Intro to Data Science Workshop

NBA SHOT DATA VISUALIZATION

Anna Matlin
amatlin@princeton.edu

Overview

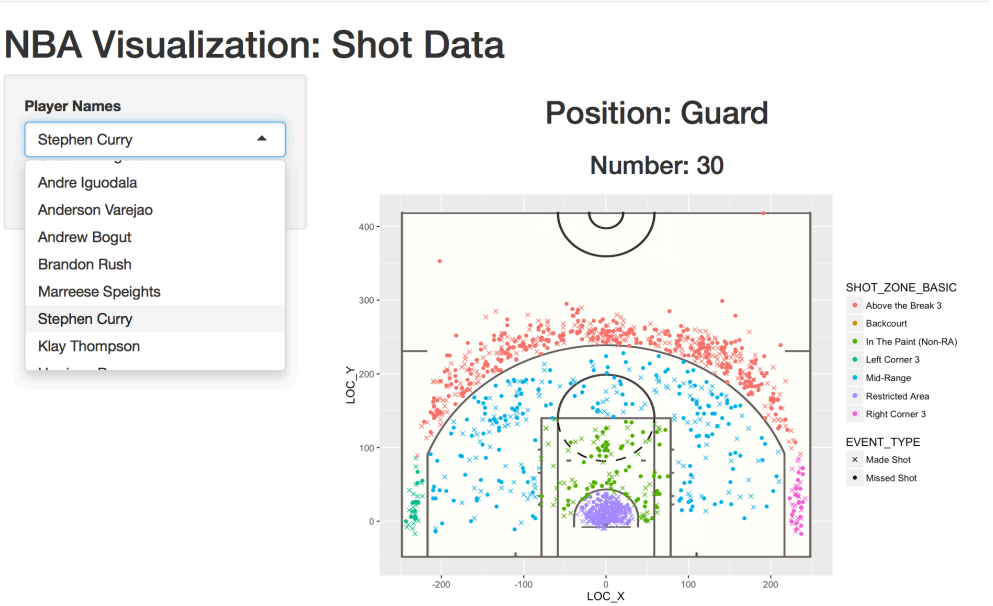
The world of sports data is vast, and significant quantities of information are publicly available online. The ability to access, capture, analyze, and illustrate that data is extremely valuable.

In this workshop, we will:

- scrape data from the web using Python
- process it in R
- guide participants (you) through the basics of making a Shiny app in R

By the end of the workshop, each participant will have made a visualization of player data as a web application.

End goal


NBA Visualization: Shot Data
Player Names
Stephen Curry
Andre Iguodala
Anderson Varejao
Andrew Bogut
Brandon Rush
Marreese Speights
Stephen Curry
Klay Thompson
Position: Guard
Number: 30
LOC_Y
LOC_X
400
300
200
100
0
-200
-100
0
100
200
SHOT_ZONE_BASIC
Above the Break 3
Backcourt
In The Paint (Non-RA)
Left Corner 3
Mid-Range
Restricted Area
Right Corner 3
EVENT_TYPE
Made Shot
Missed Shot

Resource dump

- All code used in this workshop: https://github.com/amatlin/NBAvis
- Stats data retrieved from http://www.nba.com/warriors/stats
- Scraping:
  - https://www.crummy.com/software/BeautifulSoup/bs4/doc/
  - http://web.stanford.edu/~zlotnick/TextAsData/Web_Scraping_with_Beautiful_Soup.html (tutorial)
- Shiny:
  - http://shiny.rstudio.com/
  - http://shiny.rstudio.com/articles/cheatsheet.html
- Download python: https://www.python.org/downloads/
- Download R and Rstudio:
  - http://lib.stat.cmu.edu/R/CRAN/
  - https://www.rstudio.com/products/rstudio/download/
- ggplot2:
  - http://ggplot2.org/
  - https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
- **Lots of code used** from http://thedatagame.com.au/2015/09/27/how-to-create-nba-shot-charts-in-r/

Step 1: Have Python and R installed

Python needed for web scraping

R, RStudio needed for visualization + Shiny

Step 2: go to Github repository

https://github.com/amatlin/NBAvis

→ open README and follow instructions for installation of necessary packages

Step 3: Scraping

Go to http://www.nba.com/warriors/stats to get a look for what we will be scraping.

Player IDs come from image links in HTML.

Find image of player, right click and select "Inspect"

Frameworks/APIs we will be using to scrape:

→ Beautiful Soup

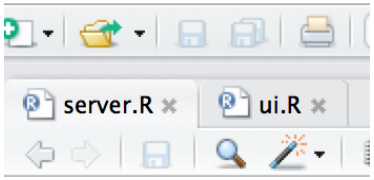
What is a Shiny app?

Check out http://shiny.rstudio.com/ for documentation + tutorials

→ "**A web application framework for R;** Turn your analyses into interactive web applications; No HTML, CSS, or JavaScript knowledge required"

→ each app is made containing two files

- **server.R**: behind the scenes calculations
- **ui.R**: what the user sees



→ designed for development in RStudio

→ Terms to keep in mind: *render*, *reactive*, *input/output*

Step 4: Start making our app

What does our data look like?

- Column names:

GRID_TYPE GAME_ID GAME_EVENT_ID PLAYER_ID PLAYER_NAME TEAM_ID TEAM_NAME PERIOD MINUTES_REMAINING SECONDS_REMAINING EVENT_TYPE ACTION_TYPE SHOT_TYPE SHOT_ZONE_BASIC SHOT_ZONE_AREA SHOT_ZONE_RANGE SHOT_DISTANCE LOC_X LOC_Y SHOT_ATTEMPTED_FLAG SHOT_MADE_FLAG

- First data point

1	Shot Chart Detail	0021400014	50	201939	Stephen Curry	1610612744			
	Golden State Warriors	1	7	29	Made Shot	Running Jump			
Shot	2PT Field Goal	Mid-Range	Left Side(L)		8-16 ft.	10	-81	72	
	1	1							

- Total data points for Steph Curry: 1341

Step 5: ggplot2

For more info check out

- http://ggplot2.org/
- https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

"ggplot2 is a plotting system for R, based on the grammar of graphics, **which tries to take the good parts of base and lattice graphics and none of the bad parts**. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it **easy to produce complex multi-layered graphics**."

Launch and run app

Questions?

Thank you!

- http://princeton-data-science.github.io/