

Preparing for Datathon

Rebecca C. Steorts, Duke University

STA 325, Supplemental Material

Agenda

- ▶ Intro to R/RStudio and EDA analysis
- ▶ Dealing with large files in R
- ▶ Visualizations with ggplot

Intro to R/RStudio and Basic Commands

- ▶ We've been over R/RStudio and EDA many times in class, but please find some extra resources just in case you find this helpful.

https:

`//github.com/mine-cetinkaya-rundel/df2017_workshops/
blob/master/intro_to_r/intro_to_R_DF2017.pdf`

How to deal with large files in R

- ▶ Avoid the following commands: `read.table` and `read.csv`
- ▶ I find the most useful command is `fread`, which is in the `library(datatable)`
- ▶ For a tutorial on reading data, see some information from DataFest done by Colin Rundel and Mine Cetinkaya-Rundel

https://github.com/mine-cetinkaya-rundel/df2017_workshops/tree/master/bigdata

Visualizations with ggplot2()

`https://github.com/mine-cetinkaya-rundel/df2017_workshops/tree/master/viz/ggplot2`