

Welcome to AEM 2850!

Week 1

AEM 2850: R for Business Analytics

Cornell Dyson

Spring 2022

Acknowledgements: Andrew Heiss, Claus Wilke

Plan for today

Why take R for Business Analytics?

Summary of key class details

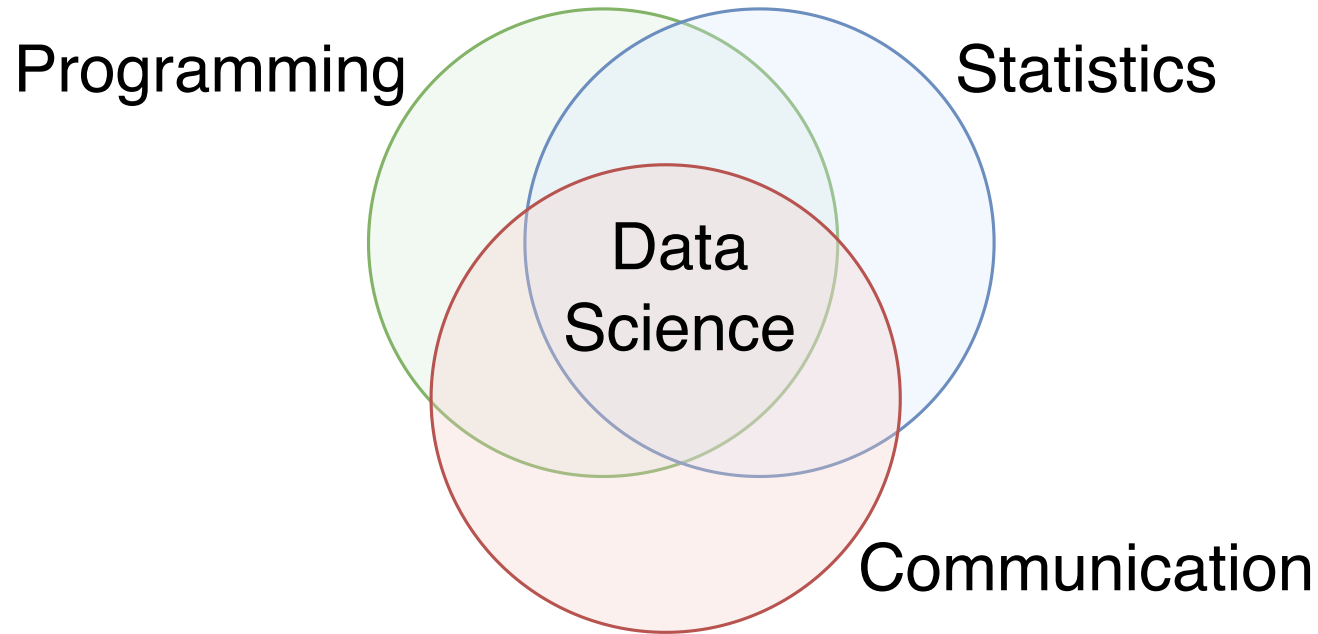
Facts, truth, and beauty

Data, truth, and beauty

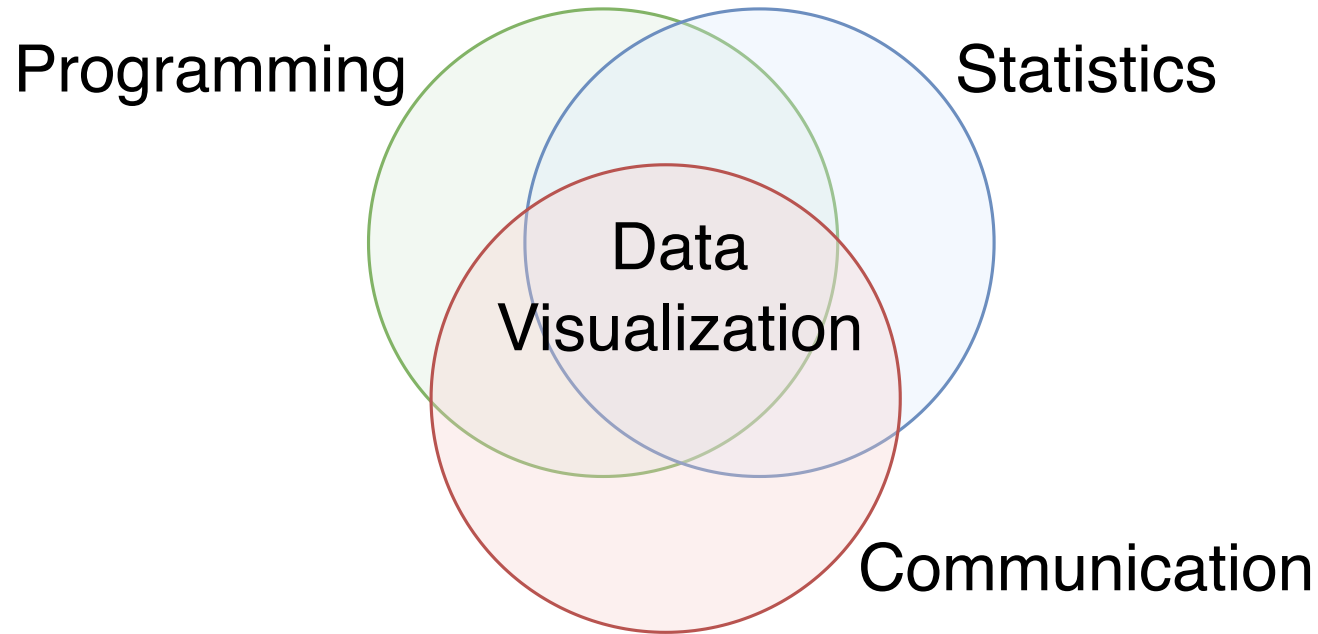
Beautiful visualizations

Why take R for Business Analytics?

Why take R for Business Analytics?



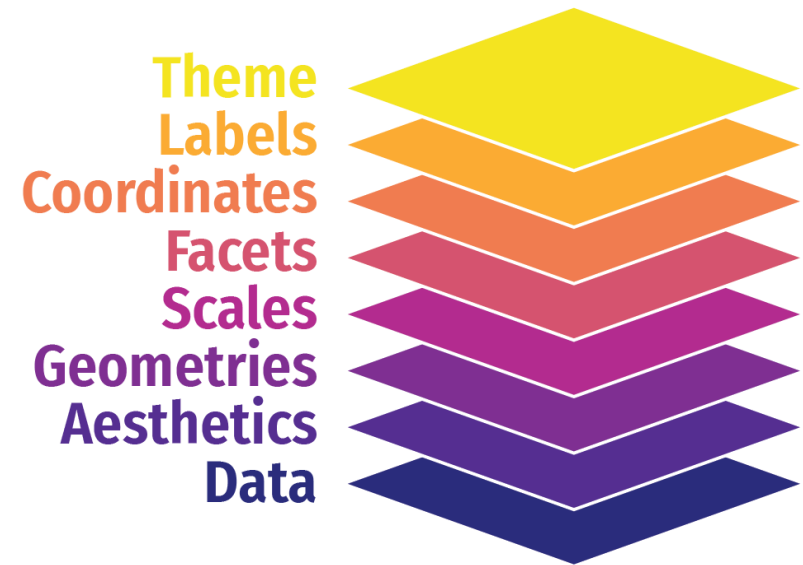
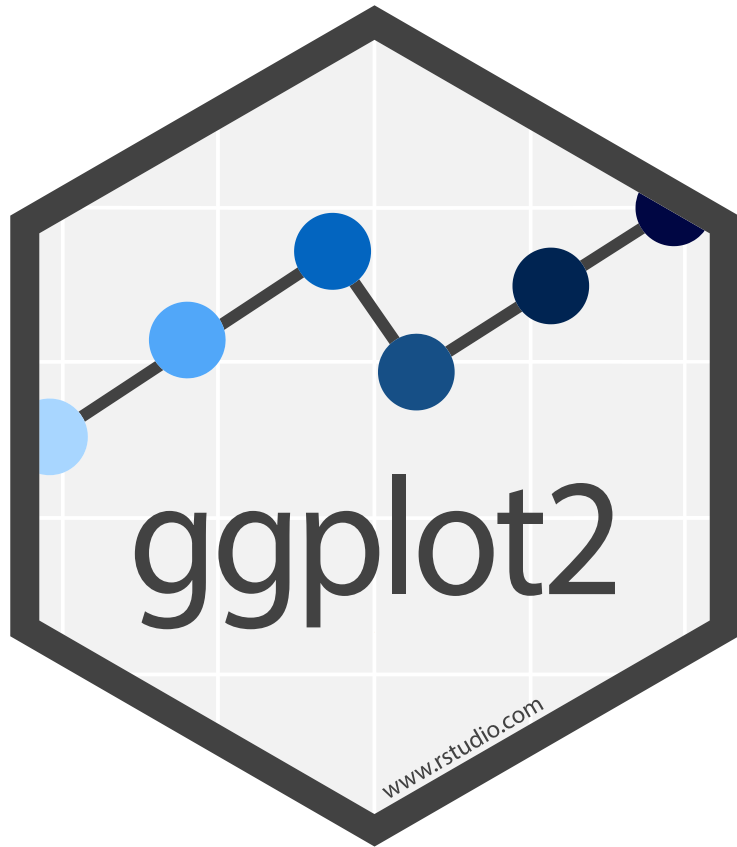
Why take R for Business Analytics?



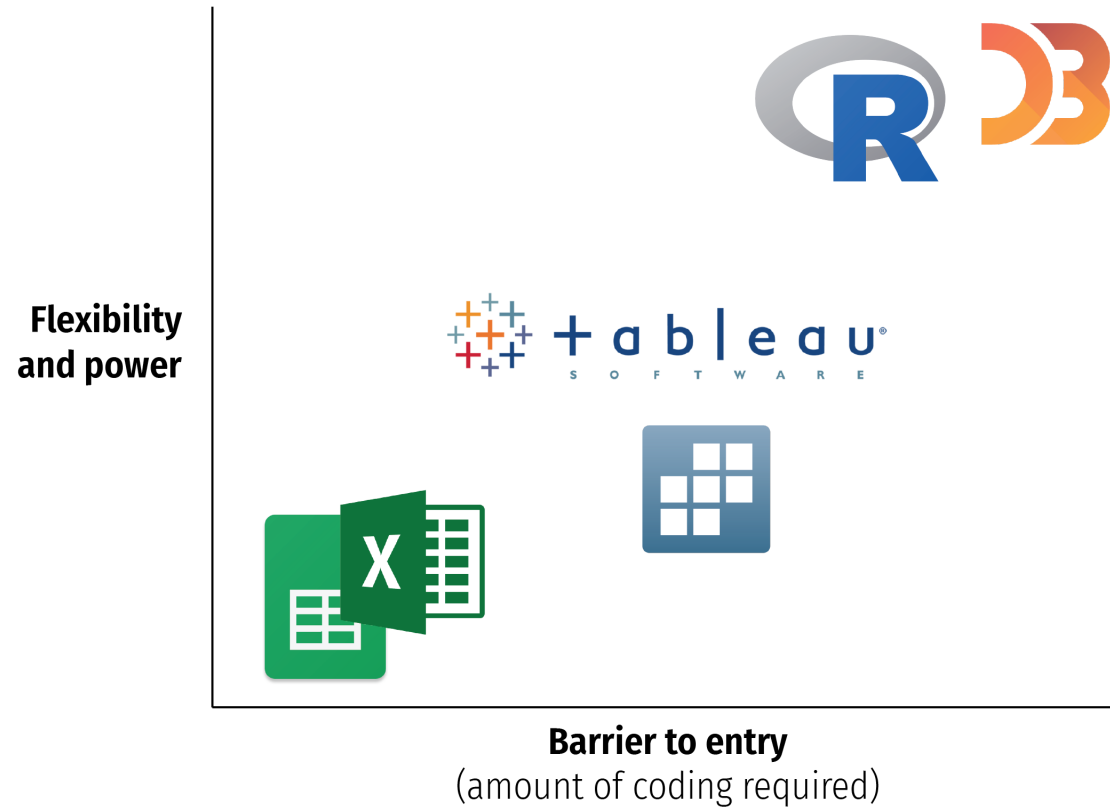
Why R for Business Analytics?



Why R for Data Visualization?



Why R for Data Visualization?



Why R for Life?

Practical tool that could help you get a job and then do said job

Open source

Huge community of users and package developers

Here are a few examples of other things you can do using R:

- Make slides like the ones you're looking at right now
- Build websites like [our course site](#)
- Write books like [R for Data Science](#)
- Make interactive web apps
- Much, much more

Class details

Preface

1. Your success in this class is important to me
2. This course is a work in progress
3. Get the semester off to a good start: **read the syllabus!**

A bit about me

A bit about our Graduate TA Hui Zhou

A bit about you

Do you have any programming experience? (zoom poll)

First course assignment will be to use **RMarkdown** to fill out a survey to tell us more about you

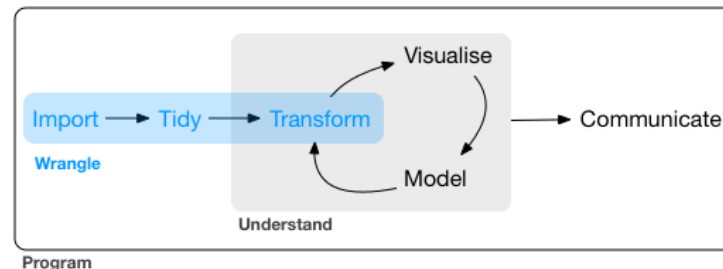
Course objectives

1. Develop basic proficiency in R programming
2. Understand data structures and manipulation
3. Describe effective techniques for data visualization and communication
4. Construct effective data visualizations
5. Utilize course concepts and tools for business applications

Plan for the semester

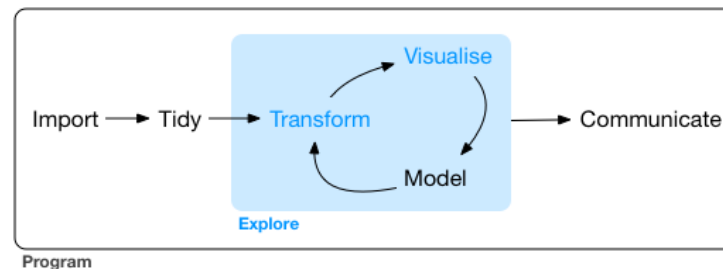
Programming Foundations

R, RStudio, Rmarkdown, and the tidyverse



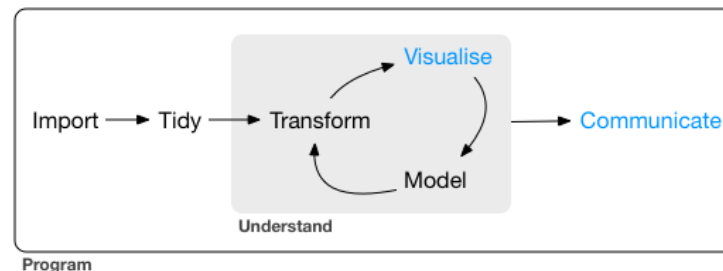
Data Visualization Foundations

the grammar of graphics, ggplot2



Special Topics (as time allows)

annotations, time, space, telling stories with data



Plan for each week

We will follow the same general process each week:

- Do readings listed on the before Tuesday (**example: Week 1**)
- **Tuesday:** come to class, where I will summarize material for that week's topic
- **Wednesday:** submit reflection on canvas by 11:59pm (starting with Week 2)
- **Thursday:** come to class, where we will work through hands-on examples
- Work on the lab before the next Tuesday's class, attending office hours as needed
- **The following Monday:** submit lab on canvas by 11:59pm (starting with Week 1)

Assignments

We will have frequent, short assignments to develop and cement new skills.

No prelims, no final exam.

- **Reflections** are short weekly writing assignments, intended as an easy way to get points and a nudge to engage with readings
- **Labs** are short weekly homework assignments that require you to practice programming
- **Projects** are intended to synthesize and reinforce individual skills, and to provide examples of their application to business and life more generally

Assignment	Points	Percent
Reflections (14 x 5 points each)	70	16%
Labs (14 x 10 points each)	140	32%
Mini project 1	50	11%
Mini project 2	50	11%
Final project	130	30%

Contacting us

Office hours:

- Mondays 1:30pm - 3:30pm: TA Hui Zhou at cornell.zoom.us/j/4786955504
- Tuesdays 11:00am - 12:00pm: Prof. Gerarden using [the class zoom link](#)
- Other times by appointment: Prof. Gerarden, at aem2850.youcanbook.me

Email:

You can also reach us by email. The best approach is to email both of us at the same time. You can do that with one click [here](#). Please read the syllabus for tips on how to make the most of email.

Course websites

Site for accessing course materials: (↓)

aem2850.toddgerarden.com

Site for submitting work: (↑)

canvas.cornell.edu/courses/38623

- viewing announcements
- viewing grades
- for convenience, you can also view and navigate the course site through canvas (Home, Syllabus)

Sucking

"The bad news is whenever you're learning a new tool, for a long time you're going to suck. It's going to be very frustrating.

But, the good news is that that is typical, it's something that happens to everyone, and it's only temporary.

Unfortunately, there is no way to go from knowing nothing about a subject to knowing something about a subject and being an expert in it without going through a period of great frustration and much suckiness.

But remember, when you're getting frustrated, that's a good thing, that's temporary, keep pushing through, and in time [it] will become second nature."

Hadley Wickham, author of `ggplot2`, *R for Data Science*, and much more

I *promise* you can succeed in this class. Don't hesitate to get help from me, TAs, office hours, and your peers.

Questions about the class?

Facts, truth, and beauty

What is truth?

Core principles of the universe?

Underlying trends in society?

Something transcendental?

Reality?

How do we find truth?

Science!



Neil deGrasse Tyson 
@neiltyson



The good thing about Science is that it's true whether or not you believe in it.

10:41 AM · Jun 14, 2013 · [TweetDeck](#)

14.3K Retweets **8.3K** Likes



761



14.3K



8.3K



But wait!

Beware of scientism!

"... promotion of science as the best or only objective means by which society should determine normative and epistemological values"

Science is not the only way

Art

Music

Literature

Religion

Nature

In chat: Name one thing that is not factual...

...but still reveals truth

the office

SOUR

What does this have to do with AEM 2850?

Truth does not require science or facts

Facts alone do not necessarily reveal truth

Truth comes from aesthetic combination of **content** and **form**

There is no single ideal combination of **content** and **form** for all **audiences**

Keep this in mind as a guiding principle for analyzing, visualizing, and communicating data

Data, truth, and beauty

Just show me the data!

```
head(my_data, 10)
```

```
## # A tibble: 10 × 2
##       x     y
##   <dbl> <dbl>
## 1  55.4  97.2
## 2  51.5  96.0
## 3  46.2  94.5
## 4  42.8  91.4
## 5  40.8  88.3
## 6  38.7  84.9
## 7  35.6  79.9
## 8  33.1  77.6
## 9  29.0  74.5
## 10 26.2  71.4
```

```
mean(my_data$x)
```

```
## [1] 54.26327
```

```
mean(my_data$y)
```

```
## [1] 47.83225
```

```
cor(my_data$x, my_data$y)
```

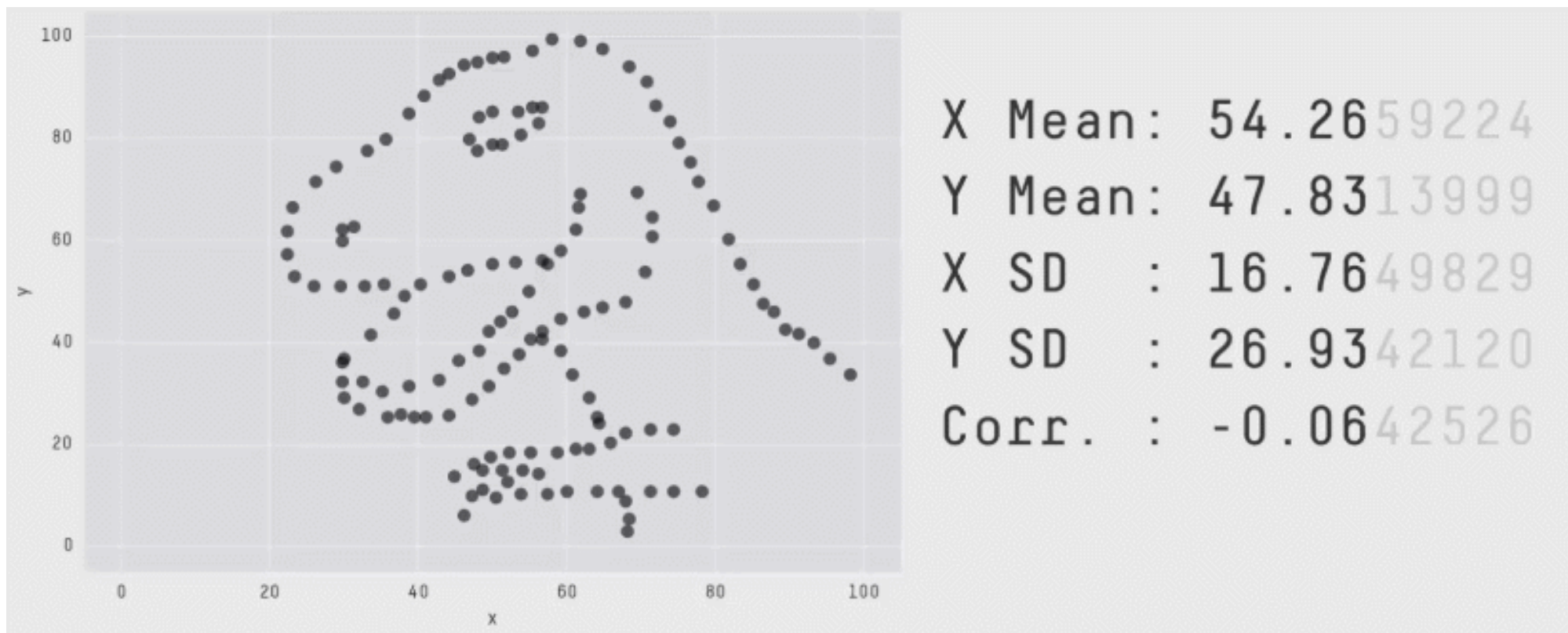
```
## [1] -0.06447185
```

Seems reasonable

Seems reasonable

No correlation

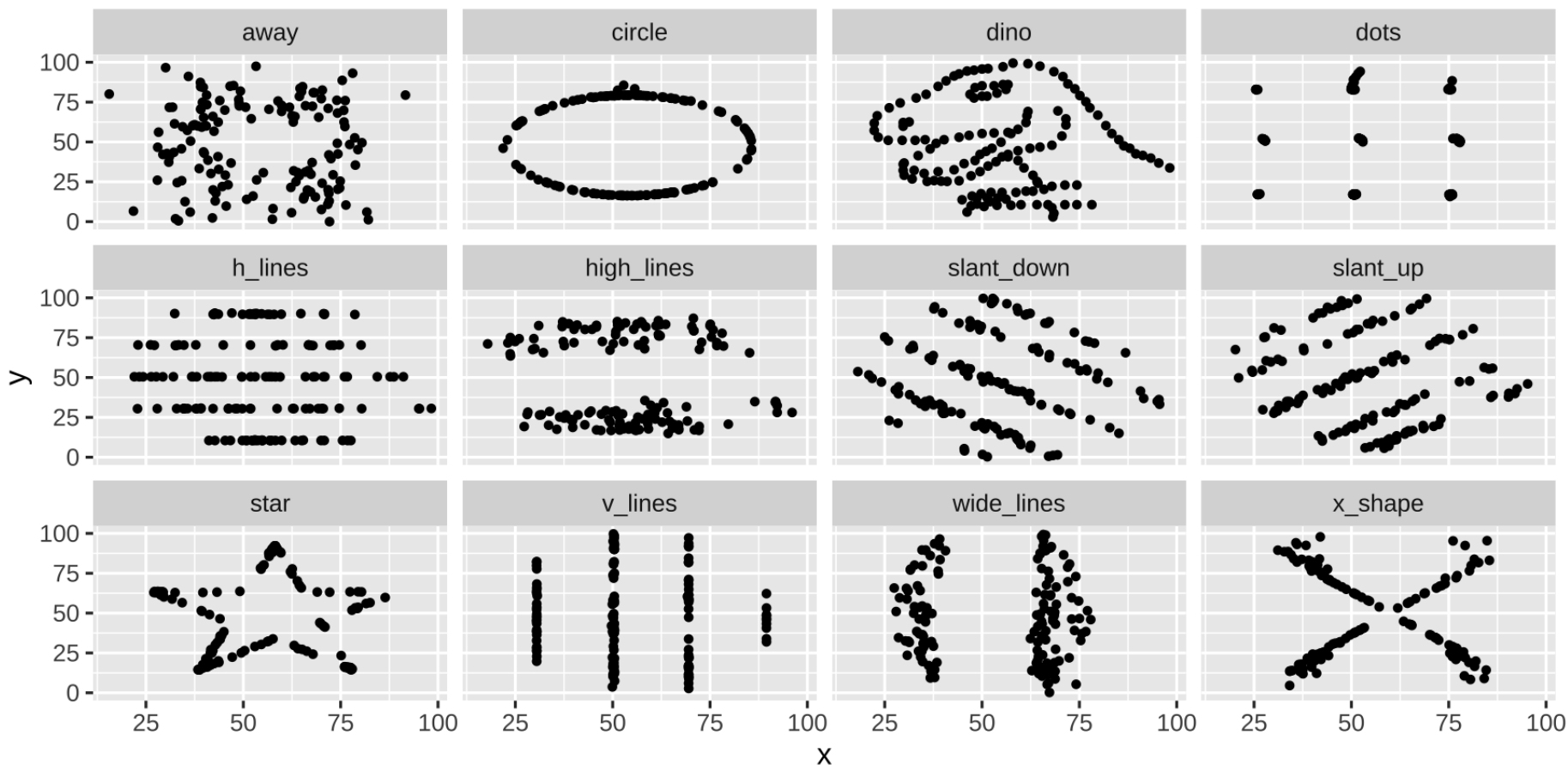
Oh no!



The Datasaurus Dozen

Raw data is not enough

Each of these has the same mean, standard deviation, variance, and correlation



Beautiful visualizations

What makes a great visualization?

Truthful

Functional

Beautiful

Insightful

Enlightening

Alberto Cairo, *The Truthful Art*

What makes a great visualization?

"Graphical excellence is the **well-designed presentation of interesting data**—a matter of substance, of statistics, and of design ... [It] consists of complex ideas communicated with clarity, precision, and efficiency. ... [It] is that which **gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space** ... [It] is nearly always multivariate ... And graphical excellence requires **telling the truth about the data.**"

Edward Tufte, *The Visual Display of Quantitative Information*, p. 51

What makes a great visualization?

Good aesthetics

No substantive issues

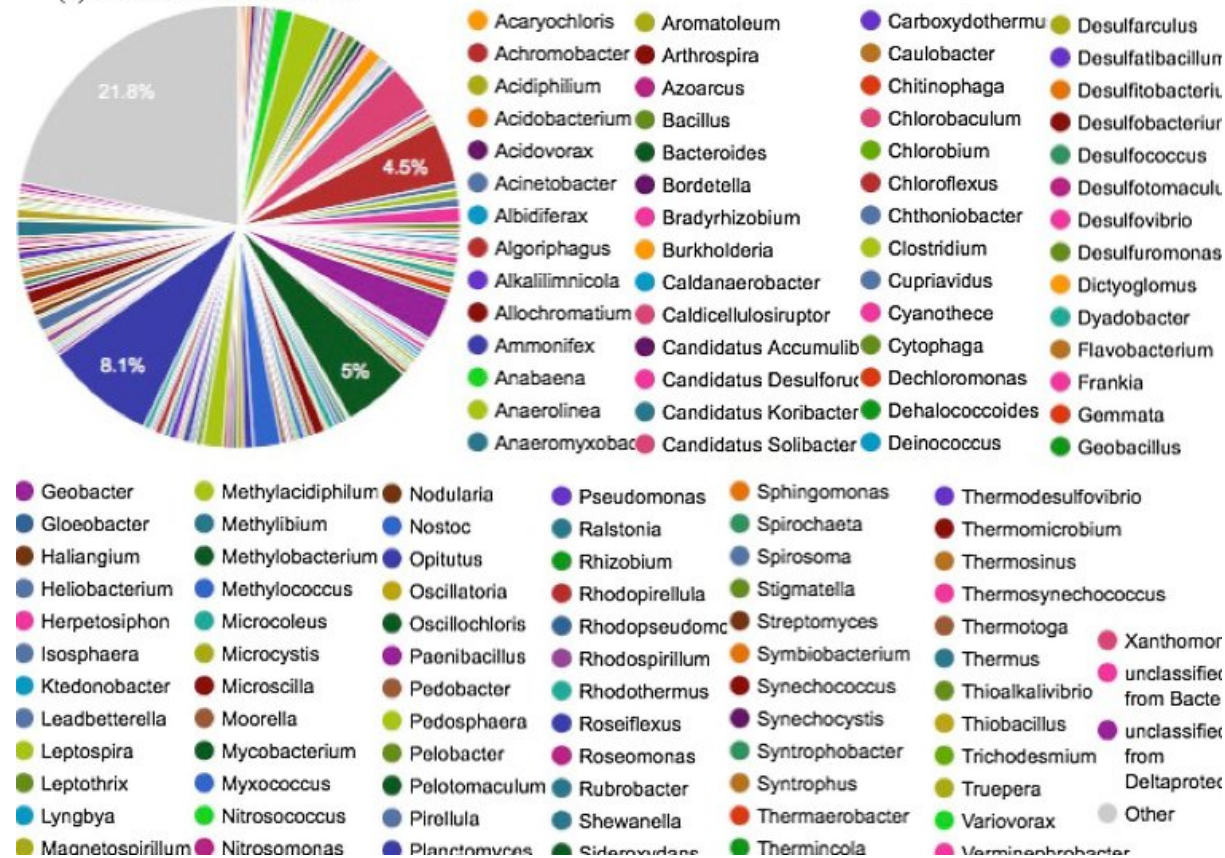
No perceptual issues

Honesty + good judgment

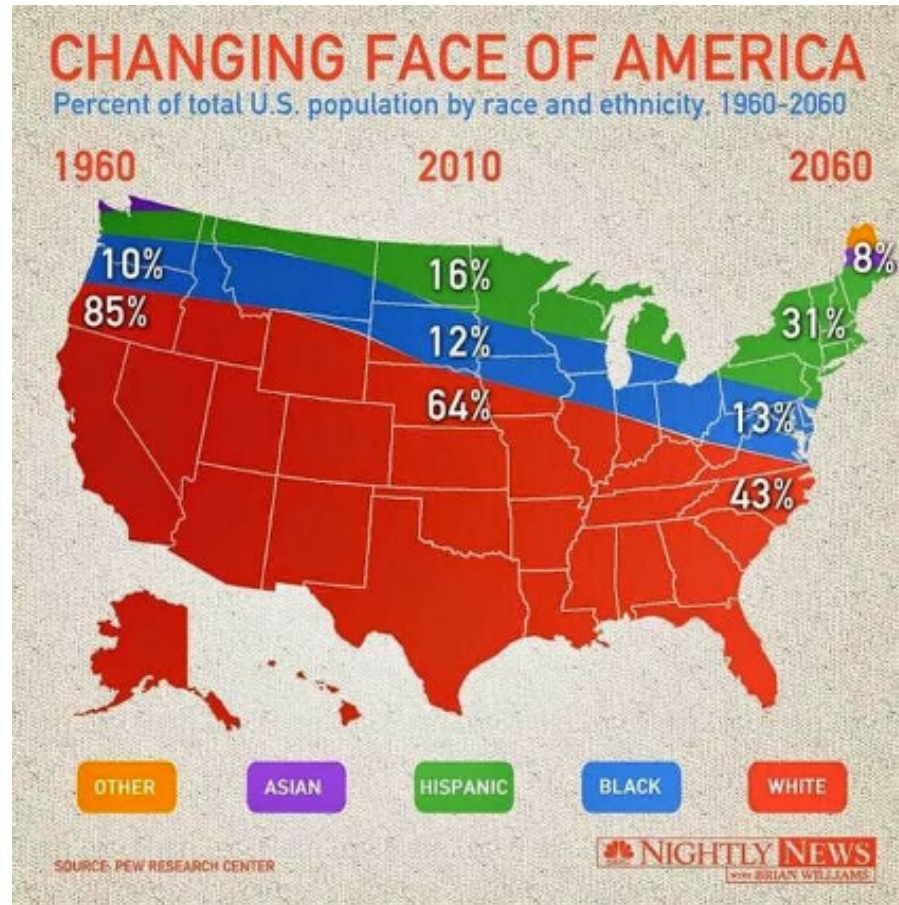
Kieran Healy, *Data Visualization: A Practical Introduction*

What's wrong?

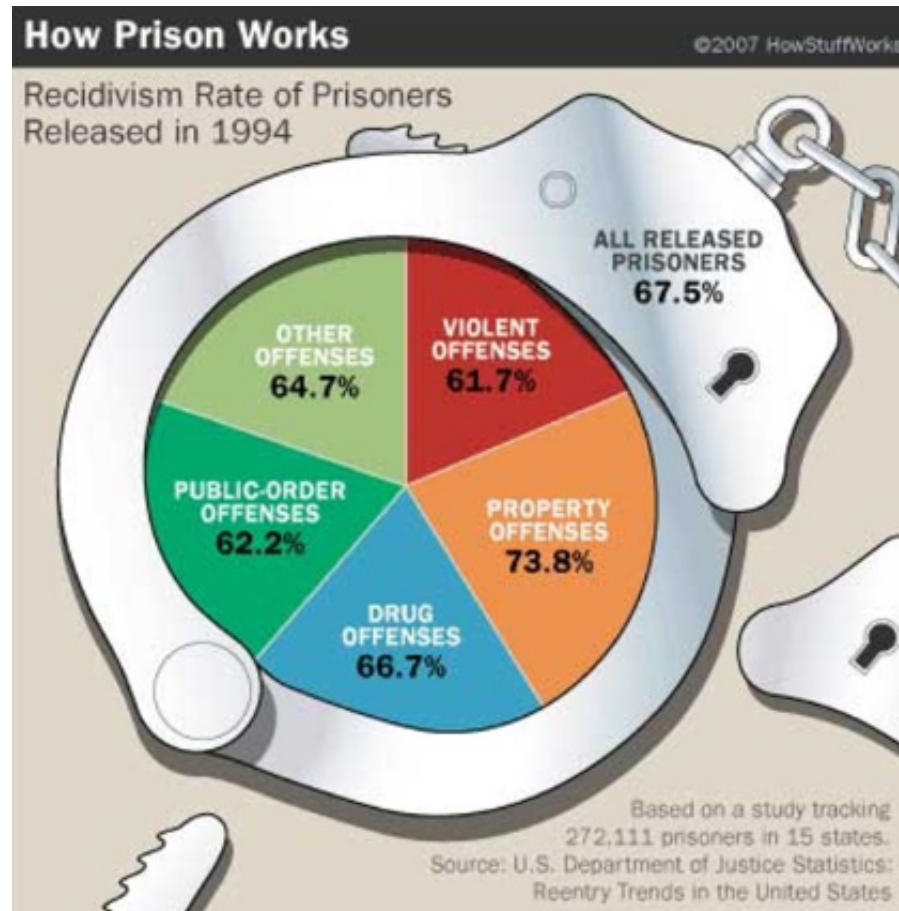
(f) Distribution of Genus



What's wrong?



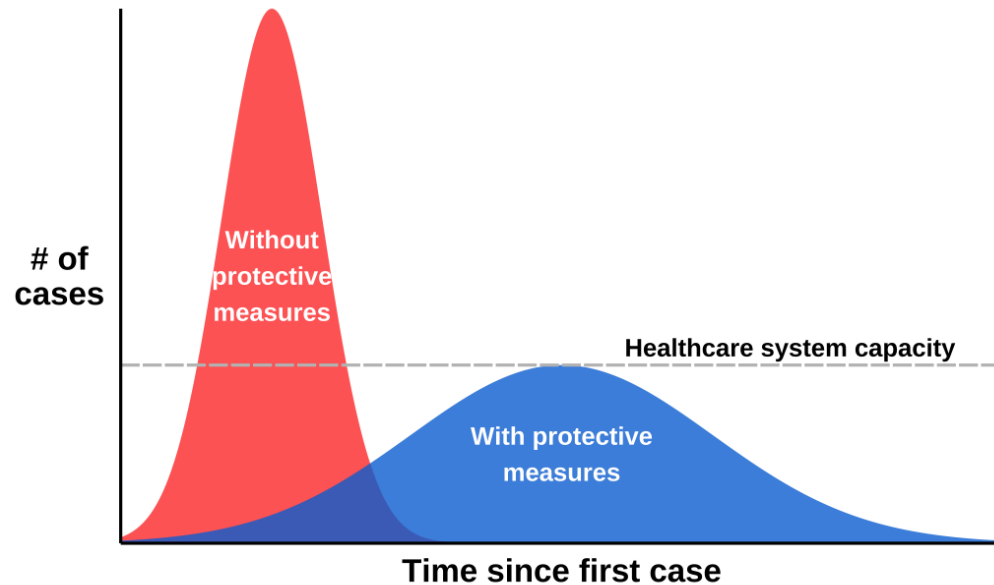
What's wrong?



What's right?

Flatten the curve!

Slow down community spread by social distancing



Adapted from the CDC and The Economist
Visit flattenthecurve.com

Carl T. Bergstrom @CT_Bergstrom · Mar 6

3. There is a lot of complicated epidemiological modeling behind this idea, but this graphic strips all of that away, and discards irrelevant details to provide a straightforward story that people find easy to grasp at a glance.

It **simplifies** and **highlights** what matters.

6 198 1.8K

[Show replies](#)

Carl T. Bergstrom @CT_Bergstrom · Mar 6

4. I've seldom seen a piece of sci-comm matter so much. We have an opportunity to flatten the [#COVID19](#) [#coronavirus](#) epidemic curve by aggressive social distancing and other measures.

But people don't understand what the point is, if the virus is going to circulate broadly.

8 313 2K

[Show replies](#)

Carl T. Bergstrom @CT_Bergstrom · Mar 6

5. This graph provides the answer, powerfully and concisely.

And because of that, it has exploded across twitter and other media. I've used it myself a number of times. This graph is changing minds, and by changing minds, it is saving lives.

6 196 1.5K

Thread by Carl T. Bergstrom

Plan for the rest of this week

Office hours:

- Tuesdays 11:00am - 12:00pm: Prof. Gerarden using [the class zoom link](#)
- Other times by appointment: Prof. Gerarden, at aem2850.youcanbook.me

Thursday:

- Intro to [R](#), [RStudio](#), and [RMarkdown](#)
- You will need your computer for coding exercises
- Keep an eye out for a canvas announcement with instructions for getting set up on [RStudio.cloud](#)