Introduction to Data Science
SKKU University, Summer 2015

Prof. Jevin D. West
University of Washington
Lecture 4 – July 2, 2015

Attendance

please sign in to class every day

Agenda

- 9:30 – 10:30 Group Project
- 10:30 – 10:40 Break
- 10:40 – 11:15 Black Box Exercise
- 11:15 – 11:45 Government Data
- 11:45 – 12:00 Questions

Logistics

- Attendance: sign-in to class every day
- Repository naming: firstname_lastname
- Repository naming (groups): groupname
  - In your team repository, create a file your team members and your email addresses.
- Share your repository to me (jevinw)
- Class materials can be found in this repository:
  - https://github.com/jevinw/SKKU_DataScience_2015

Group Project

GroupProject.pdf

https://github.com/jevinw/SKKU_DataScience_2015/assignments/

Group Project Schedule

- July 3: Dataset Identified (1 paragraph)
- July 7: Question Identified (1 page)
- July 10: Preliminary Analysis (2-3 pages)
- July 15: Final Paper (4-5 pages)
- July 17: Final Presentation (~ 8 minutes)

Group Project

Example Presentations

Example Papers

Data

Homework

- **Group Project**
  - ❑ Identify a dataset for your group project
  - ❑ Describe the dataset in 1 paragraph
  - ❑ Push the paragraph description to your team repository
  - ❑ **DUE** at the beginning of class on July 3

- **Assignment #1**
  - ❑ Identify an open *government* data set (this is different than your group data set)
  - ❑ Create two plots in R describing some aspect of the data. These can be any kind of plot (e.g., scatter plot, histogram, bar plot, etc)
  - ❑ Push your R script and your two plots to your Github repository
  - ❑ **DUE** at the beginning of class on July 3
  - ❑ You can work in groups on the data but you need to create your own plots