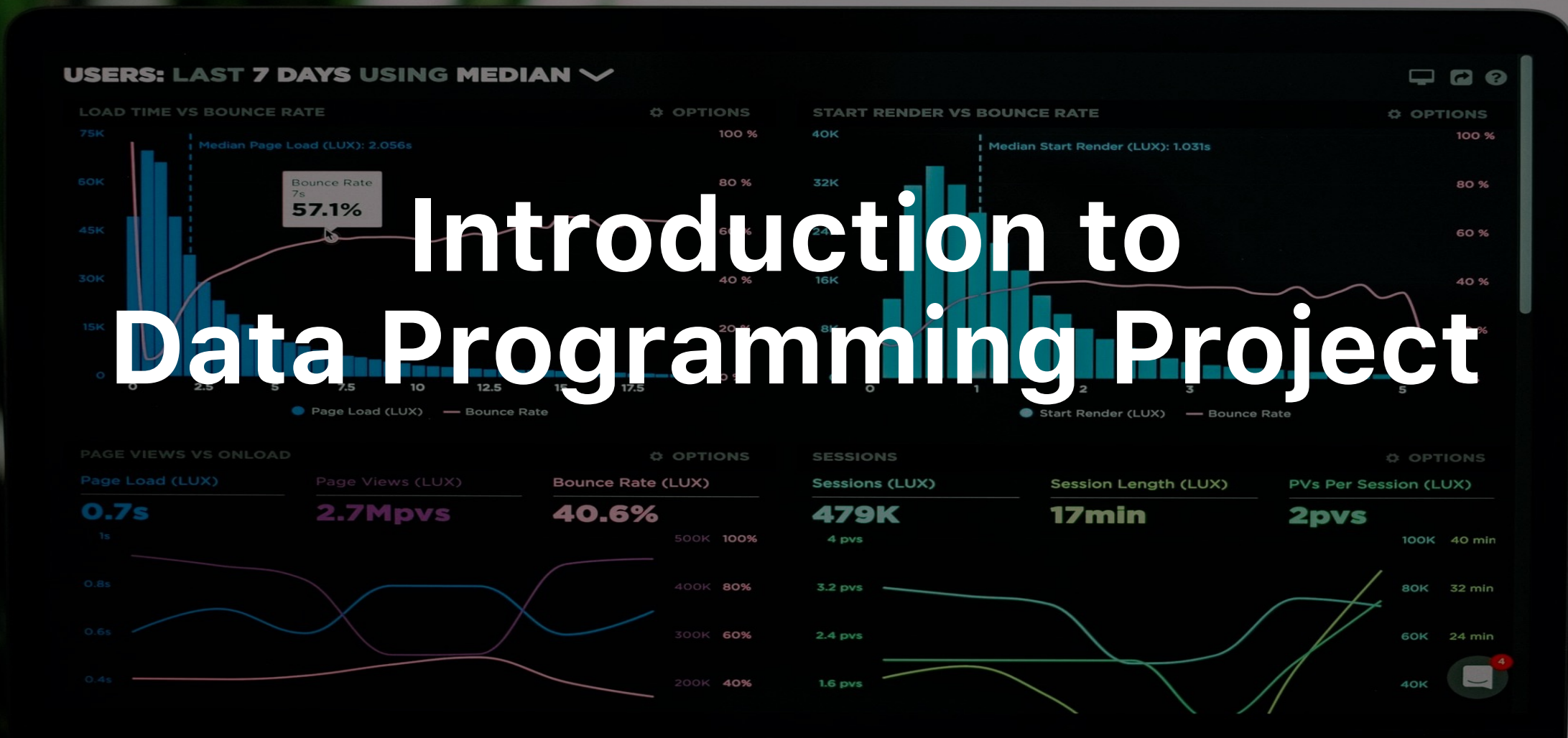
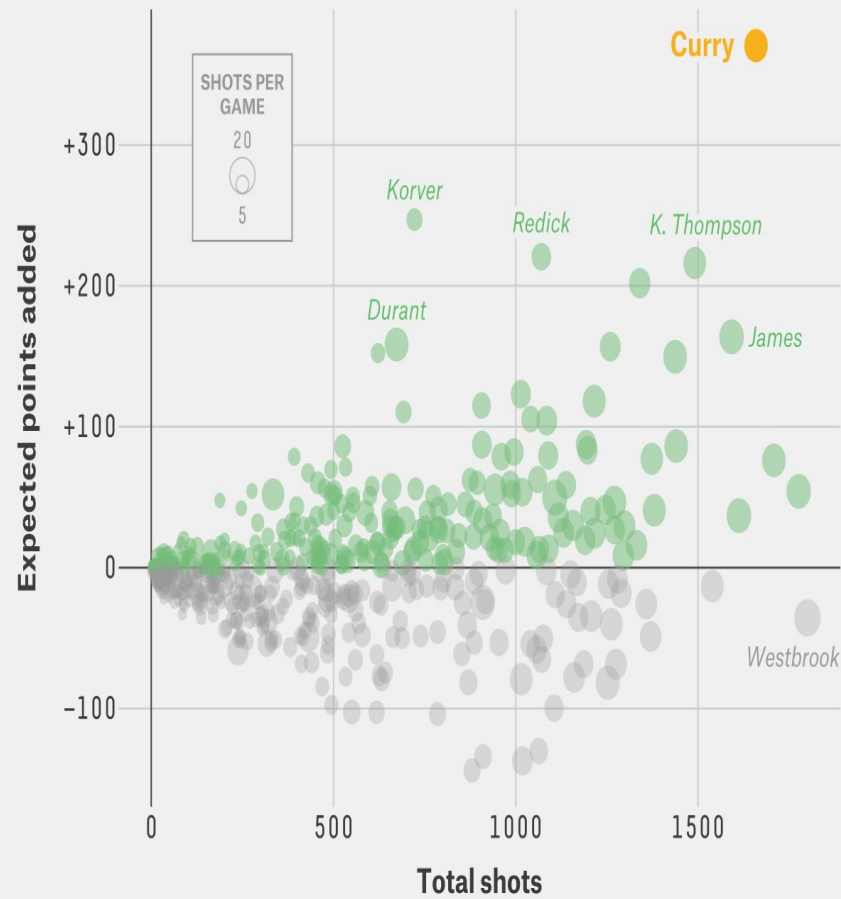


Introduction to Data Programming Project



Curry Is The Most Valuable Shooter (By A Lot)

Shooting value added (based on distance, shot clock and defender distance) vs. shots, by player; last season through Nov. 28, 2015

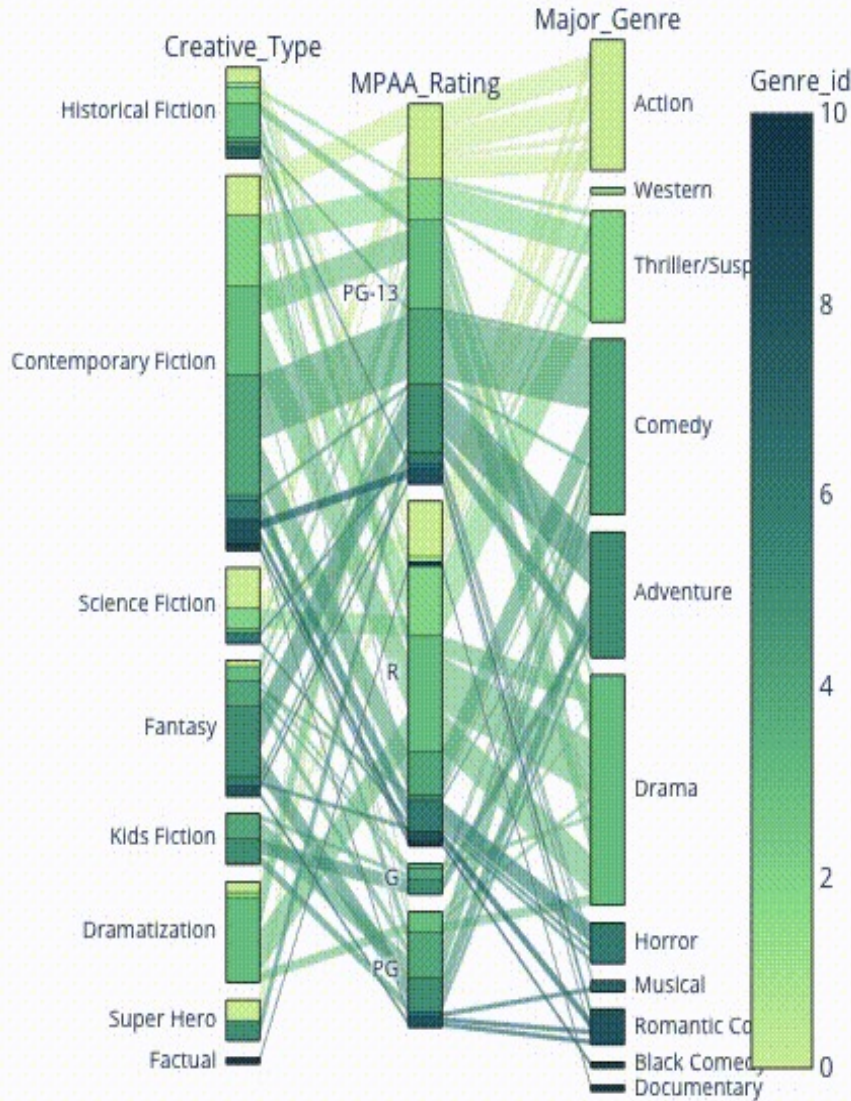


FIVETHIRTYEIGHT

BASED ON DATA PROVIDED BY NYLON CALCULUS

Guidelines

- **Data Programming Project (40% of grade)**
 - Analyze a real-world issue in analytics
 - Describe topic and objectives of the project
 - Locate and compile a dataset from a reputable source in order to tackle the chosen issue
 - Provide an overview of all selected variables in the dataset
 - Multiple analyses, visualizations, and interpretations of the data and findings
 - Must provide all R code and packages utilized
- Required structure
 - ~15-minute presentation
 - PowerPoint, Keynote, Google Slides, or equivalent
 - **Presentation due Tuesday, May 4**



Guidelines

- This will be an individual project!
- Suggested format
 - **Present using 16-21 slides**
 - Introduction & Background (3-4 slides)
 - Methods & Data Collection (3-4 slides)
 - Results (6-8 slides)
 - Discussion (3-4 slides)
 - References (1 slide)
 - APA preferred



Guidelines

- **Introduction & Background (3-4 slides)**
 - *What will you be studying?*
 - Provide background information
 - Summarize previous research
 - Visit Google Scholar and other reputable sites
 - *What is the goal of the project?*
 - List your research objectives
 - *What will your project attempt to answer?*
 - List your research question(s) and/or hypothesis(es)
 - Should be based on previous research!



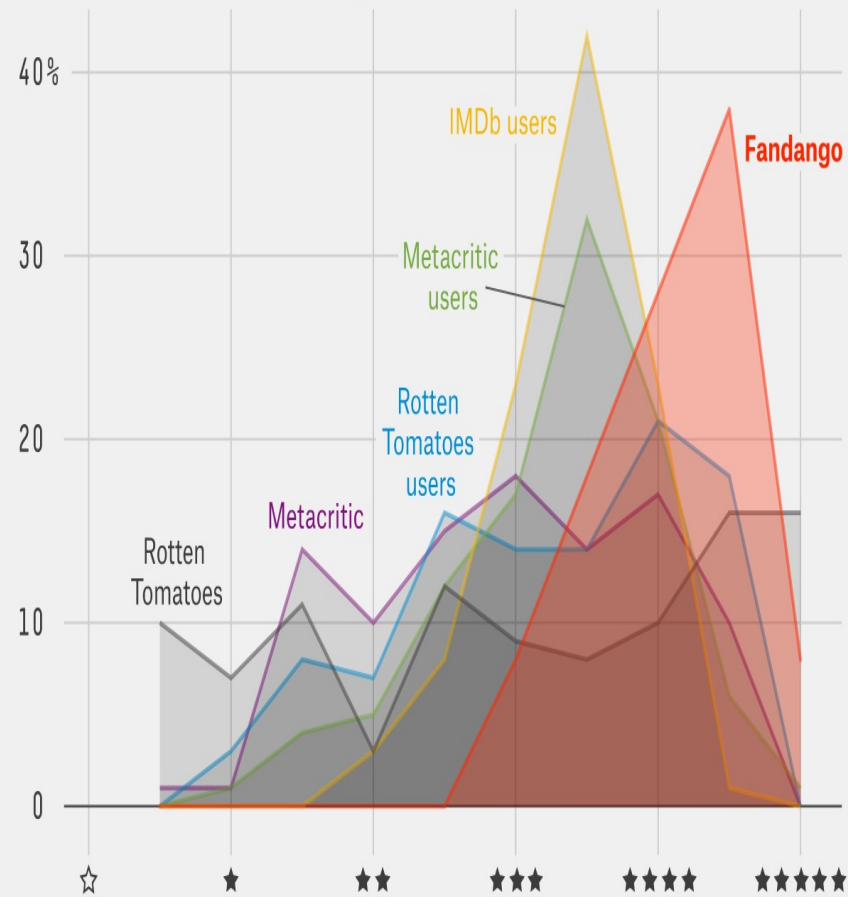
Guidelines

- **Methods & Data Collection (3-4 slides)**

- *What data will you use to answer your research question(s) and/or test your hypothesis(es)?*
 - Locate a real dataset (Kaggle, Dataverse, Data.gov, The World Bank, etc.)
 - Describe it!
 - Source, data collection process, and all variables utilized
 - Provide summary (e.g., descriptive) statistics and plots (e.g., histograms) as well

Fandango LOVES Movies

Normalized ratings distribution of 146 films in theaters in 2015 that had 30+ reviews on Fandango.com



FIVETHIRTYEIGHT

SOURCE: FANDANGO.COM, IMDB, METACRITIC, ROTTEN TOMATOES

Guidelines

- **Results (8-10 slides)**

- *What did you find?*

- Effectively explain statistical results
 - Include multiple tables and graphs
 - As you present, provide all the code used to perform the analyses and create the graphs



Guidelines

▪ Discussion (2-3 slides)

- *What should we take away from your project?*
 - Review your findings
- *Based on your analyses, what was the answer to your research question(s)?*
- *Did you find support for your hypothesis(es)?*
- *What explains your findings?*
 - Note strengths and limitations of your project
 - Practical implications and an explanation of how managers could use your findings
 - Directions for future research



Guidelines

- **References (1 slide)**

- *What outside sources did you use to put together the presentation?*
 - Cite everything!

- **Formatting**

- Slides should be well-organized and adhere to a proper format

- **Delivery**

- Speak with fluctuation in volume and inflection to maintain audience interest and emphasize key points
- Be well-versed with the information
- Do not only read from the slides!

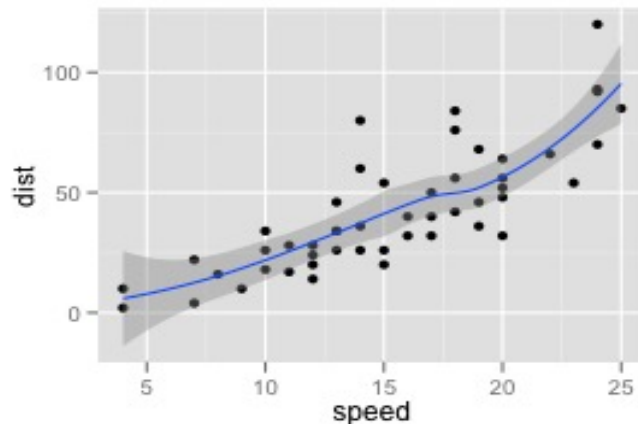
R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0   Min.   : 2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```

```
qplot(speed, dist, data = cars) + geom_smooth()
```



Guidelines

- **Supplementary Materials**
 - *How can we replicate your project's analyses?*
 - Provide all code utilized
 - Properly-commented Knitted R HTML document or equivalent
 - Should be easy to follow and reproducible in RStudio with no major issues



For Next Time

- **Full rubric posted on Canvas**
- Actual presentation must be recorded beforehand!
 - Use Zoom
 - Resources posted on Canvas
- **Today's Tasks**
 - Develop research topic, question, objectives, and hypothesis(es)*
 - Search for previous research
 - Begin looking for a dataset for your project
- **Next Class (Thursday, April 1)**
 - Continue working on Data Programming Project
- **Also...**
 - All R Package Group Cheat Sheets posted on Canvas