing graphics
 - Misleading graphics
- Grammar of graphics
- Color perception
- Measures of summarization
- Truthful Storytelling
- Bias
 - Patternicity
 - Confirmation bias
 - Observability (sometimes called availability) bias

Patternicity

- Definition:
 - The human tendency to find patterns in meaningless noise
- False positive: You hear a loud noise in the bushes. You assume it is a predator and run away. It was not a predator, but a powerful wind gust. Your cost for being incorrect is a little extra energy expenditure and a false assumption.
- False negative: You hear a loud noise in the bushes and you assume it is the wind. It is a hungry predator. Your cost for being wrong is your life.
 - Shermer's 2000 book *How We Believe*

Confirmation Bias

- Definition:
 - confirmation bias (or confirmatory bias) is a tendency to search for or interpret information in a way that confirms one's preconceptions or beliefs, leading to statistical errors.
- Example
 - Seeking out information in the media to confirm beliefs about gun control
 - Confirmation bias about racial stereotypes
 - *“We are more likely to remember (and repeat) stereotype-consistent information and to forget or ignore stereotype-inconsistent information If you learn that your new Canadian friend hates hockey and loves sailing, and that your new Mexican friend hates spicy foods and loves rap music, you are less likely to remember this new stereotype-inconsistent information.”*
 - Catherine A. Sanderson (Social Psychology)

Sample selection bias

- Definition:
 - sample selection bias, (observability bias, availability bias)
 - Sample selection bias is a type of bias caused by choosing non-random data for statistical analysis where a subset of data is systematically excluded due to a particular attribute
- Example
 - works in sample design (smart phone survey example)
 - but also in visual design (truncating data, not revealing entire set of data or hiding the true sample)