

Jesse Lecy
ASU

BOOTCAMP IN R

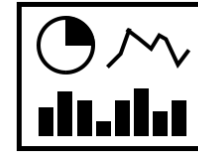
MOTIVATION

- Why should I learn R?
 - Efficiency of research
 - Public stewardship
 - Reproducibility crisis
 - It's fun!
- Can I learn R?
 - Do I need a special background?
 - How long does it take?
- How should I learn R?
 - Where to start?
 - Resources and support

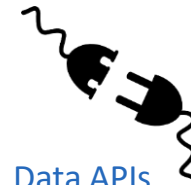
R



Stack
Overflow



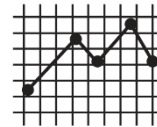
Dynamics
Web Applications
(Dashboards)



Data APIs



Statistics



Graphics



(operating system)



Open
Data



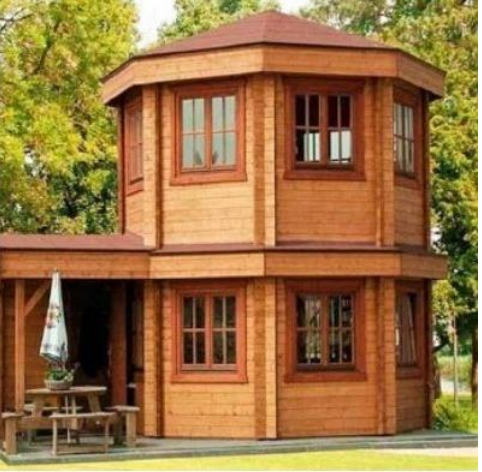
GIS

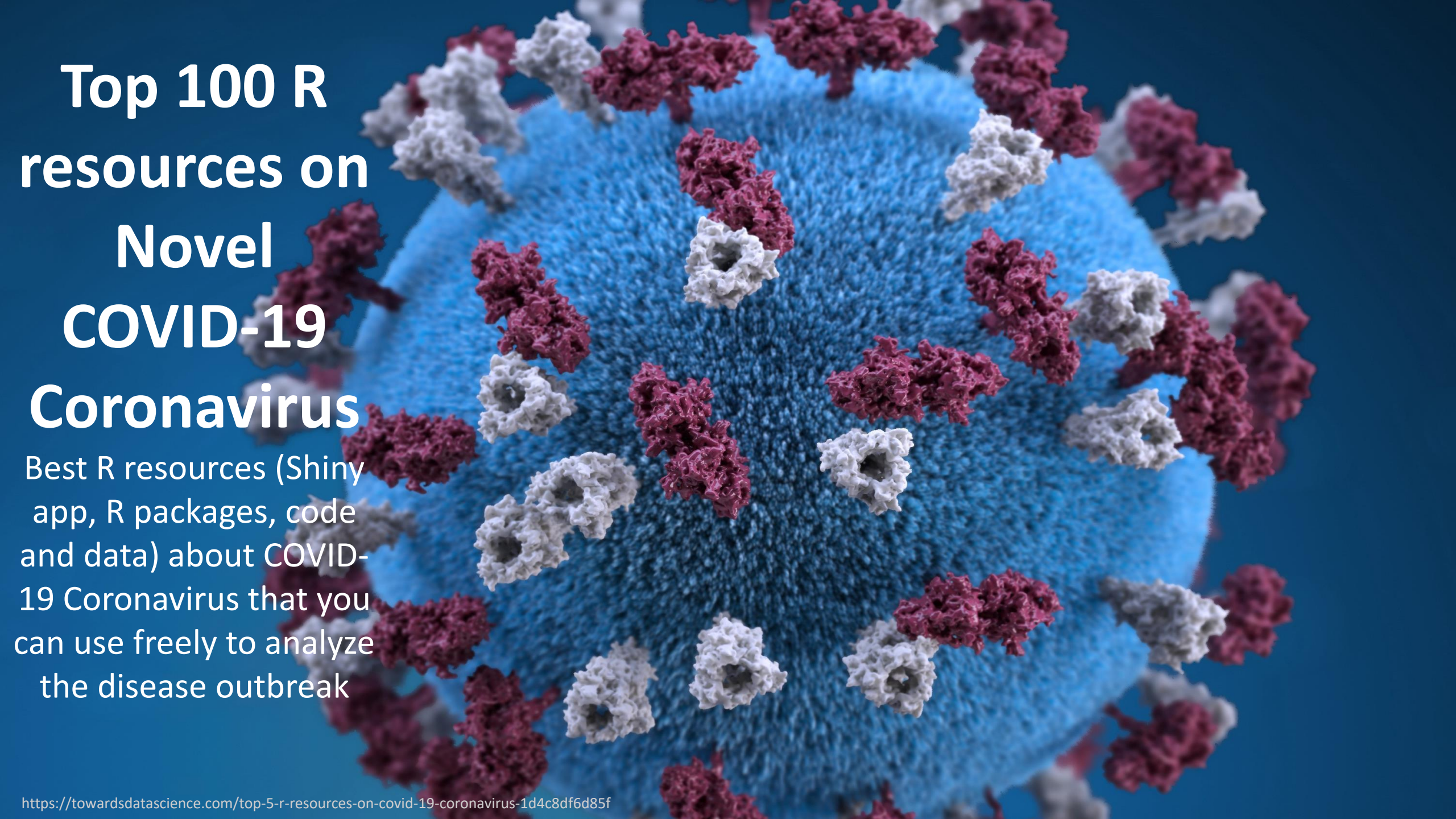


GitHub



Data-Driven
Documents





Top 100 R resources on Novel COVID-19 Coronavirus

Best R resources (Shiny app, R packages, code and data) about COVID-19 Coronavirus that you can use freely to analyze the disease outbreak



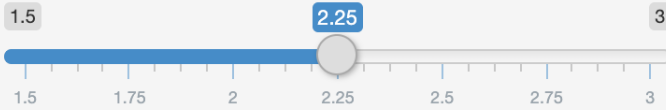
<https://towardsdatascience.com/top-5-resources-on-covid-19-coronavirus-1d4c8df6d85f>

Rapid sharing of

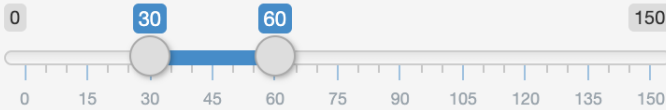
- Data (APIs, packages)
- Models
- Dashboard templates
- Code vignettes

DASHBOARD

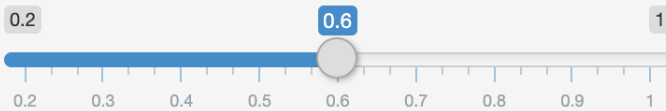
Choose a value for R_0



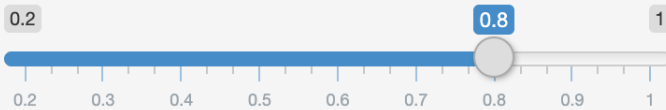
Choose the first 'social distance period' (sdp 1)



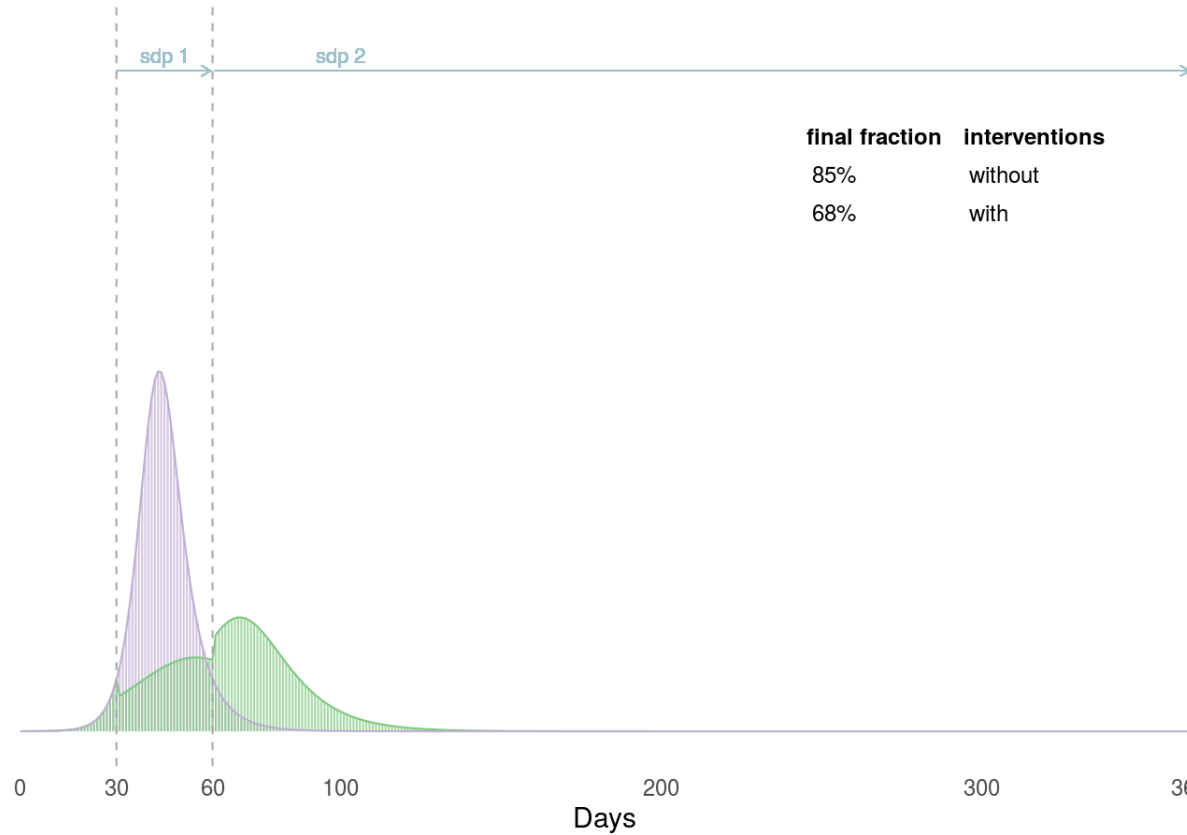
Reduction during first period (sdp 1)



Reduction after first period (sdp 2)



Daily new cases in % of the population



final fraction	interventions
85%	without
68%	with

Interventions
— without
— with

sdp: social distance period

TOTAL COUNTRIES AFFECTED: 154

Confirmed
242,708



Yesterday: 214,910

Active 147,987
(61%)



Yesterday: 122,970

Recovered
84,854 (35%)



Yesterday: 83,207

Deaths 9,867
(4.1%)



Yesterday: 8,733

Countries showing in the
geospatial map:

Afghanistan, Albania, Algeria, ▾

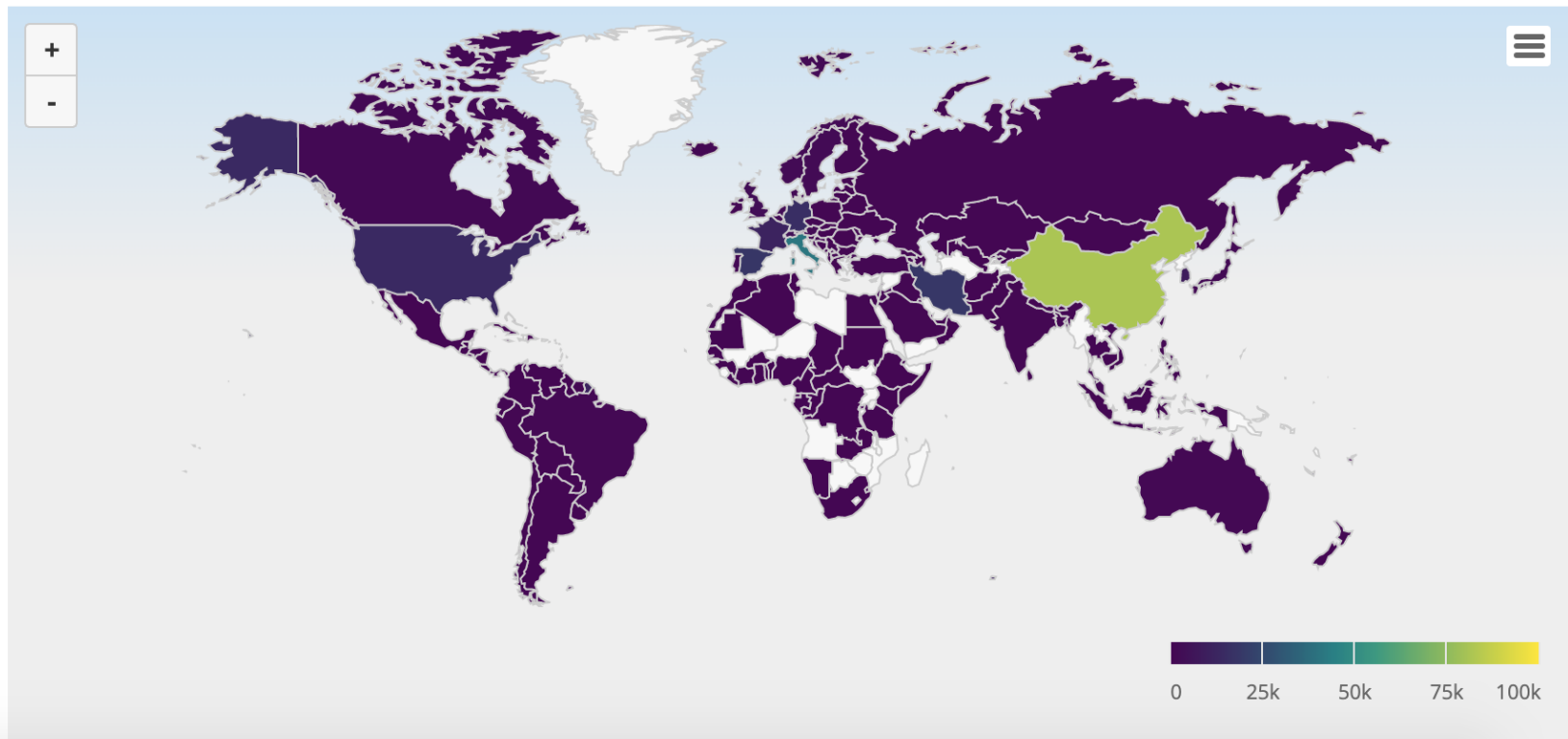
Specify the outcome to be shown in
map:

Total Cases

Recovered cases

Deaths

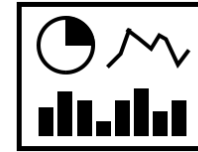
Active Cases



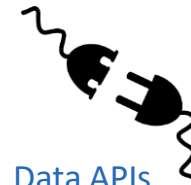
R



Stack
Overflow



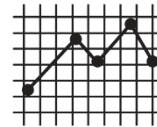
Dynamics
Web Applications
(Dashboards)



Data APIs



Statistics



Graphics



(operating system)



Open
Data



GIS



GitHub



Data-Driven
Documents

| R

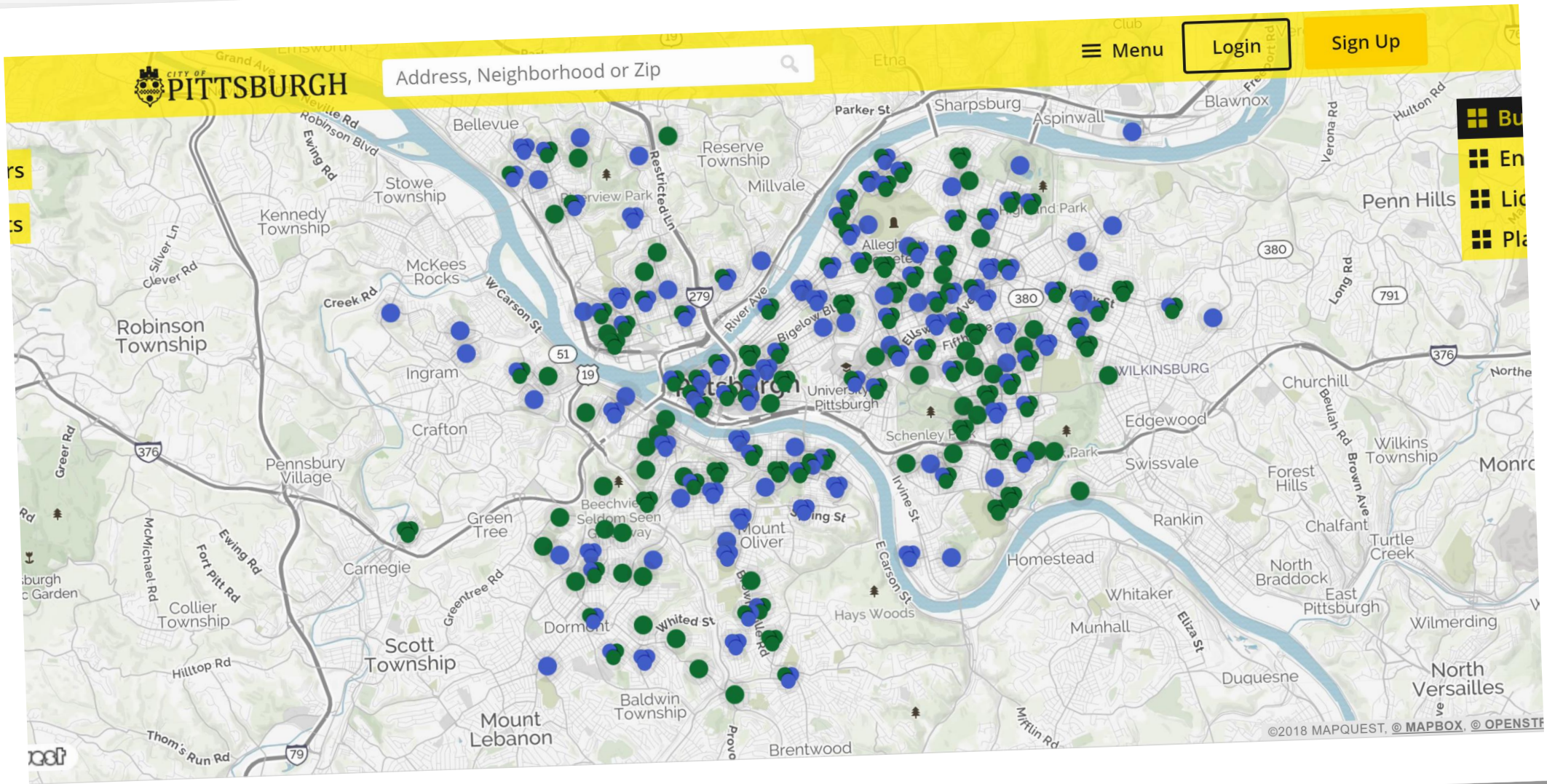
Atlas TI: \$700

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 ###

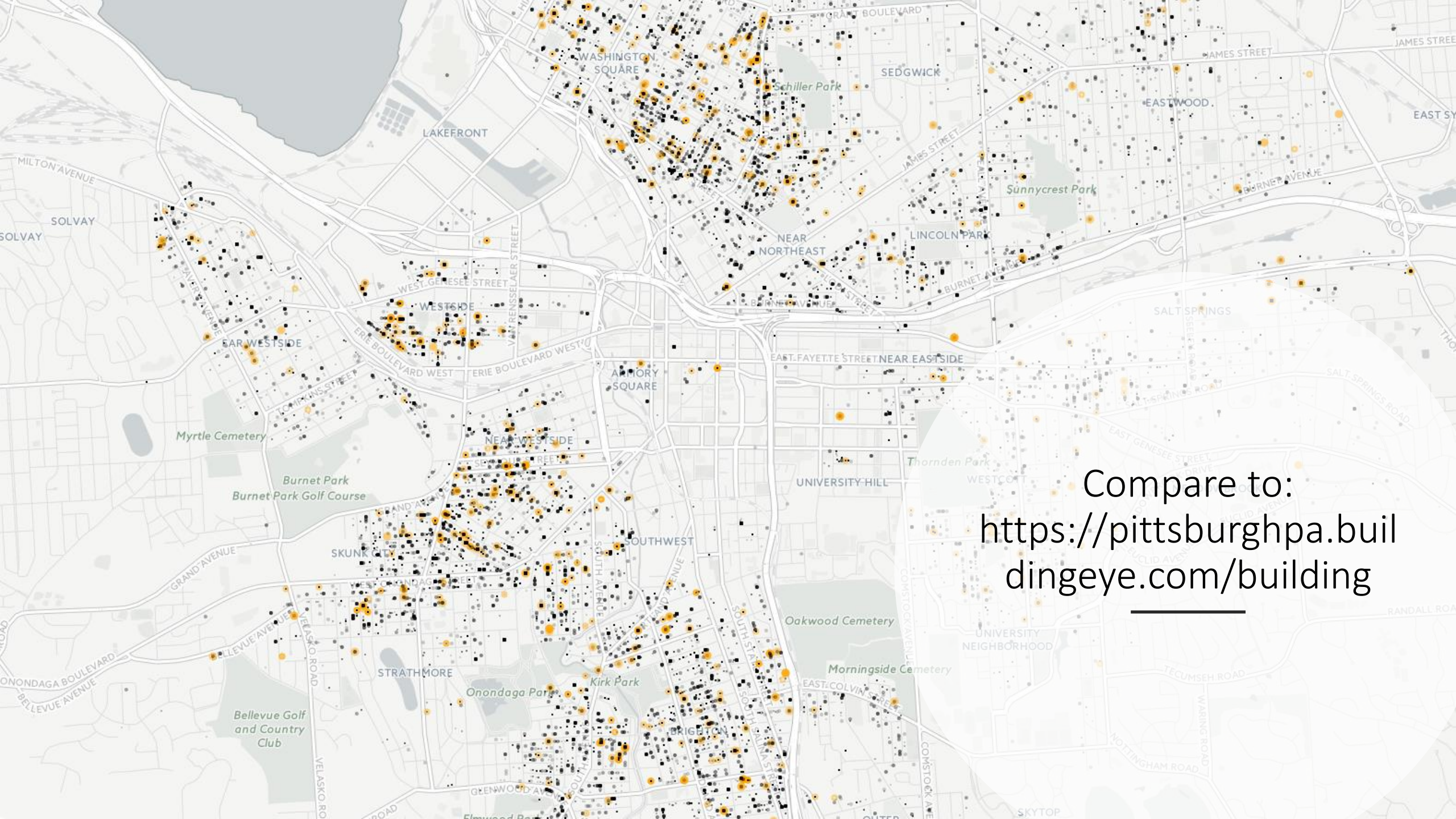
syr.map <- leaflet(data=lat.lon ) %>%
  addProviderTiles("CartoDB.Positron", tileOptions(minZoom=10, maxZoom=17)) %>%
  setView(lng=-76.13, lat=43.03, zoom=13) %>%
  setMaxBounds(lng1=-75, lat1=41, lng2=-77, lat2=45)

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>

THE DATA SCIENCE ECOSYSTEM

1. Data programming (R)
2. Packages (tools)
3. Open data / APIs
4. Data-driven documents
5. Collaboration platforms (GitHub)
6. Development Environments (R Studio)

DATA-DRIVEN DOCUMENTS

MARKDOWN



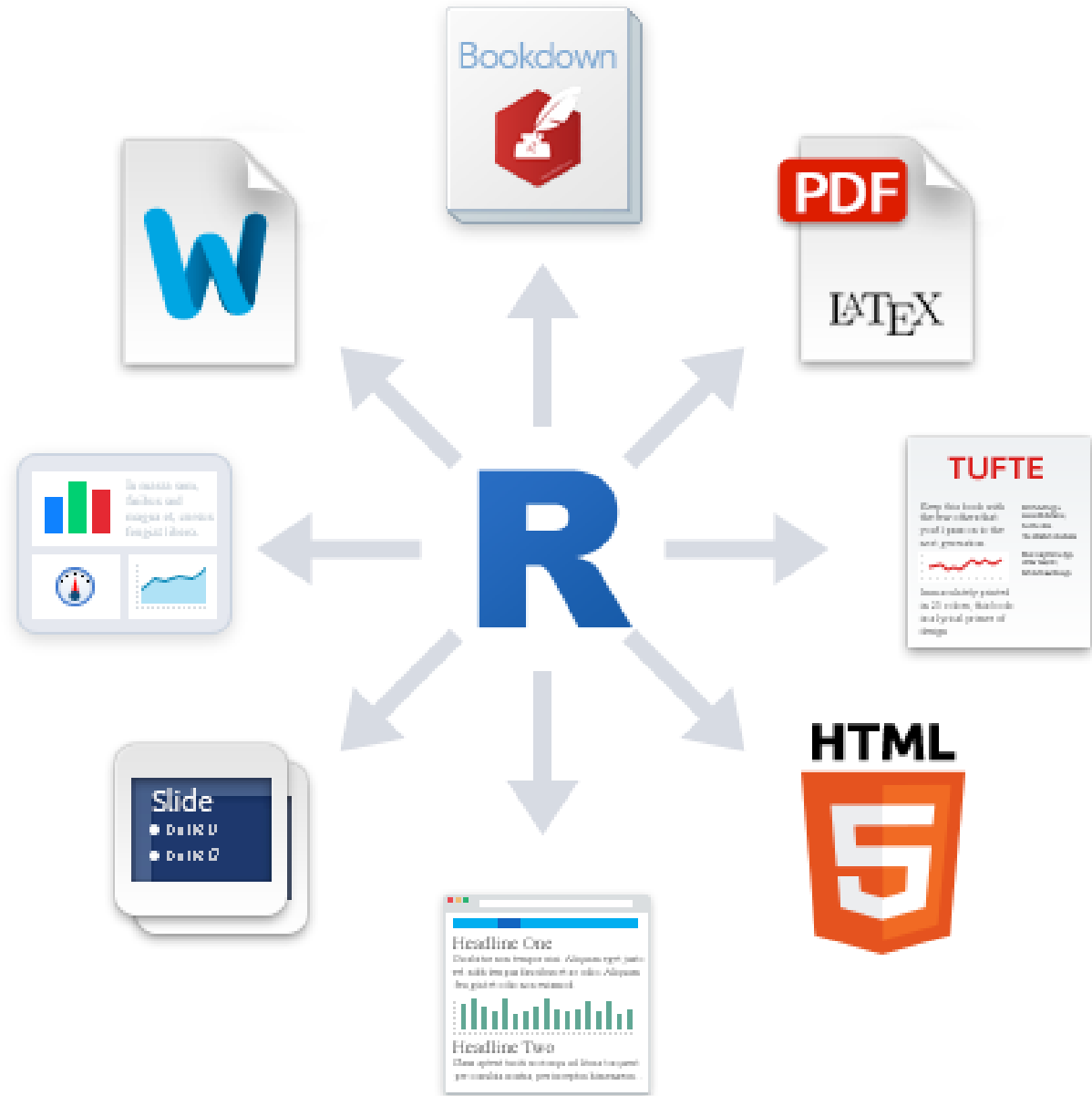
1. Reports [[html](#)] [[pdf](#)]
2. Books [[link](#)]
3. Presentations [[link](#)]
4. Tutorial [[link](#)]
5. Websites [[link](#)]
6. Dashboards [[link](#)]
7. Web Apps [[link](#)] [[link](#)] [[CDC](#)]



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

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

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




```
---  
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.

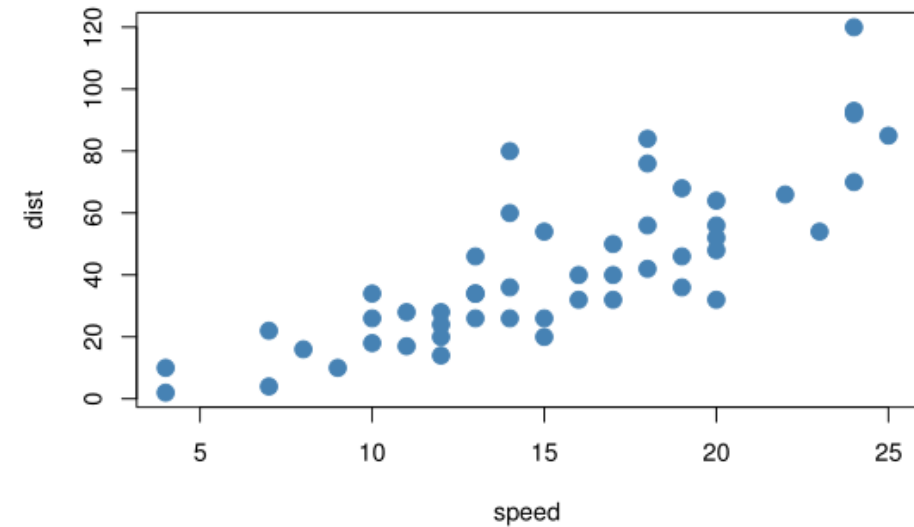
A sub-heading

Perhaps you want a list:

- Item 1
- Item 2

And here is some code:

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





RStudio

File Edit Code View Project Workspace Plots Tools Help

Go to file/function

Project: (None)

Workspace History

Load Save Import Dataset Clear All

Data

diamonds 53940 obs. of 10 variables

Values

aveSize 0.7979

clarity character [8]

p ggplot [8]

Functions

format.plot(plot, size)

```
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12            data=diamonds, color=clarity,
13            xlab="Carat", ylab="Price",
14            main="Diamond Pricing")
15
```

15:1 (Top Level) R Script

Console

```
> summary(diamonds)
  x           y           z
Min. : 0.000  Min. : 0.000  Min. : 0.000
1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910
Median : 5.700 Median : 5.710  Median : 3.530
Mean   : 5.731 Mean   : 5.735  Mean   : 3.539
3rd Qu.: 6.540 3rd Qu.: 6.540 3rd Qu.: 4.040
Max.   :10.740 Max.   :58.900  Max.   :31.800
> summary(diamonds$price)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  326   950   2401   3933   5324  18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+           data=diamonds, color=clarity,
+           xlab="Carat", ylab="Price",
+           main="Diamond Pricing")
>
> format.plot(p, size=24)
>
```

Diamond Pricing

Price

Carat

Clarity

- I1
- SI2
- SI1
- VS2
- VS1
- VVS2
- VVS1
- IF

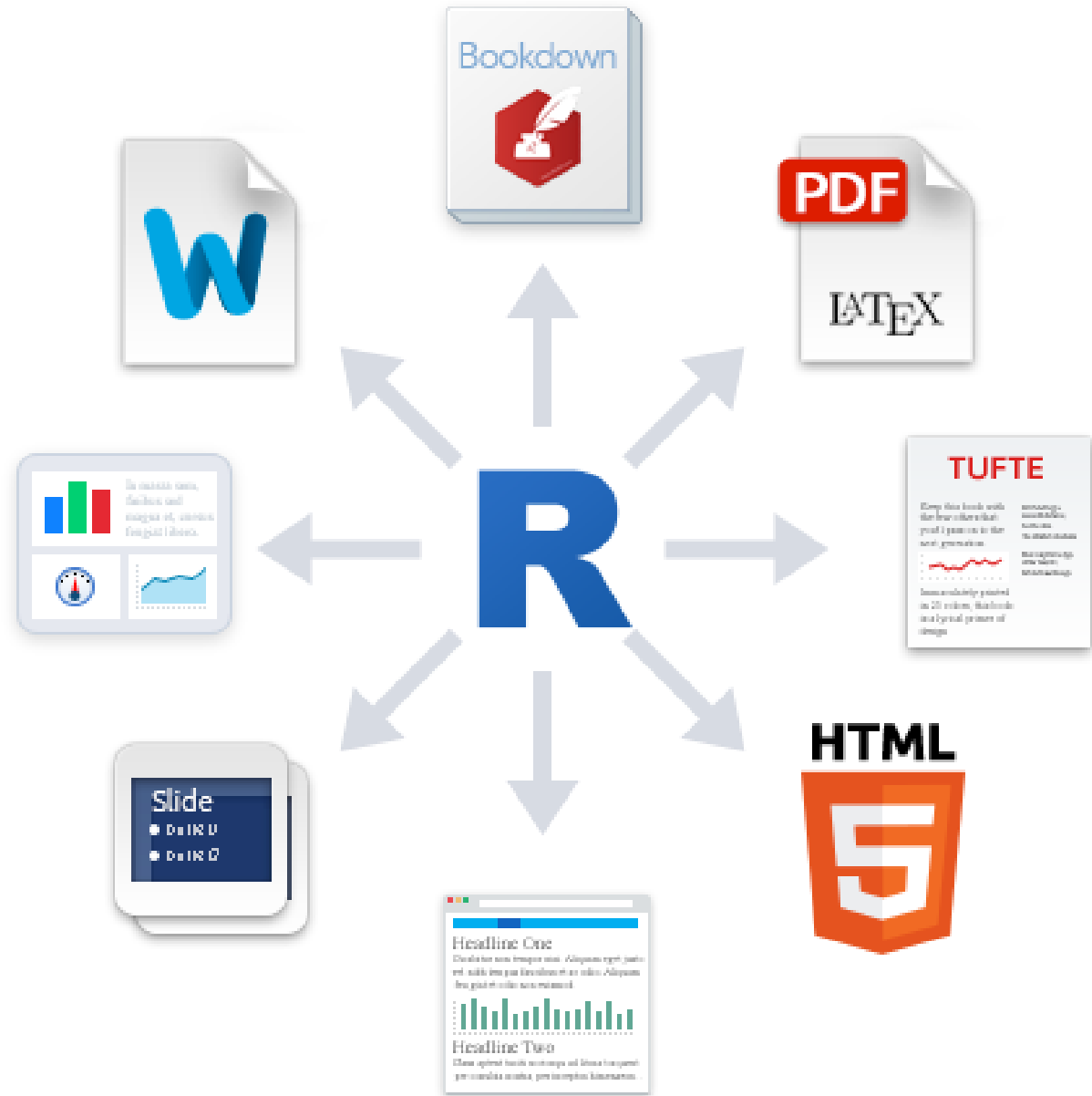
The figure is a scatter plot titled "Diamond Pricing". The x-axis is labeled "Carat" and ranges from 0.0 to 3.5 with major ticks every 0.5. The y-axis is labeled "Price" and ranges from 0 to 15000 with major ticks every 5000. The plot shows a positive correlation between carat weight and price. The data points are colored according to their clarity grade, as indicated by the legend on the right. The legend includes categories: I1 (red), SI2 (orange), SI1 (yellow), VS2 (green), VS1 (teal), VVS2 (blue), VVS1 (purple), and IF (pink). The plot shows that for a given carat weight, higher clarity grades (like IF and VVS) generally result in higher prices. The plot is displayed in the "Plots" pane of the RStudio interface.

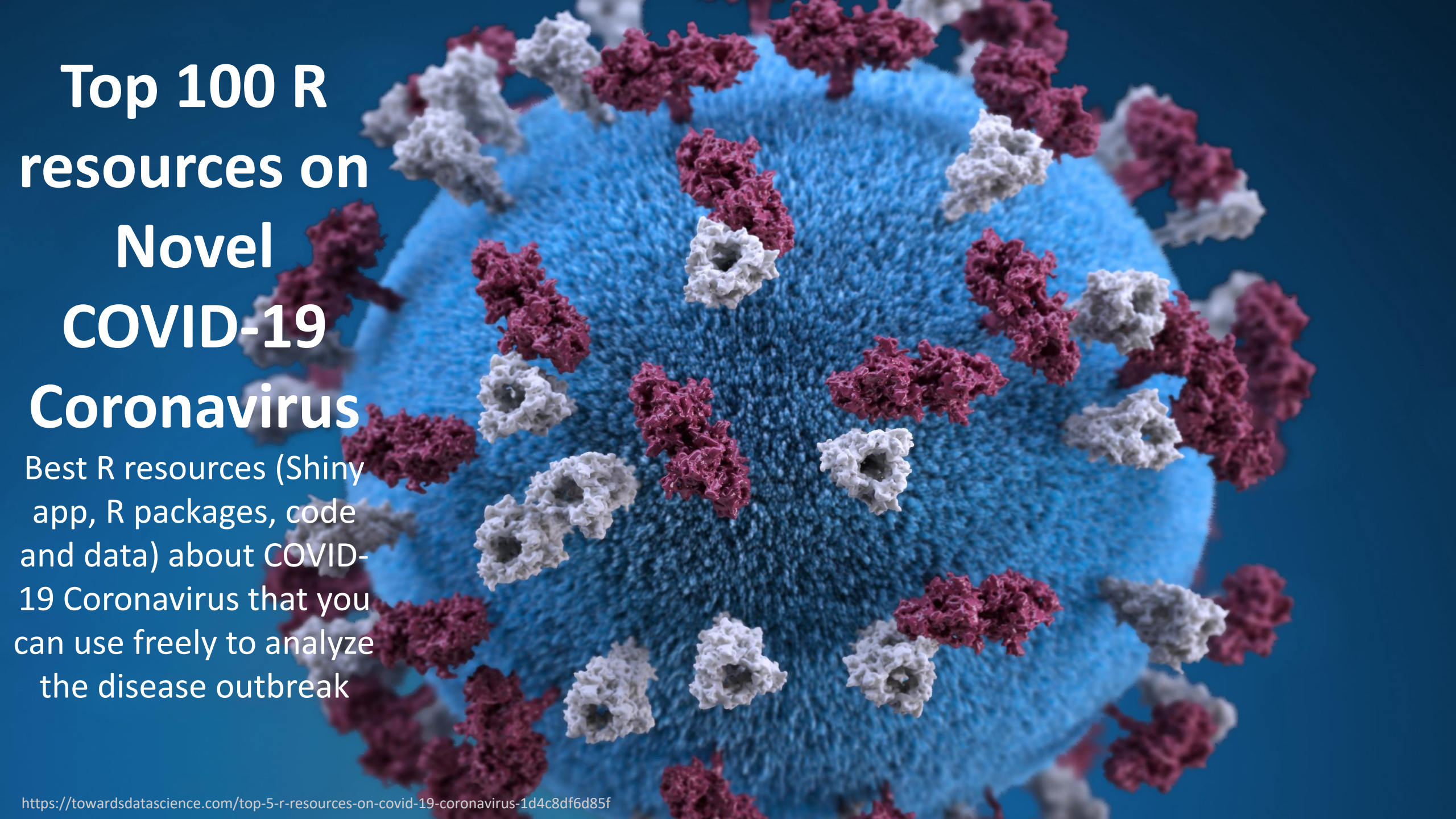


```
---  
title: "MARKDOWN DEMO"  
output: pdf_document  
---
```

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

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



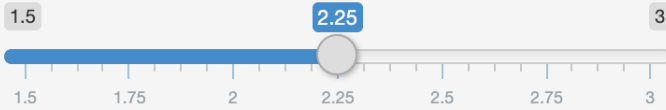


Top 100 R resources on Novel COVID-19 Coronavirus

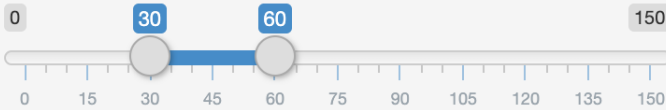
Best R resources (Shiny app, R packages, code and data) about COVID-19 Coronavirus that you can use freely to analyze the disease outbreak

DASHBOARD

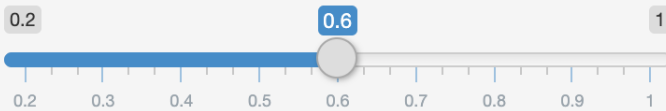
Choose a value for R_0



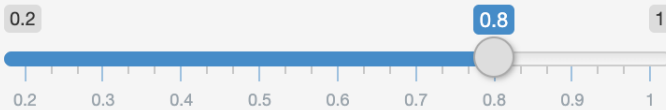
Choose the first 'social distance period' (sdp 1)



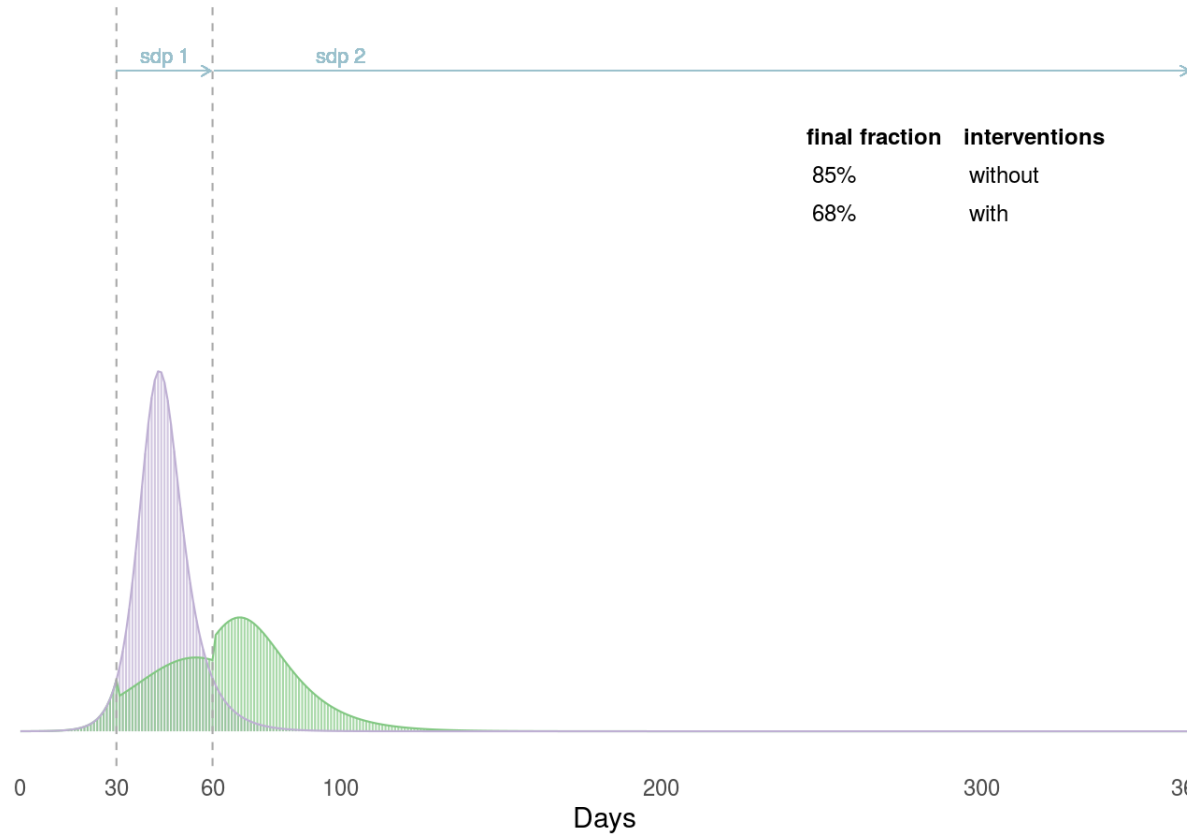
Reduction during first period (sdp 1)



Reduction after first period (sdp 2)



Daily new cases in % of the population



final fraction interventions

85% without

68% with

Interventions

— without

— with

sdp: social distance period

TOTAL COUNTRIES AFFECTED: 154

Confirmed
242,708



Yesterday: 214,910

Active 147,987
(61%)



Yesterday: 122,970

Recovered
84,854 (35%)



Yesterday: 83,207

Deaths 9,867
(4.1%)



Yesterday: 8,733

Countries showing in the
geospatial map:

Afghanistan, Albania, Algeria, ▾

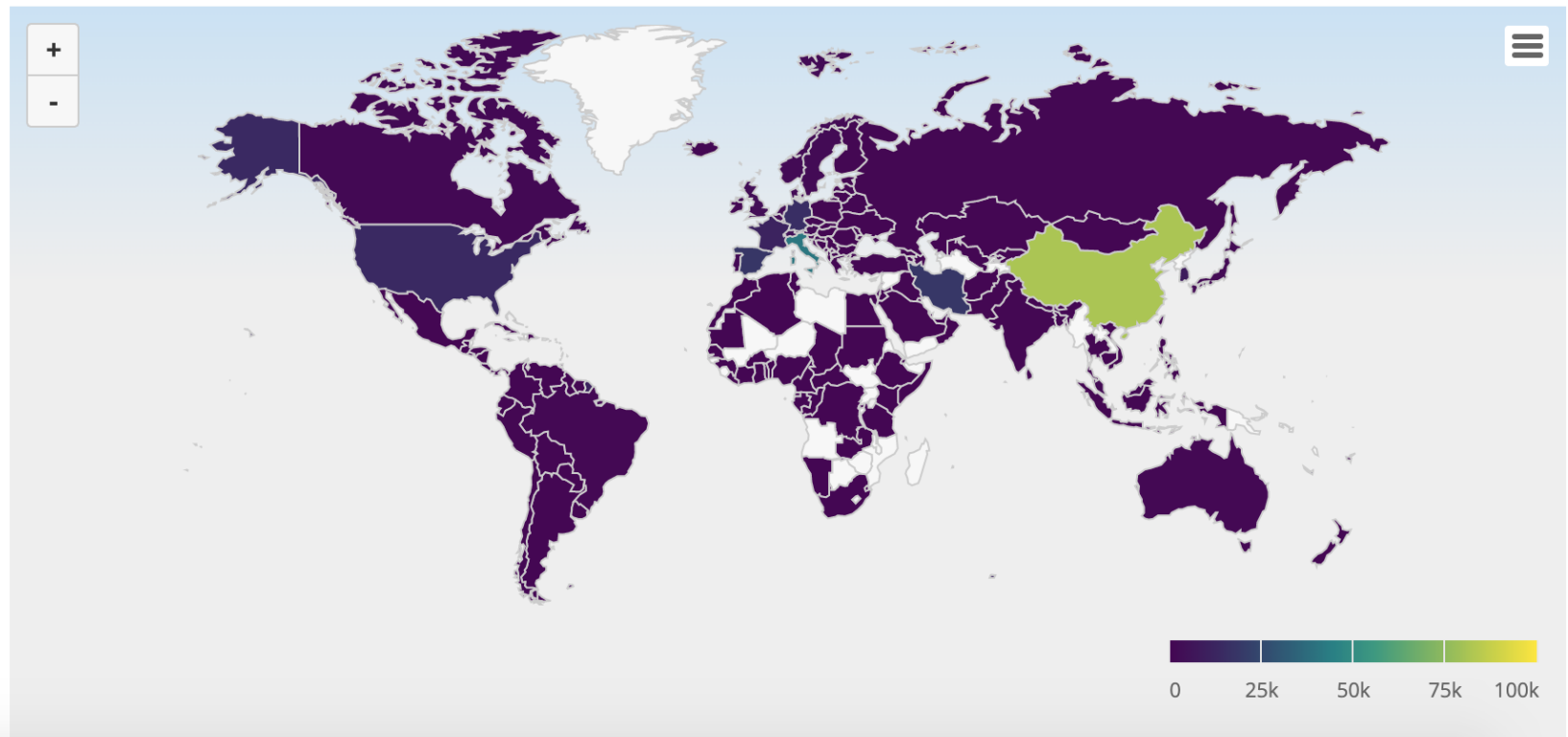
Specify the outcome to be shown in
map:

Total Cases

Recovered cases

Deaths

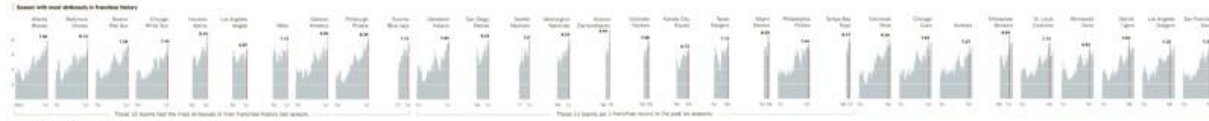
Active Cases





IN THE CLASSROOM

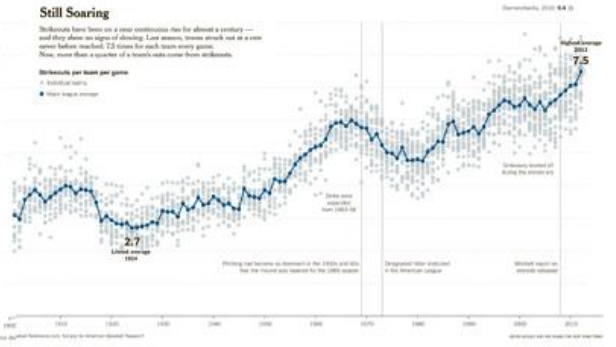
BASEBALL PREVIEW



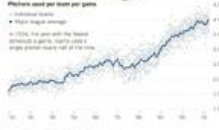
In Majors, a Swing and a Mystery

From First Sports Page
 During a game and get another one going. "I want to see a pitcher who can't throw a ball," says the author. "I want to see a pitcher who can't throw a ball." The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball.

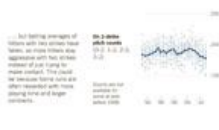
Still Soaring
 The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball.



Fresh Arms Keep Coming
 The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball.



Two Strikes, No Caution
 The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball.



Making Big Money, Finally, for Not Striking Out

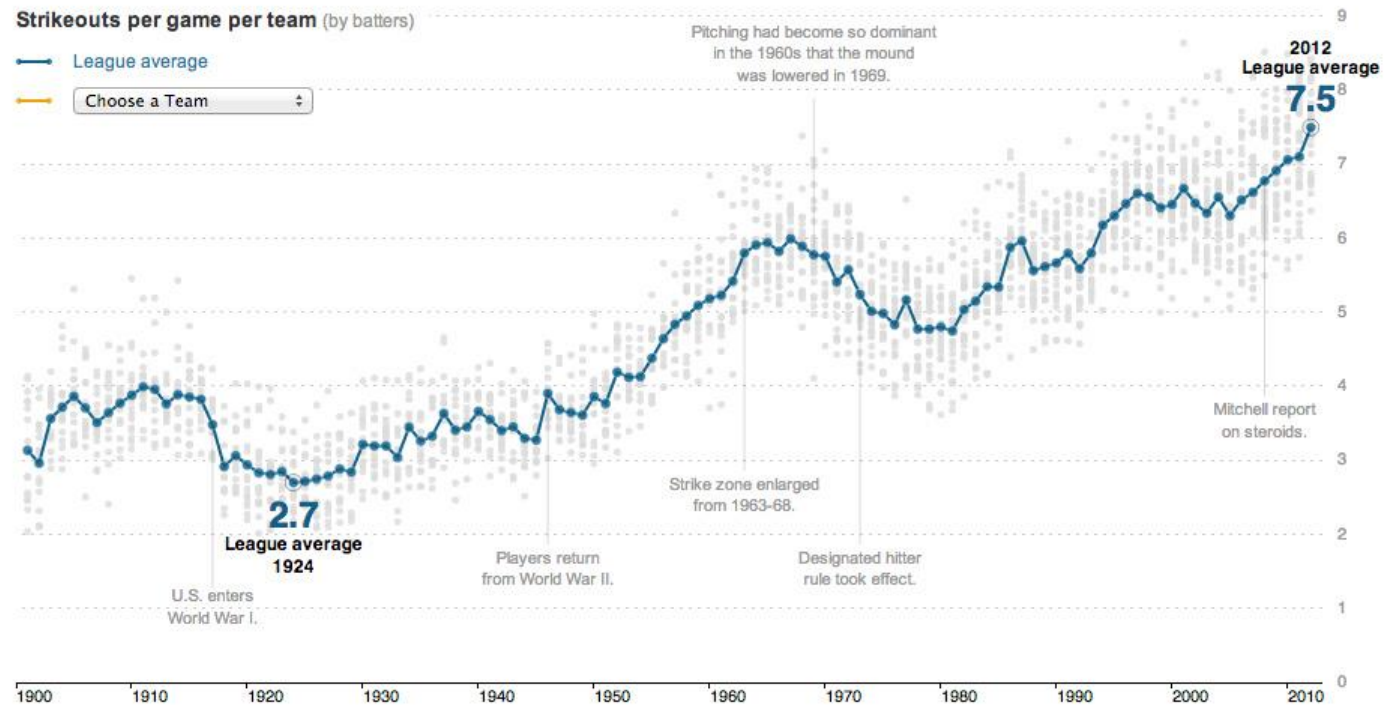
BY FRANK BRONF
 The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball. The author is talking about a pitcher who can't throw a ball.



Left: Springer, with one of the lowest strikeout rates in baseball, signed a three-year, \$12 million deal with the Mets. He had never had more than a one-pitch at-bat.

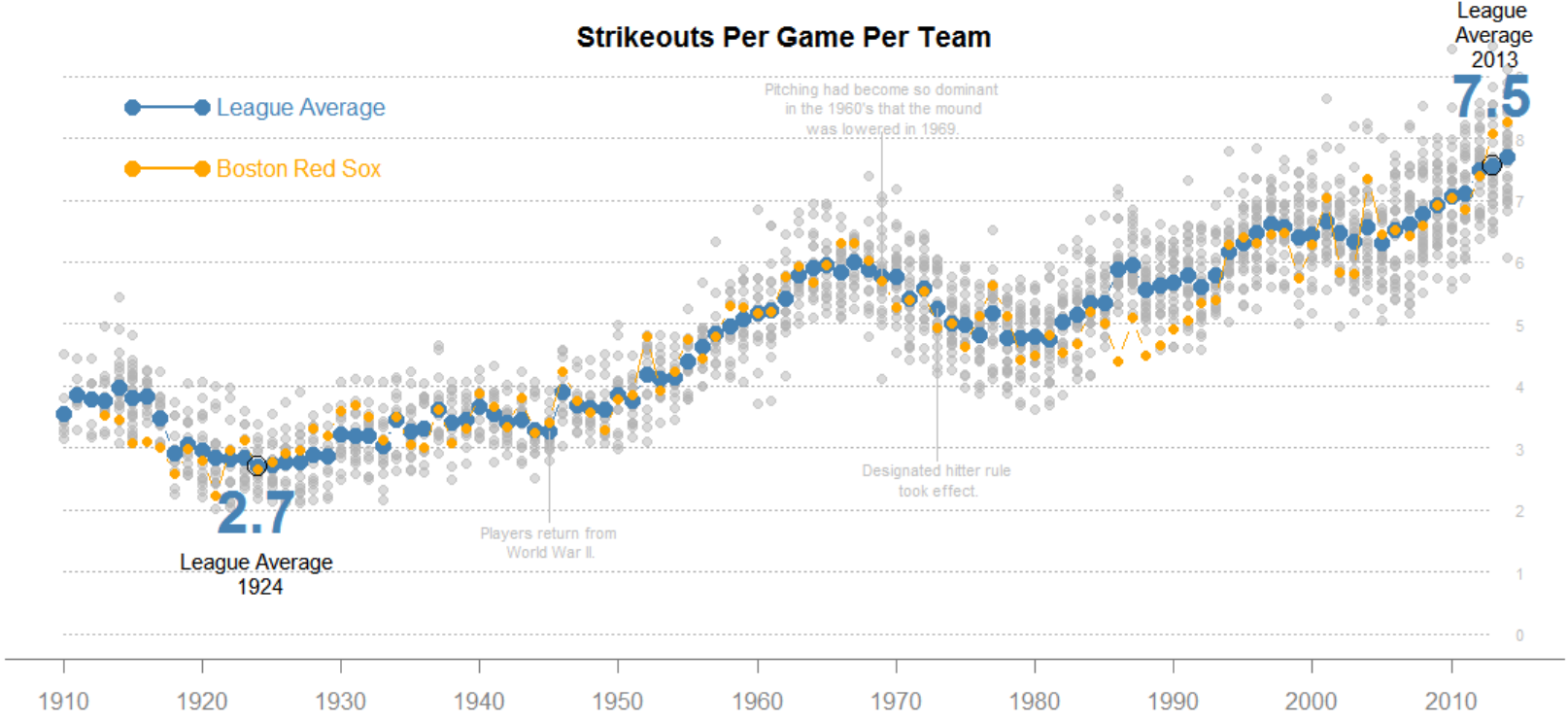
The acceptance of strikeouts as a necessary byproduct of production is now ingrained.

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

