



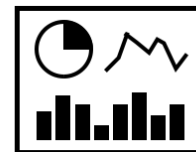
Foundations of Data Science

THE DATA SCIENCE “ECOSYSTEM”

R



Stack
Overflow



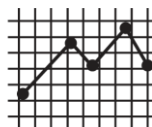
Dynamics
Web Applications
(Dashboards)



Data APIs



Statistics



Graphics



(operating system)



Open
Data



GIS



GitHub



Data-Driven
Documents



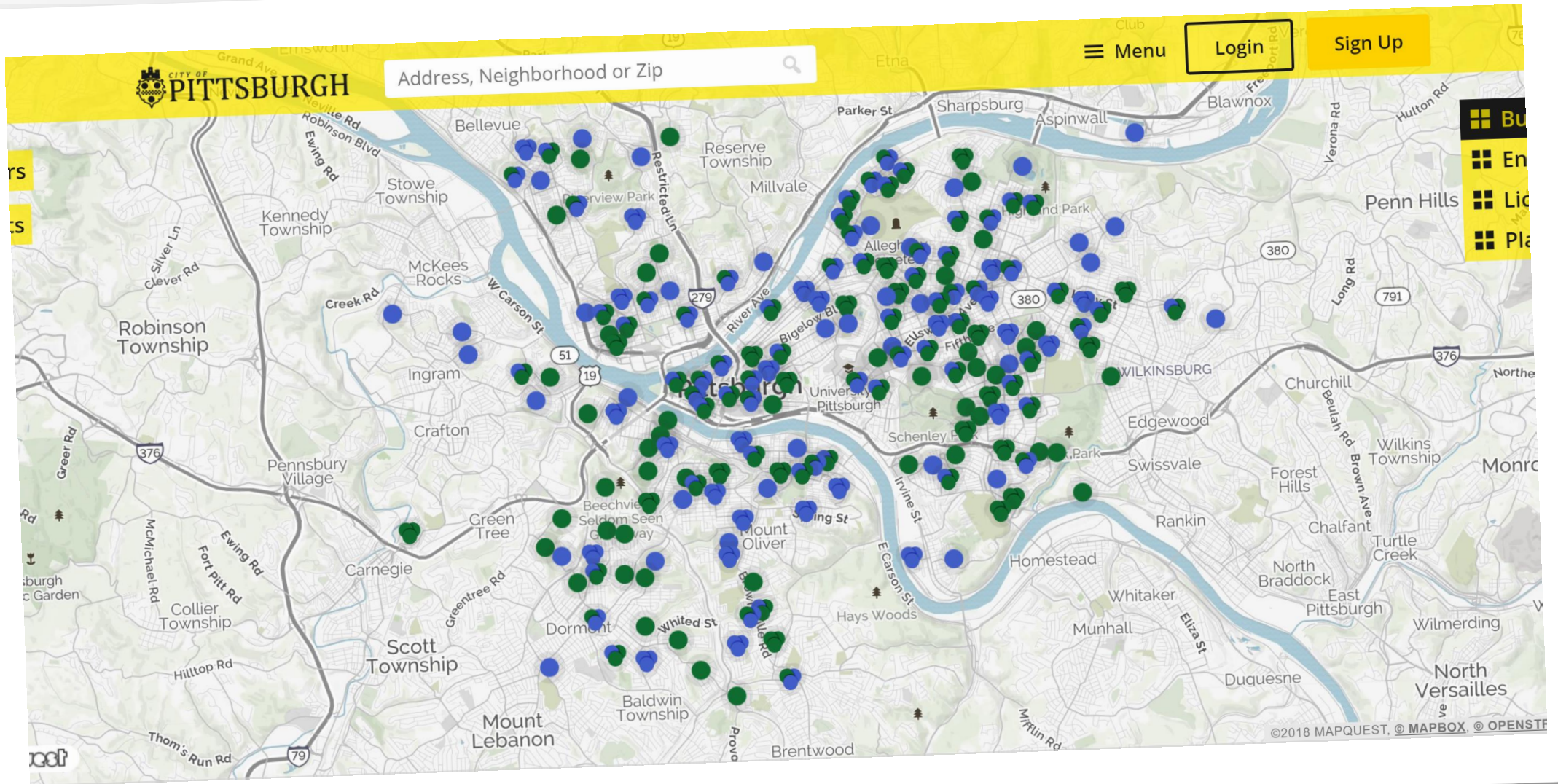
Atlas TI:\$700

Stata License:\$1200

SPSS:\$1200 - \$7800

ARCGIS:\$7000

R:FREE !



\$66,000 license per year!



```
install.packages( "leaflet" )
library( leaflet )

### LOAD DATA ###

code.violations <- read.csv("https://raw.githubusercontent.com/subartle/orangespot/master/data/code%20violations.csv")
lat.lon <- code.violations[ c("lat","lon") ]
lat.lon <- na.omit( lat.lon )

### CREATE MAP ###

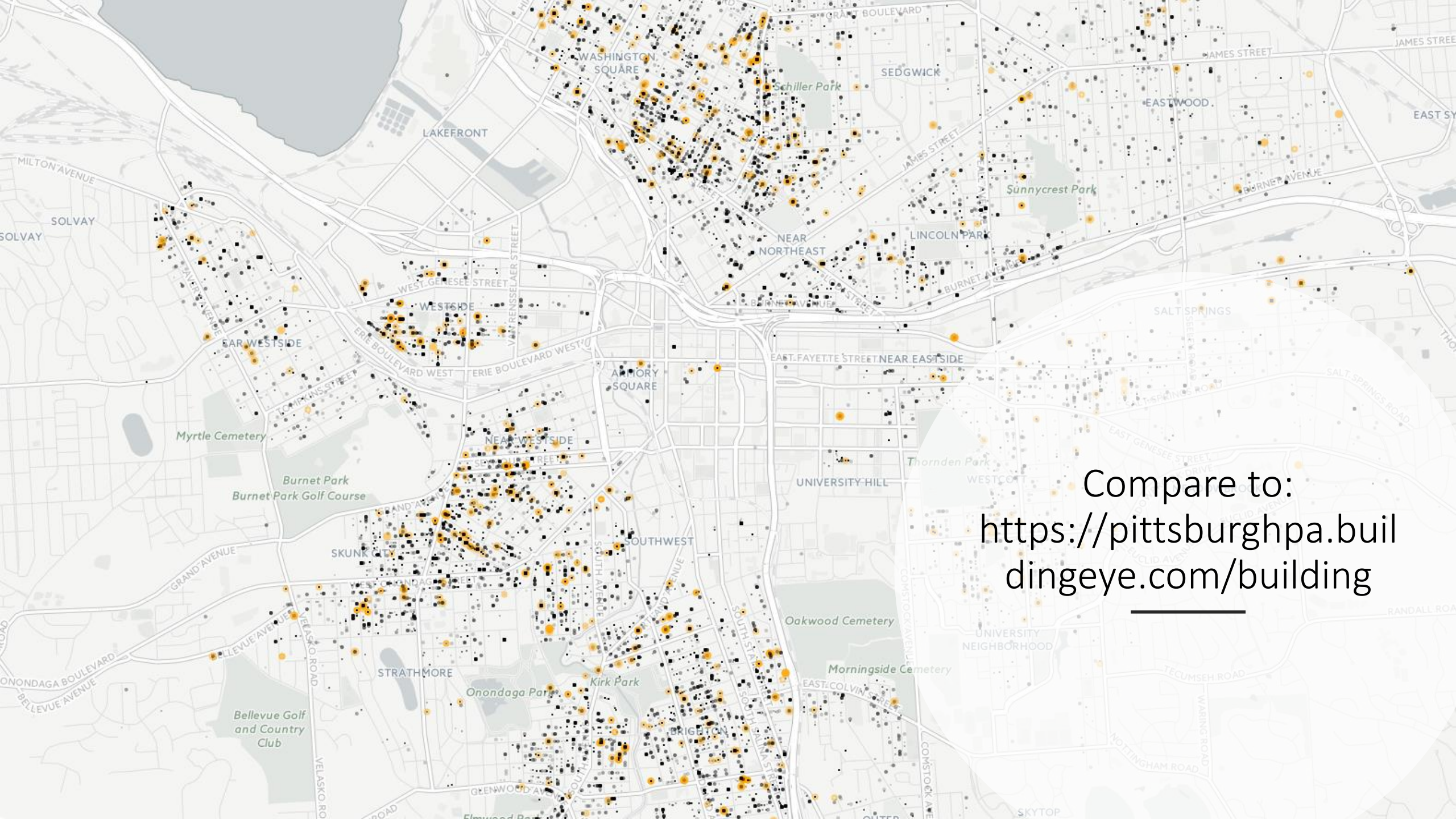
syr.map <- leaflet(data=lat.lon ) %>%
  addProviderTiles("CartoDB.Positron", tileOptions(minZoom=10, maxZoom=17)) %>%
  setView(lng=-76.13, lat=43.03, zoom=13) %>%
  setMaxBounds(lng1=-75, lat1=41, lng2=-77, lat2=45)

syr.map

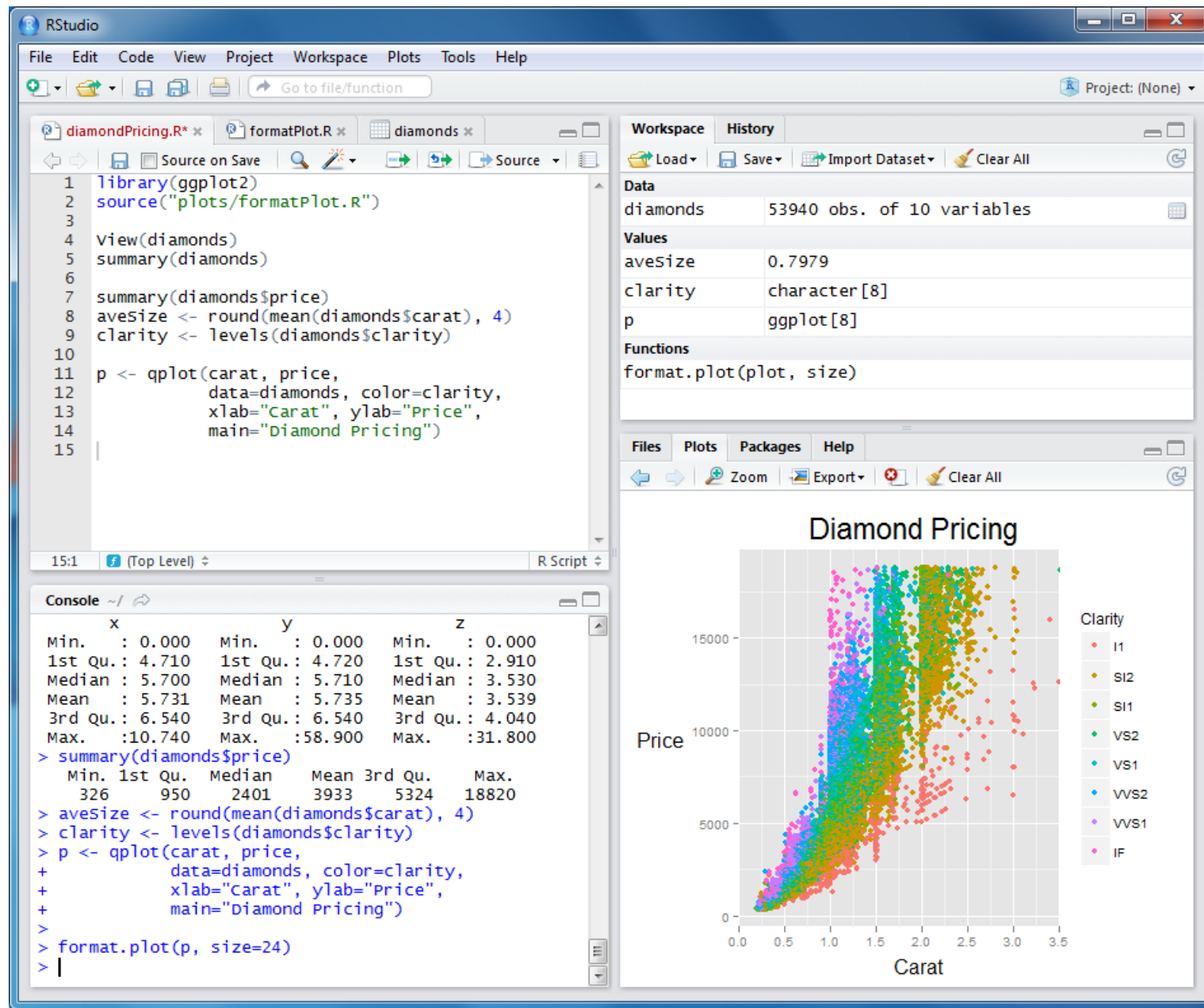
### COLOR CODE BED BUGS ###

col.vec <- ifelse( code.violations$Complaint.Type == "Bed Bugs", "orange", NA )
addCircles( syr.map, lng = lat.lon$lon, lat = lat.lon$lat, col=col.vec )
```

(you can build a similar platform in R with 12 lines of code)



Compare to:
<https://pittsburghpa.buildingeye.com/building>



DATA-DRIVEN DOCUMENTS

MARKDOWN



1. Reports [[html](#)] [[pdf](#)]
2. Books [[link](#)]
3. Presentations [[link](#)]
4. Tutorial [[link](#)]
5. Websites [[link](#)]
6. Dashboards [[link](#)]
7. Web Apps [[link](#)] [[link](#)] [[CDC](#)]

```
---  
title: "MARKDOWN DEMO"  
output: pdf_document  
---
```

Here is the heading

Here is some text.

A sub-heading

Perhaps you want a list:

- Item 1
- Item 2

And here is some code:

```
```{r}  
plot(cars, col="steel blue", cex=1.5, pch=19)
```
```

render



MARKDOWN DEMO

Here is the heading

Here is some text.

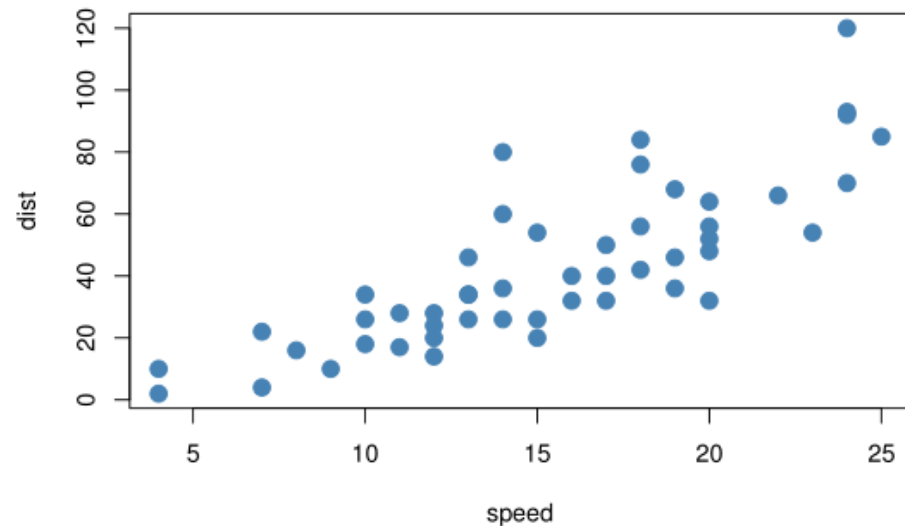
A sub-heading

Perhaps you want a list:

- Item 1
- Item 2

And here is some code:

```
plot( cars, col="steel blue", cex=1.5, pch=19 )
```

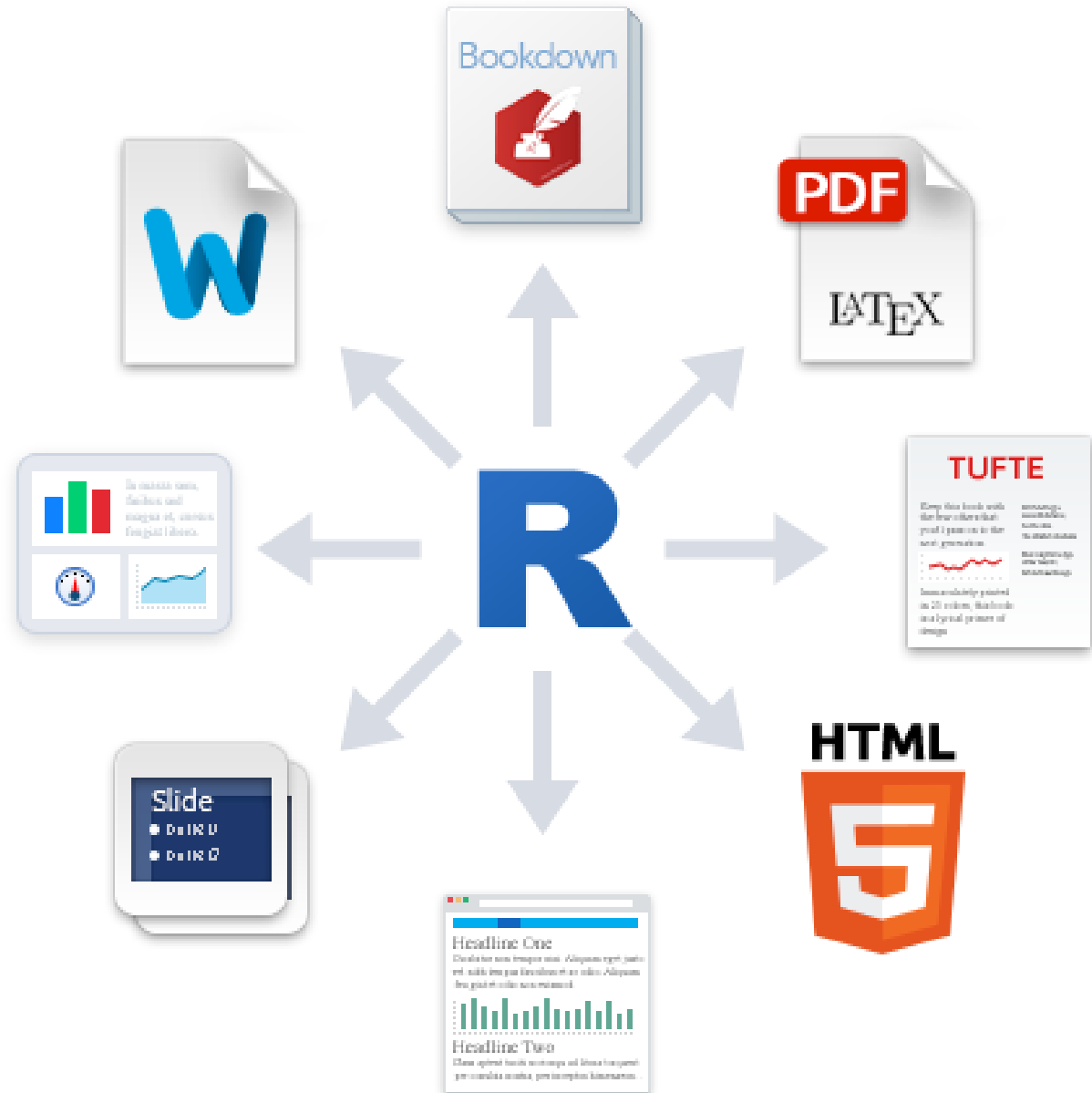




```
---  
title: "MARKDOWN DEMO"  
output: pdf_document  
---
```

```
---  
output: html_document  
---
```

```
---  
output: flexdashboard::flex_dashboard:  
---
```





EXAMPLE LABS

[illegible]

Printed First Name Page



Andrew Friedman, the Tampa Bay Rays' general manager, has found plenty of veteran, longtime relievers with high inflation rates, while developing a newer rotation through the farm system. (Last year's Tampa Bay staff included the four-time Cy Younger as Johnny's first. Friedman finished off

to help and control the action, and there's a lot of things you can do with it. But if a better struggle with violence, for example, then all a politician is making, the better, after all, is to make what pitch is coming.

"The time would have been called a bull test by I'd have been willing, and if he had made contact, there's a high probability it would have been a definite he wouldn't have done anything with."

"Supercomputer is researched in our game, and I think that's good, and there's various pieces of our game where that's really important. But it's definitely a commitment for console and players to figure out how to find their optimal business."

Flare Shogren is harder for him to say.

"Because they're taught to throw the ball over the plate. And now, whether it's P.E. or whatever the case may be, they may not be as concerned about using the middle of the plate as they have in the past because they know they're not going to bring it off the wall every time."

"It works now great, but if a guy's throwing it down the middle, who wants to catch it?"

"His team."

For the modern times, that is, never said they don't.

Plachette cost per team per game

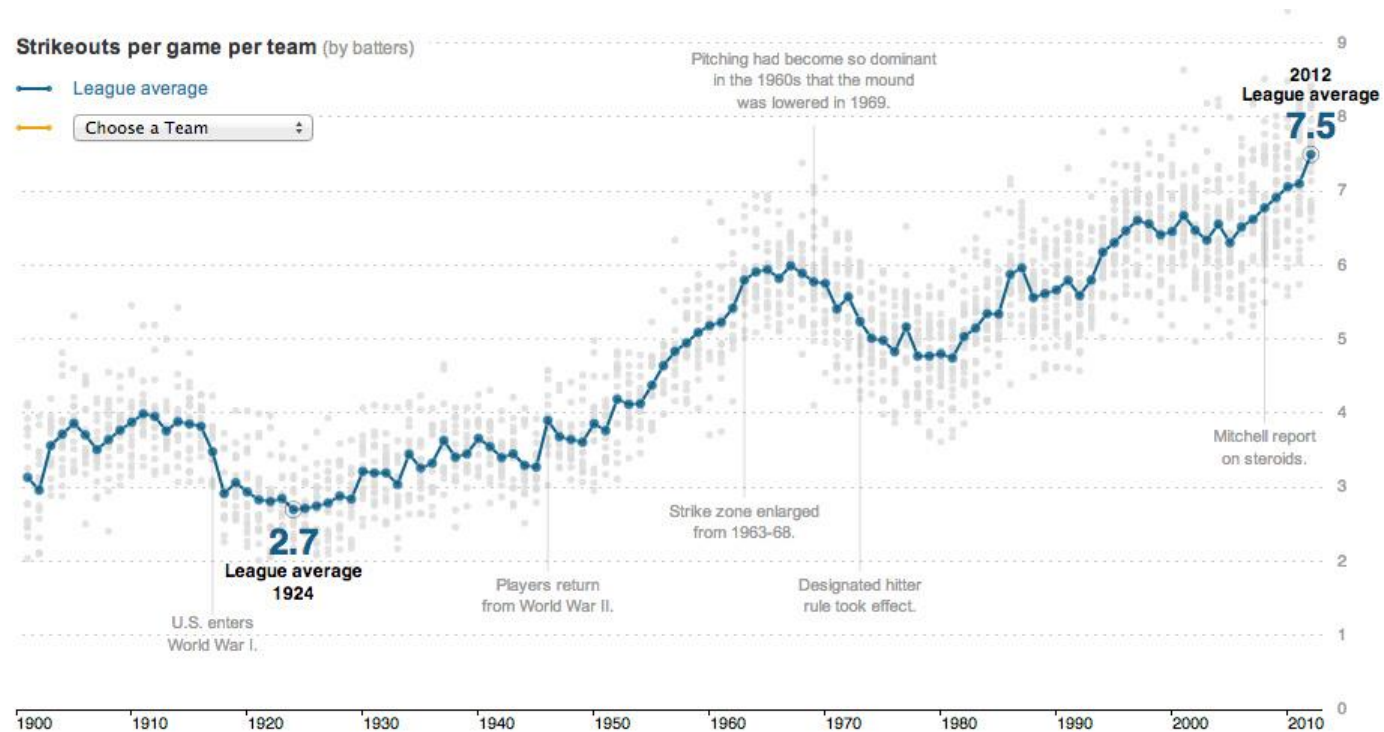
- Individual teams
- Major league average

In 1974, the year with the lowest amount of games, teams used a single pitcher nearly half of the time.

The graph shows a steady increase in the number of people aged 65 and older in the U.S. population from 1980 to 2020. The number of people grows from approximately 20 million in 1980 to over 40 million in 2020.

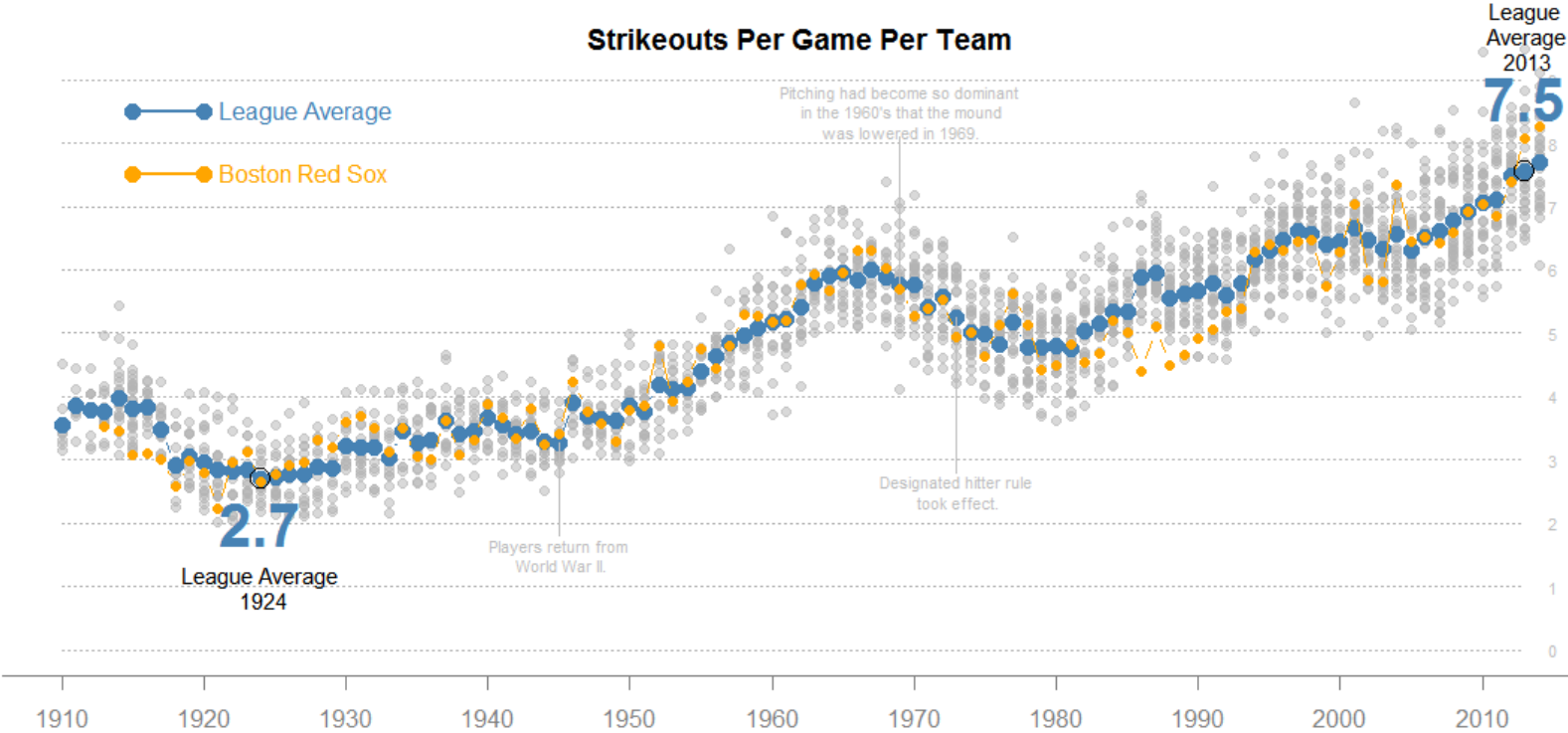
By IVAN SEPPER

FROM THE NYT



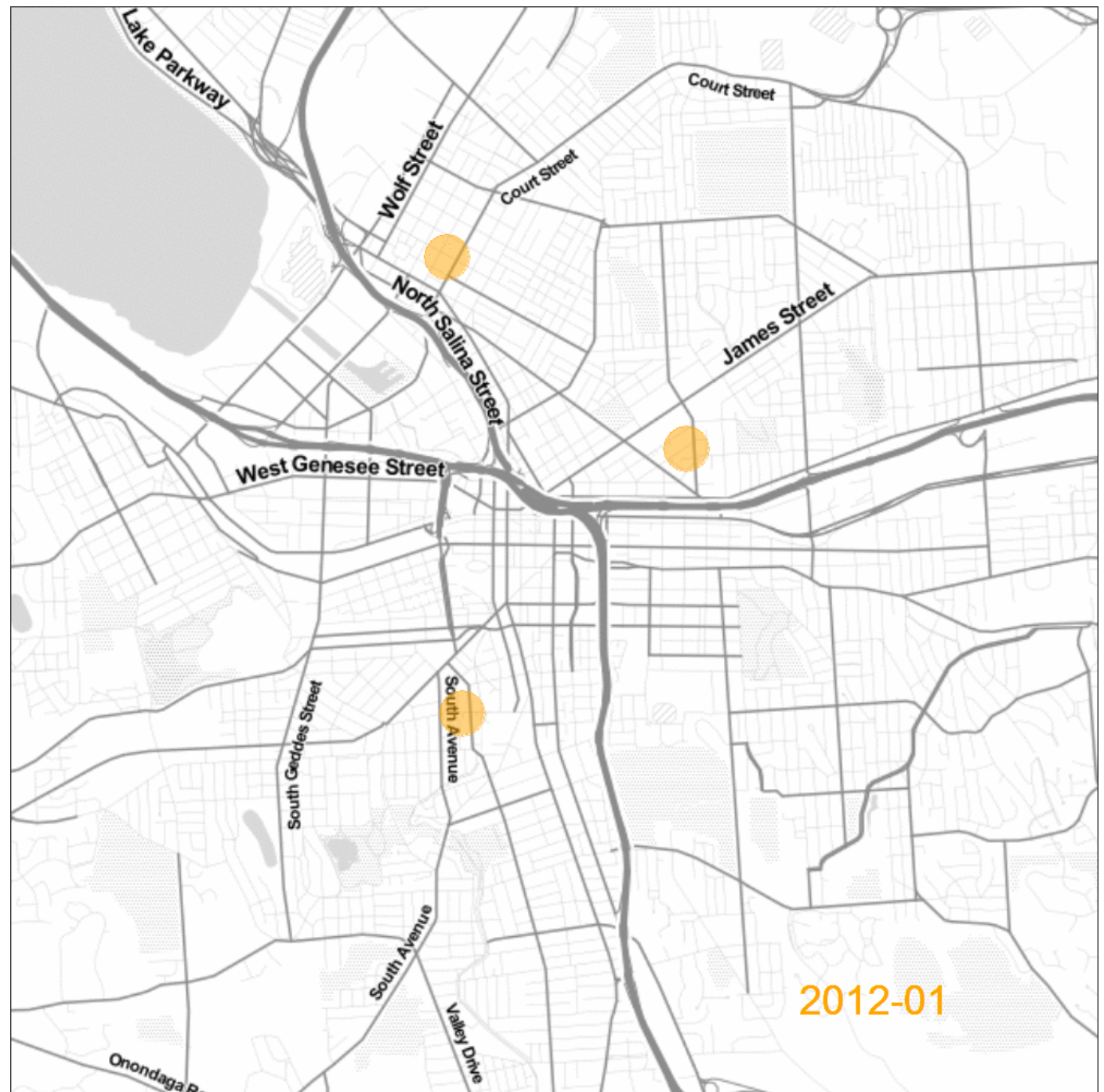
<http://chartsnthings.tumblr.com/post/47670081904/climate-change-crowbars-and-strikeouts>

FROM A CLASS ASSIGNMENT



ANIMATIONS

BED BUG
CASES IN
SYRACUSE, NY
2012-2016



PROGRAM DESIGN

ONE YEAR SCHEDULE

SEMESTER 1

A

Program Eval 1

Data Science 1

SWK 623 - Eval

Data Practicum

SEMESTER 2

Program Eval 2

Data Science 2

Program Eval 3

Data Science 3

SEMESTER 3

Capstone
(3 credits,
15 weeks)

Elective

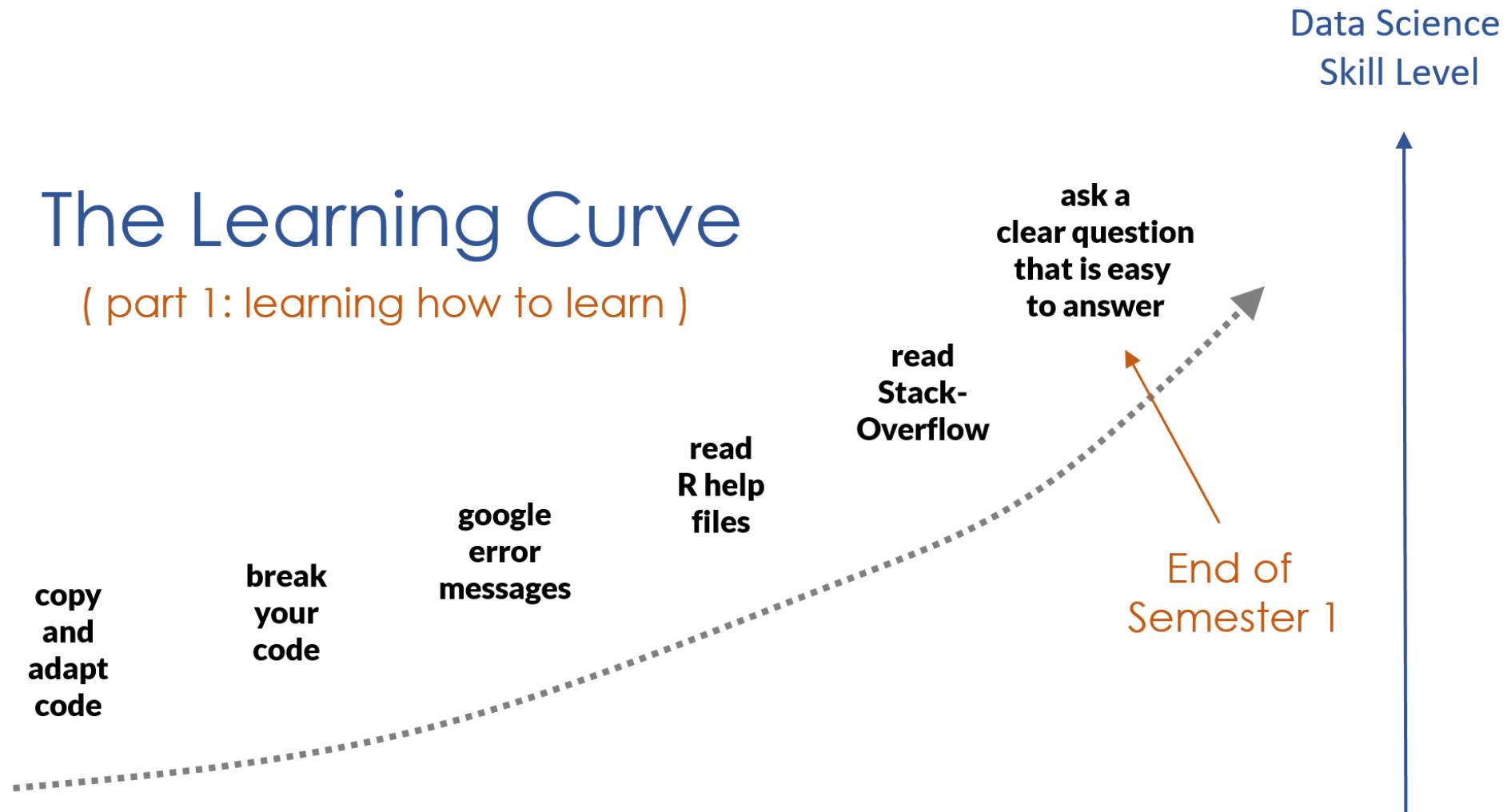
Elective

THE LEARNING CURVE

CPP 526 → PRACTICUM → CPP 527 → CPP 528

The Learning Curve

(part 1: learning how to learn)



“VERBS AND NOUNS”

arguments (function input)



```
makeCookies <- function( flour, eggs, sugar )  
{  
  
  batter <- mix( flour, eggs, sugar )  
  cookies <- bake( batter, temp=400 )  
  
  return( cookies )  
  
}
```



} “recipe”

return value (function output)



GETTING HELP

GETTING HELP (20 MINUTE RULE)

- Work in groups
- Google errors – read StackOverflow  question about functions
- Post questions to GitHub  question about labs
- Attend review sessions
- Schedule office hours

REVIEW SESSIONS

GRADES

GRADES

- 60% Labs (weekly - pass/fail)
- 10% Discussions - YellowDig
- 30% Final Project - Dashboard

