Introduction to the Grammar of Graphics with R

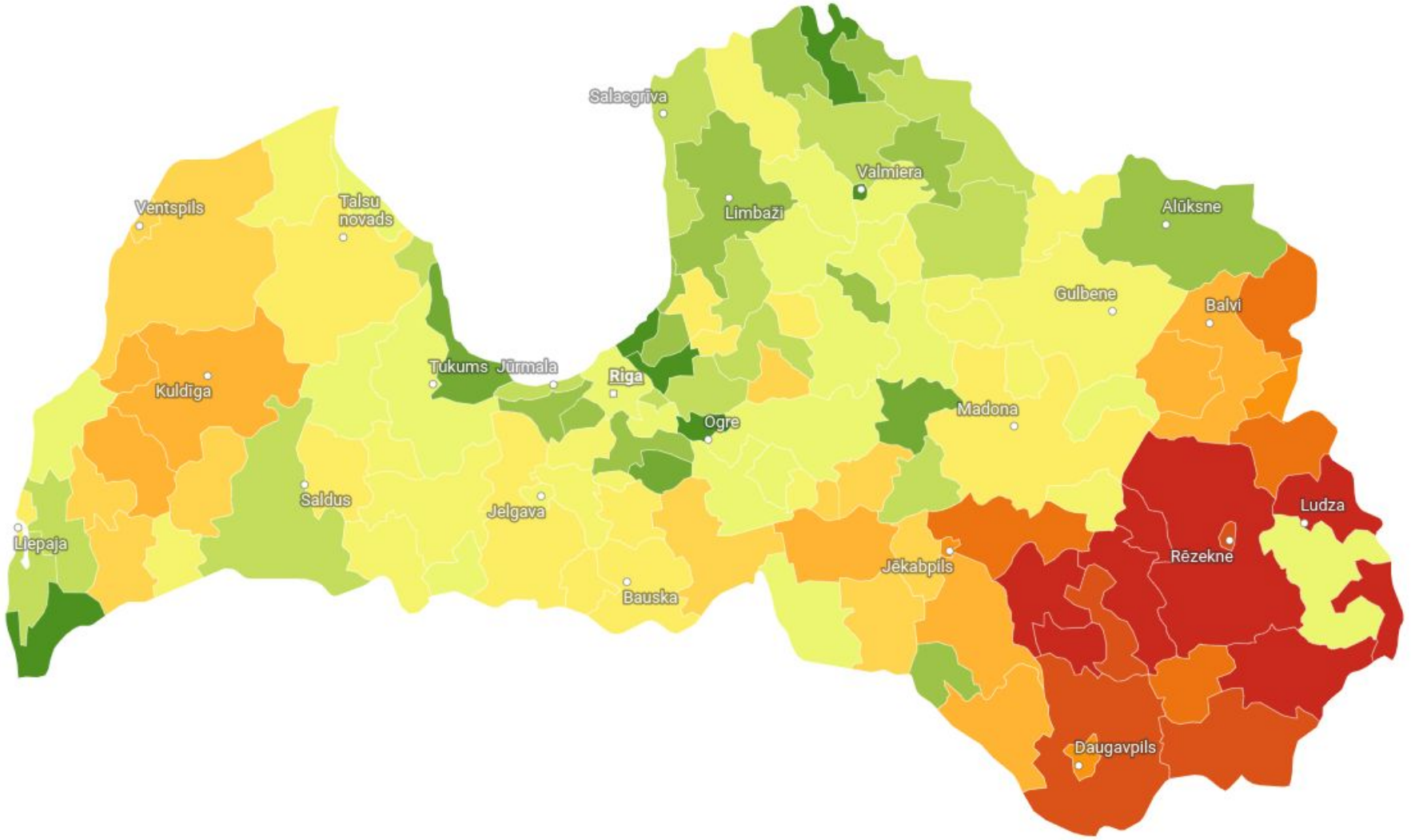
Sergio Uribe, PhD, DDS
Leading Researcher, Bioinformatics Lab, RSU, & BBCE, RTU, *Latvia*
Assoc Prof, Dept of Conservative Dent and Oral Health, RSU, Latvia
Associate Professor, Universidad Austral de Chile, *Chile*

1. **Motivation**
2. The grammar of a plot
3. Essentials data transformations
4. Polishing the plot
5. Learn more

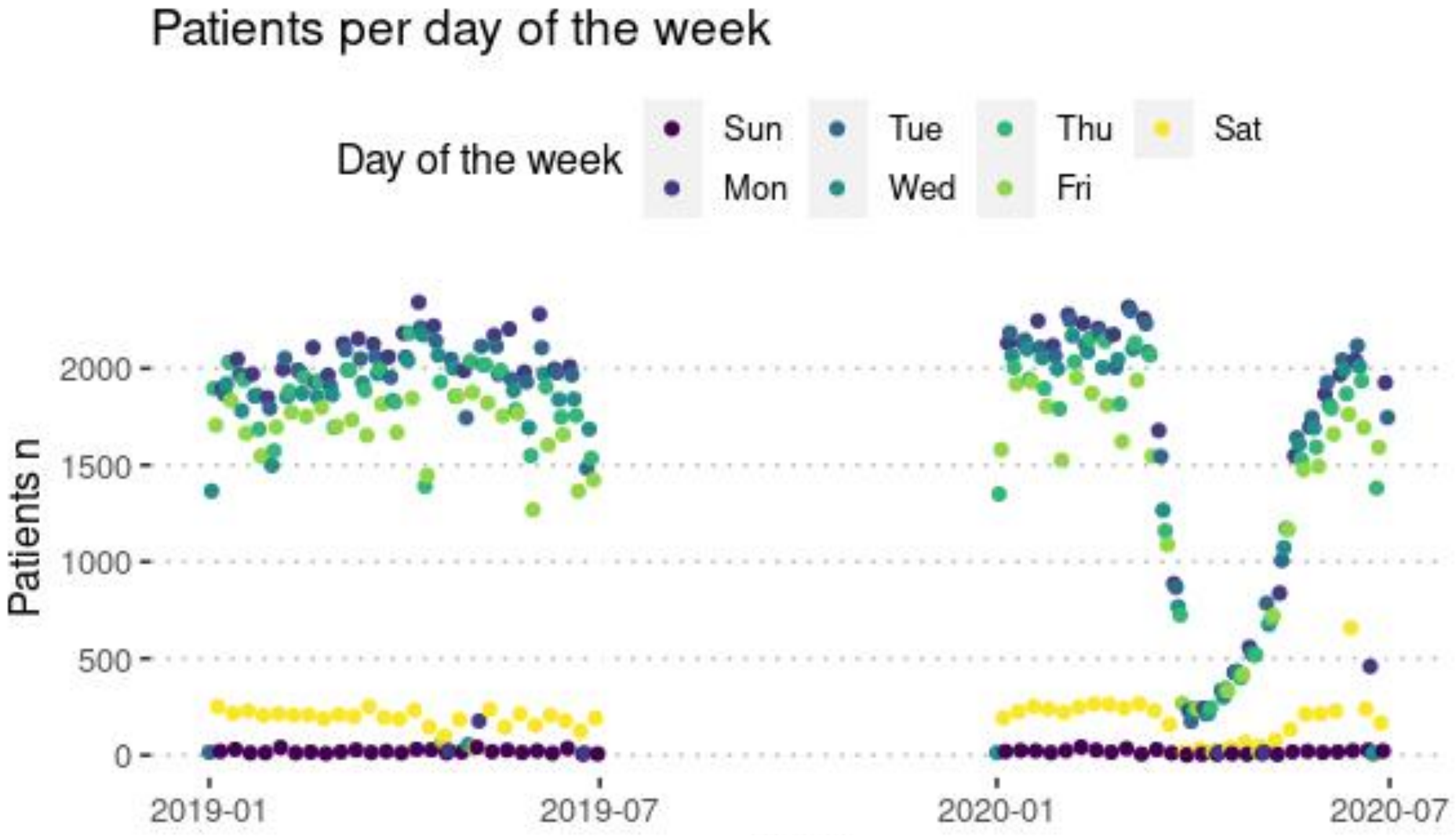

Dr. Sanjana Curtis!
@sanjanacurtis
50% of a phd is making plots
3:01 AM · Oct 17, 2020 · Twitter for iPhone
56 Retweets
39 Quote Tweets
1,235 Likes
Chelsea Parlett-Pelleriti @ChelseaParlett · Oct 17, 2020
Replying to @sanjanacurtis
and the other half is remaking them
1
39

Pret Covid-19 vakcinēto iedzīvotāju īpatsvars Latvijas novados (dalījumā pirms administratīvi teritoriālās reformas) 01.10.2021.

28303234363840424446485O



Impact of COVID19 on the provision of dental care in the national health service in Latvia



Why to create plots?

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

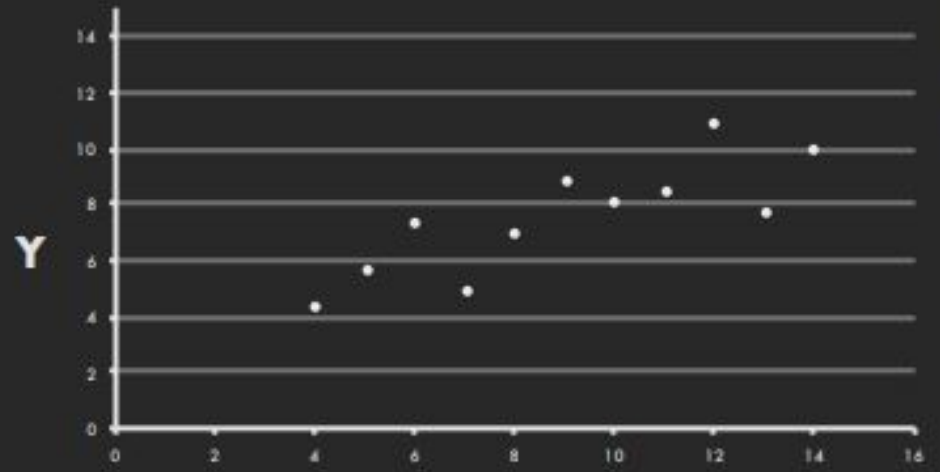
$u_X = 9.0$ $\sigma_X = 3.317$
$u_Y = 7.5$ $\sigma_Y = 2.03$

Linear Regression

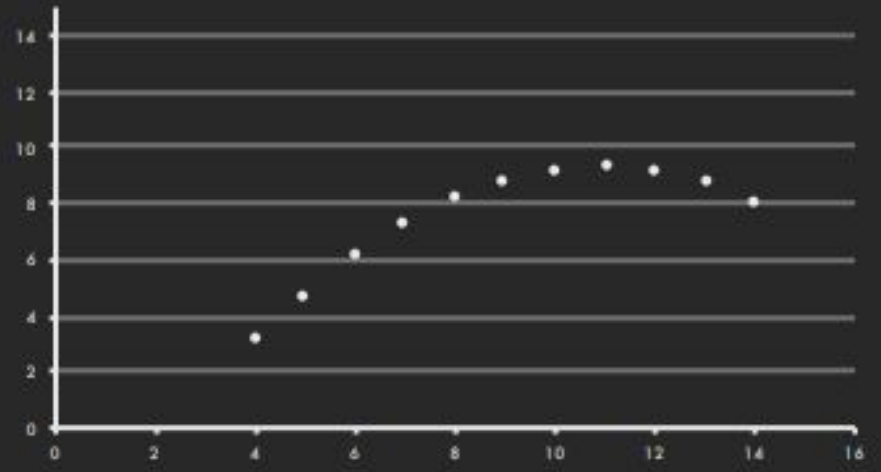
$Y = 3 + 0.5\ X$
$R^2 = 0.67$

[Anscombe 73]

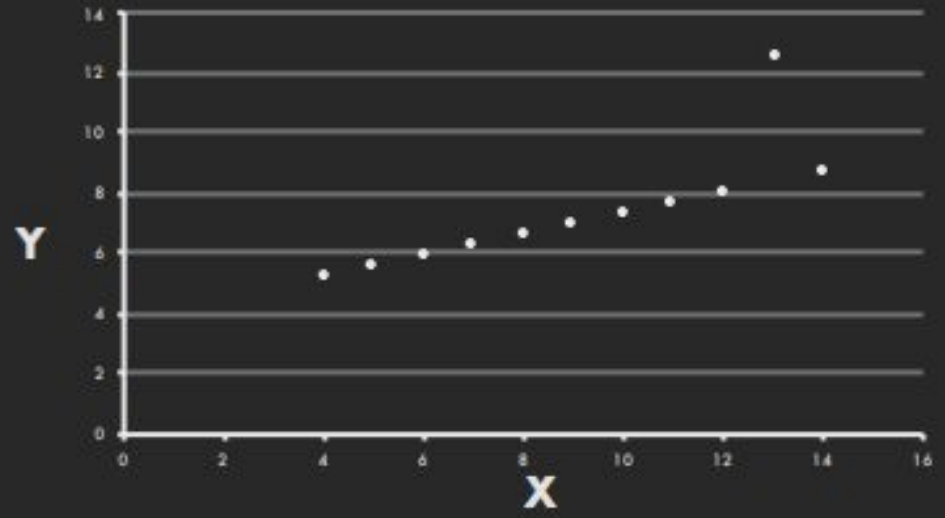
Set A



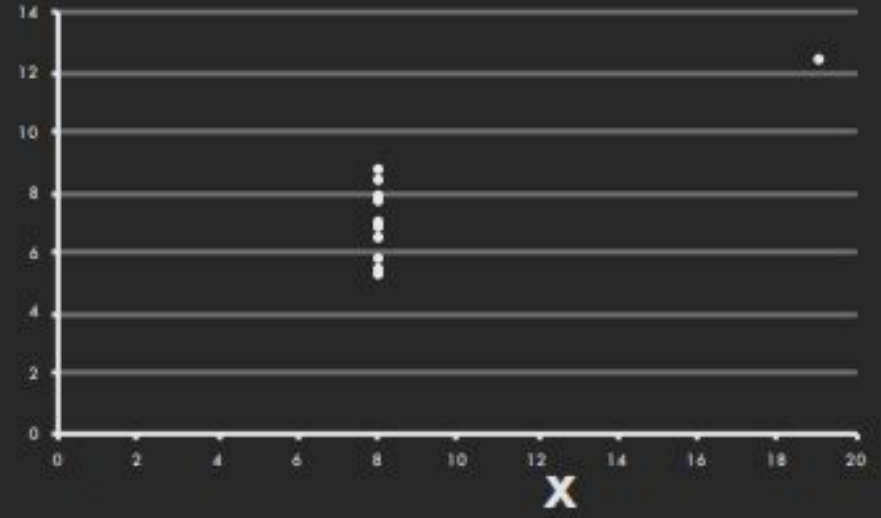
Set B



Set C



Set D

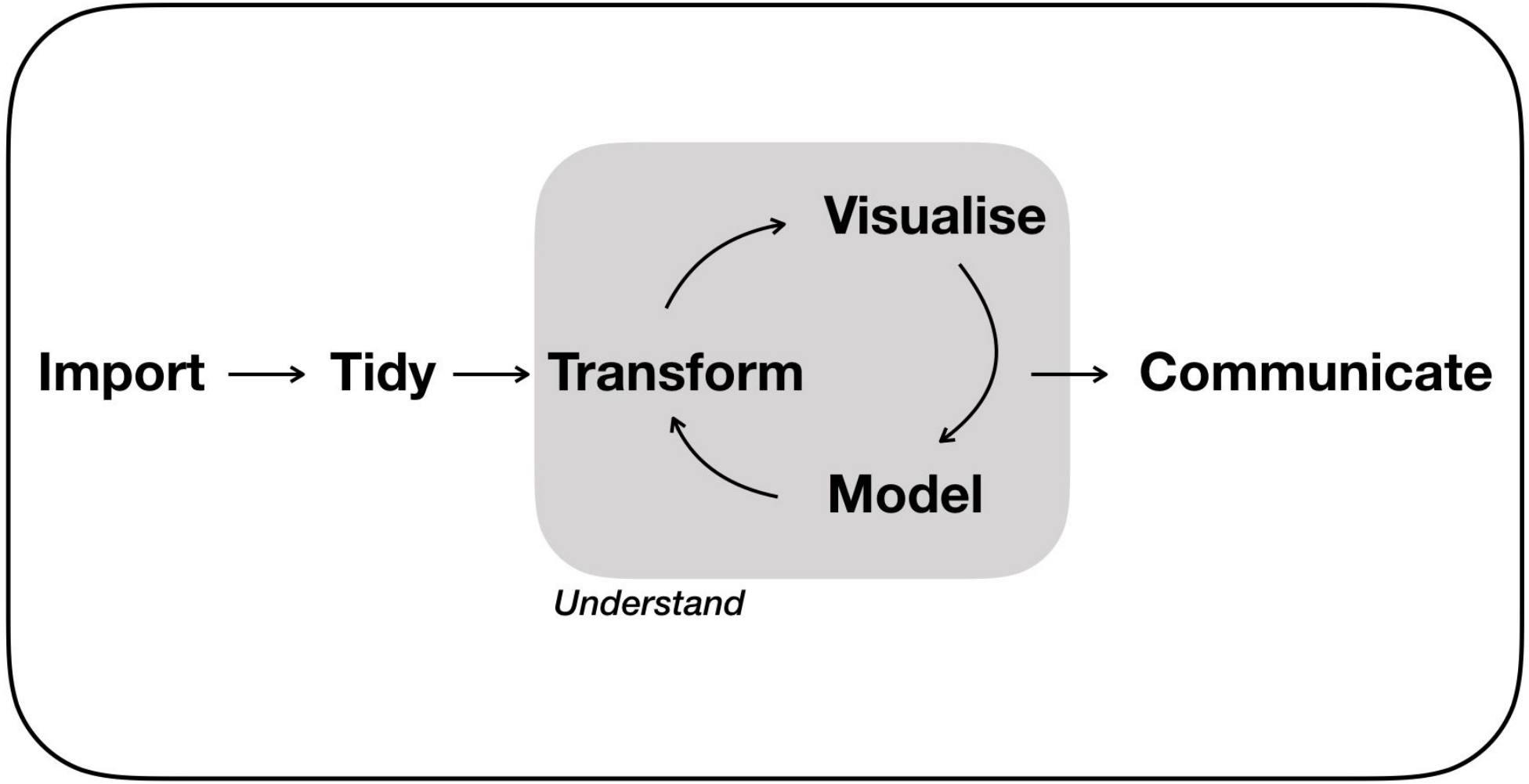


Two rules to create good visualizations

Do you have a **clear** question?

Do you **know** your data?

Data Science Workflow

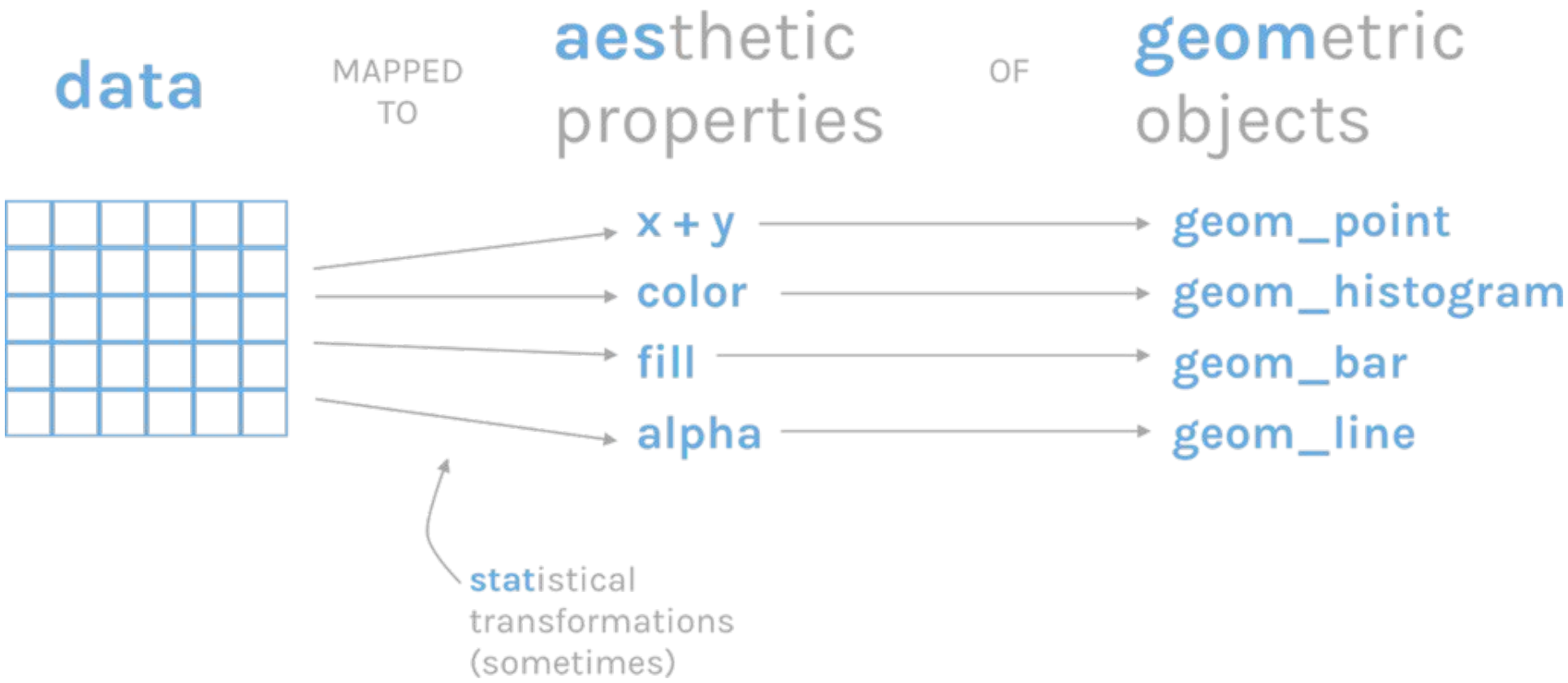


1. Motivation
2. **The grammar of a plot**
3. Essentials data transformations
4. Polishing the plot
5. Learn more

Graphic systems available for R

Base
Lattice
ggplot2

The grammar of graphics of plots: ggplot

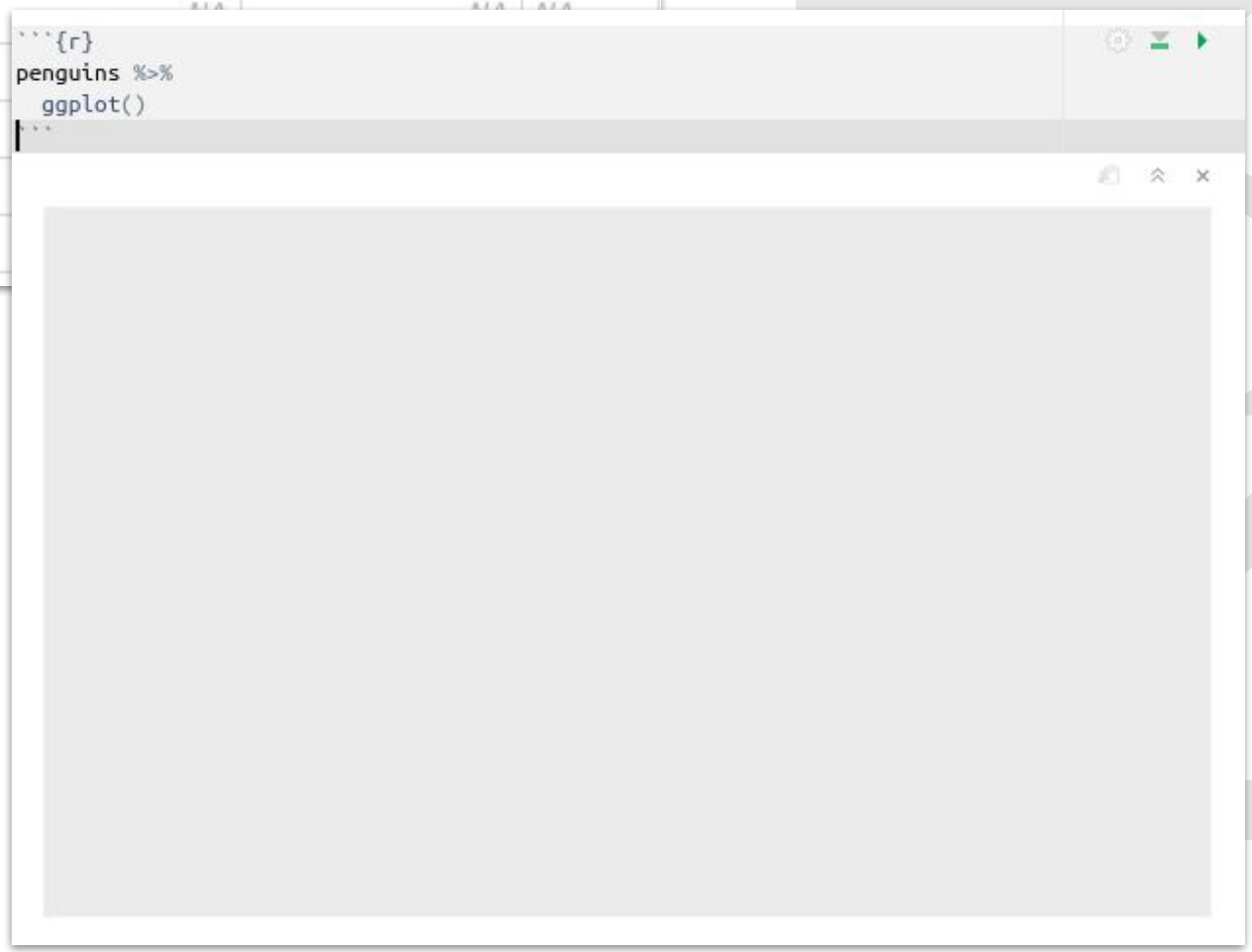


The grammar of graphics of plots: ggplot

To create a graph you need three things:

1. **Data**
2. Map some aesthetics
3. A geometry of the mapping

bill_depth_mm	flipper_length_mm	body_mass_g	sex
18.7	181	3750	male
17.4	186	3800	female
18.0	195	3250	female
NA			
19.3			
20.6			
17.8			
19.6			

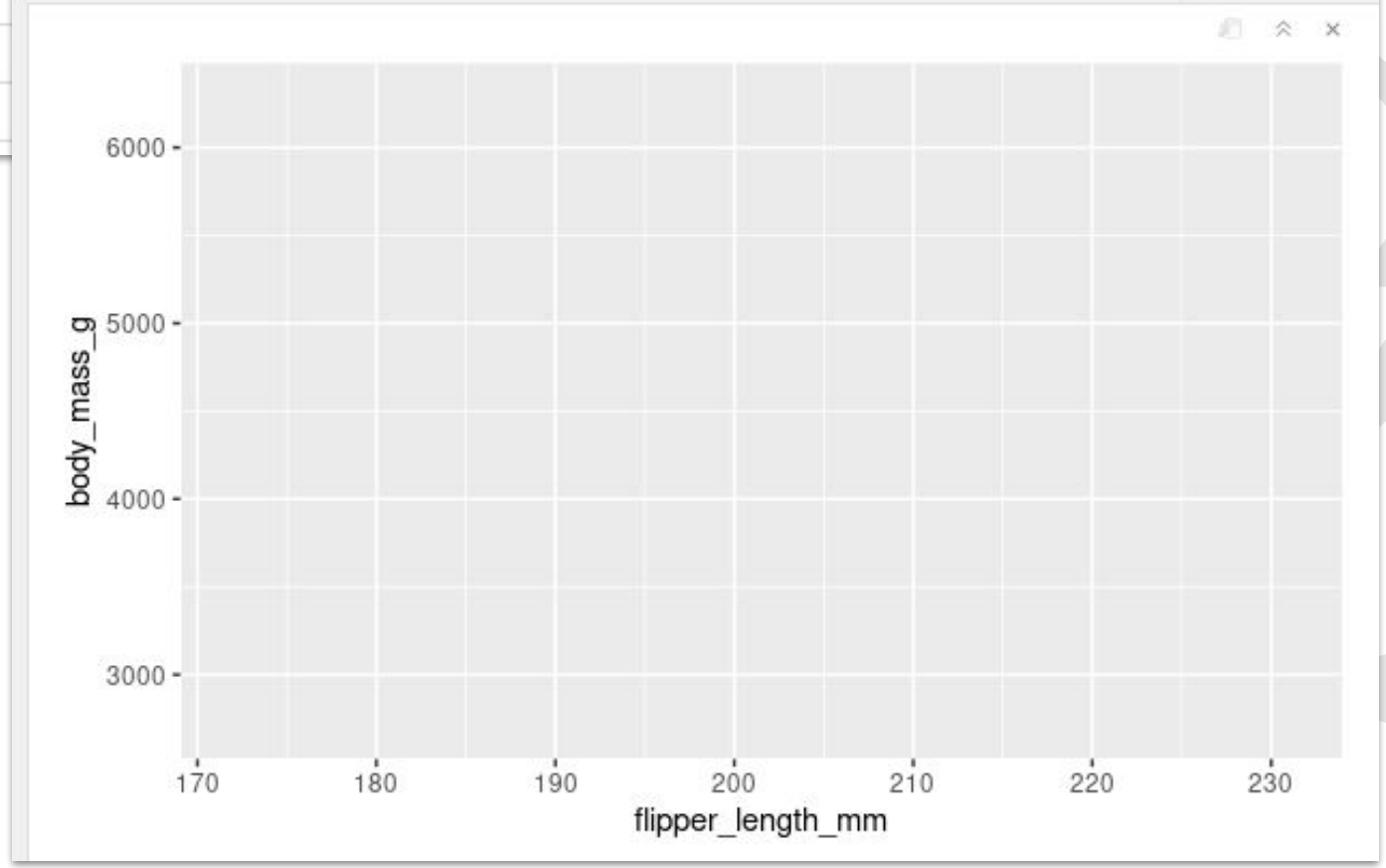


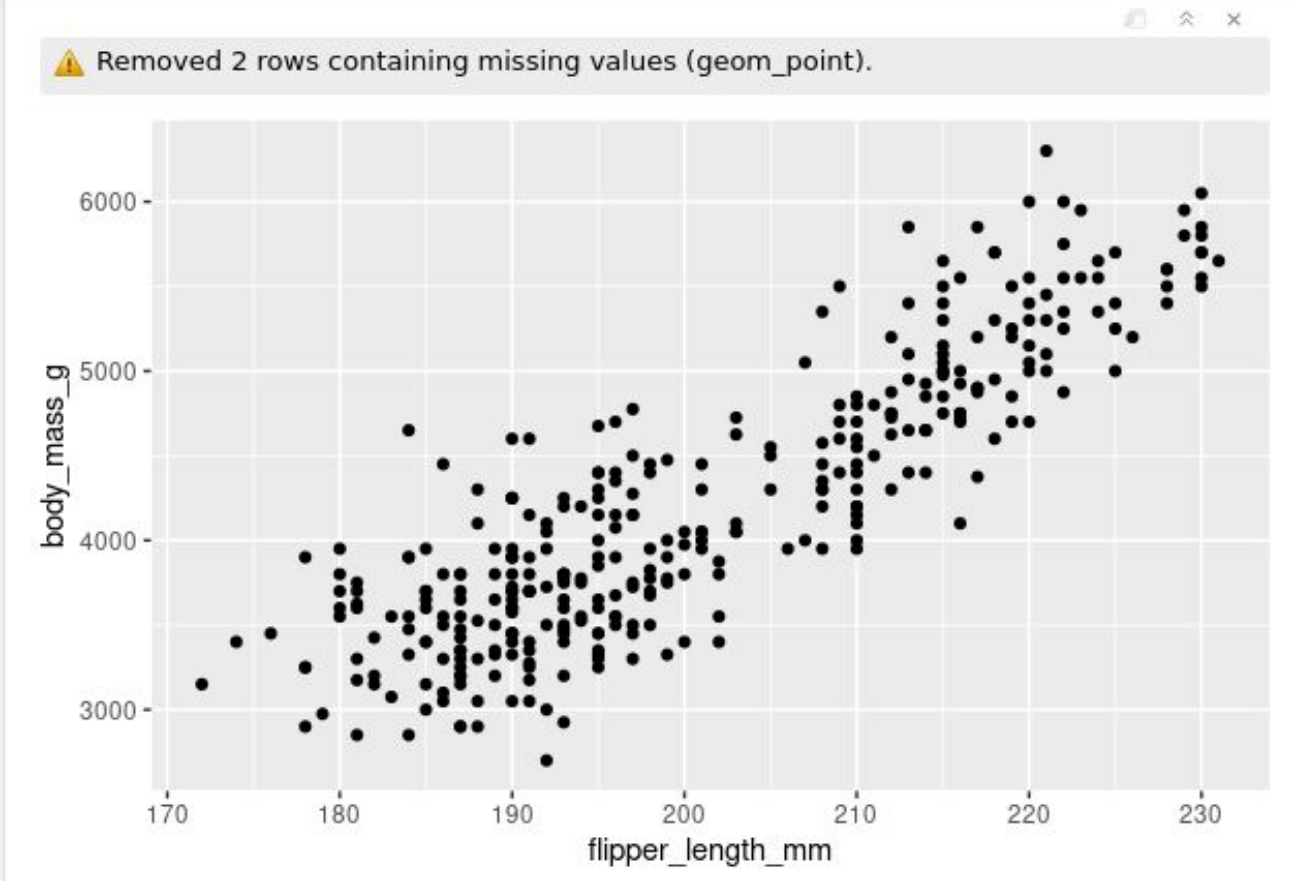
The grammar of graphics of plots: ggplot

To create a graph you need three things:

1. **Data**
2. **Map some aesthetics**
3. A geometry of the mapping

bill_depth_mm	flipper_length_mm	body_mass_g	sex
18.7	181	3750	male
17.4	186	3800	female
18.0			
NA			
19.3			
20.6			
17.8			
19.6			

```
```{r}
penguins %>%                                    # the data
  ggplot(mapping = aes(x = flipper_length_mm,   # the x variable
                       y = body_mass_g))        # the y variable
```
```



The grammar of graphics of plots: ggplot

To create a graph you need three things:

1. **Data**
2. **Map some aesthetics**
3. **A geometry of the mapping**

bill_depth_mm	flipper_length_mm	body_mass_g	sex
18.7	181	3750	male
17.4	186	3800	female
18.0			
NA			
19.3			
20.6			
17.8			
19.6			

```
```{r}
penguins %>%                                    # the data
  ggplot(mapping = aes(x = flipper_length_mm,   # the x variable
                       y = body_mass_g)) +      # the y variable
  geom_point()                                  # the geometry of the mapping
```
```



Let's start! Useful shortcuts

		
Insert chunk of code:	Ctrl+Alt+I	Cmd+Option+I
Insert pipe operator:	Ctrl+Shift+M	Cmd+Shift+M
Attempt completion:	Tab	Tab

1. Motivation
2. The grammar of a plot
3. **Essentials data transformations**
4. Polishing the plot
5. Learn more

Essential data manipulations

Data > then
    make some transformation
    plot the result

Essential data manipulations 1

```
Data %>%
    make some transformation
        filter(...) %>%
        select(...) %>%
        mutate(...) %>%
        arrange(...) %>%
        summarise (...) %>%
    plot the result
```

Read https://r4ds.had.co.nz/transform.html

Essential data manipulations 2: reshaping the dataset

country	year	metric
x	1960	10
x	1970	13
x	2010	15
y	1960	20
y	1970	23
y	2010	25
z	1960	30
z	1970	33
z	2010	35

```
pivot_wider(names_from = "year",
            names_prefix = "yr",
            values_from = "metric")
```

country	yr1960	yr1970	yr2010
x	10	13	15
y	20	23	25
z	30	33	35

```
pivot_longer(cols = yr1960:yr2010,
             names_to = "year",
             names_prefix = "yr"
             values_to = "metric")
```

read https://r4ds.had.co.nz/tidy-data.html#pivoting

1. Motivation
2. The grammar of a plot
3. Essentials data transformations
4. **Polishing the plot**
5. Learn more

RĪGAS STRADIŅA UNIVERSITĀTE

RSU

1. Motivation
2. The grammar of a plot
3. Essentials data transformations
4. Polishing the plot
5. **Learn more**

1. Motivation
2. The grammar of a plot
3. Essentials data transformations
4. Polishing the pilot
5. **Learn more**

from Data to Viz

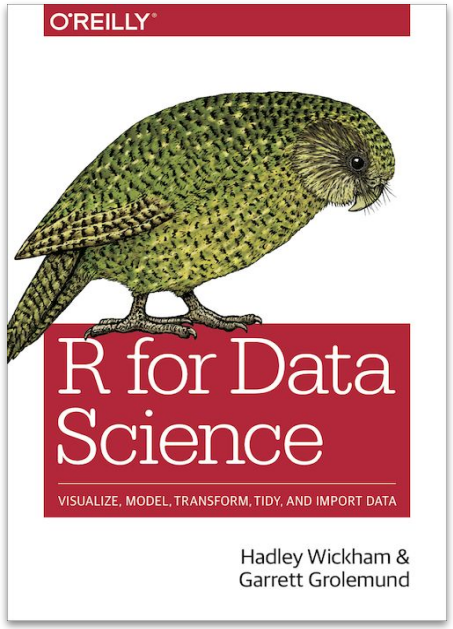
From Data to Viz leads you to the most appropriate graph for your data. It links to the code to build it and lists common caveats you should avoid.

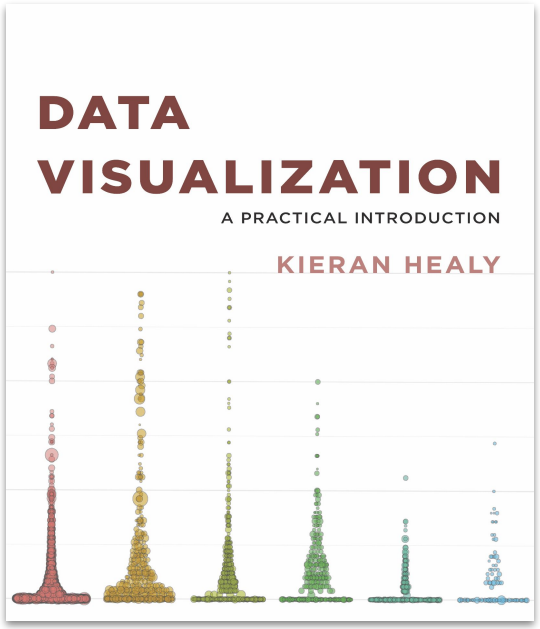
EXPLORE

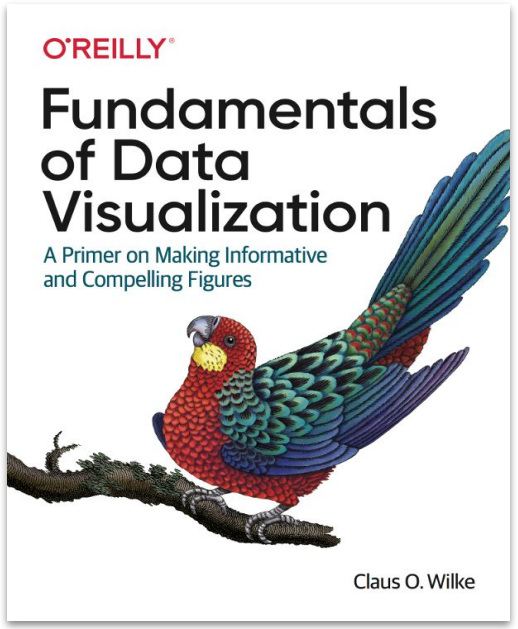
Learn more

Which graph? https://www.data-to-viz.com/
How to? https://r-graph-gallery.com/
Some inspiration https://github.com/rfordatascience/tidytuesday

Books

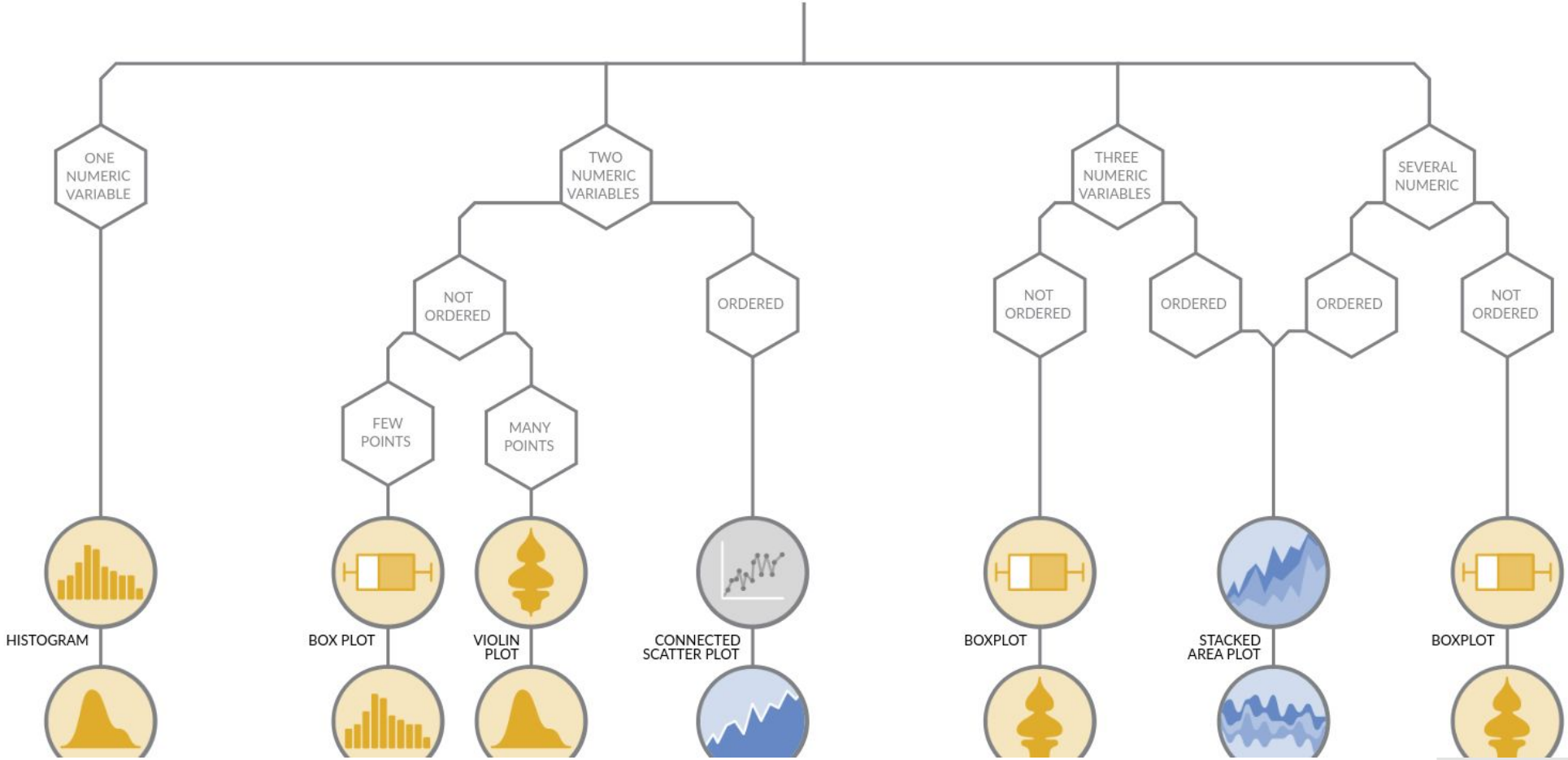






What kind of data do you have? Pick the main type using the buttons below. Then let the decision tree guide you toward your graphic possibilities.

Numeric | Categoric | Num & Cat | Maps | Network | Time series



The R Graph Gallery

Welcome the R graph gallery, a collection of charts made with the R programming language. Hundreds of charts are displayed in several sections, always with their reproducible code available. The gallery makes a focus on the tidyverse and ggplot2. Feel free to suggest a chart or report a bug; any feedback is highly welcome. Stay in touch with the gallery by following it on Twitter or Github. If you're new to R, consider following this course.

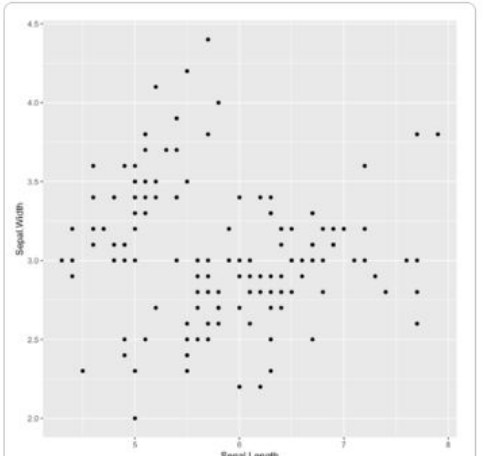
Distribution

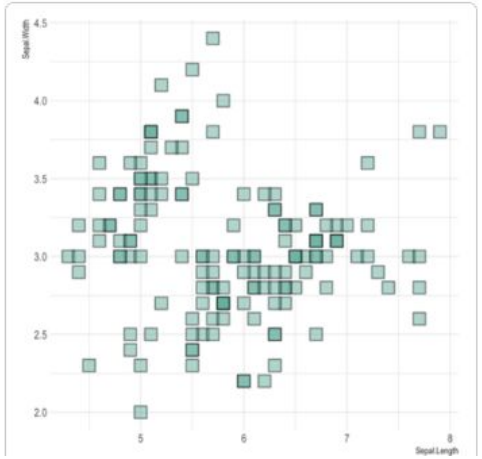
Scatterplot

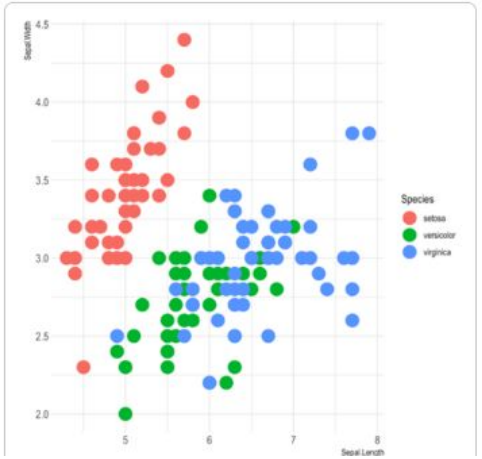
A Scatterplot displays the relationship between 2 numeric variables. Each dot represents an observation. Their position on the X (horizontal) and Y (vertical) axis represents the values of the 2 variables. Using ggplot2, scatterplots are built thanks to the `geom_point` geom. If you're not familiar with ggplot2 at all, try this course as an introduction.

USING THE `GGPLOT2` PACKAGE

Scatterplots are built with ggplot2 thanks to the `geom_point()` function. Discover a basic use case in graph #272, and learn how to custom it with next examples below.



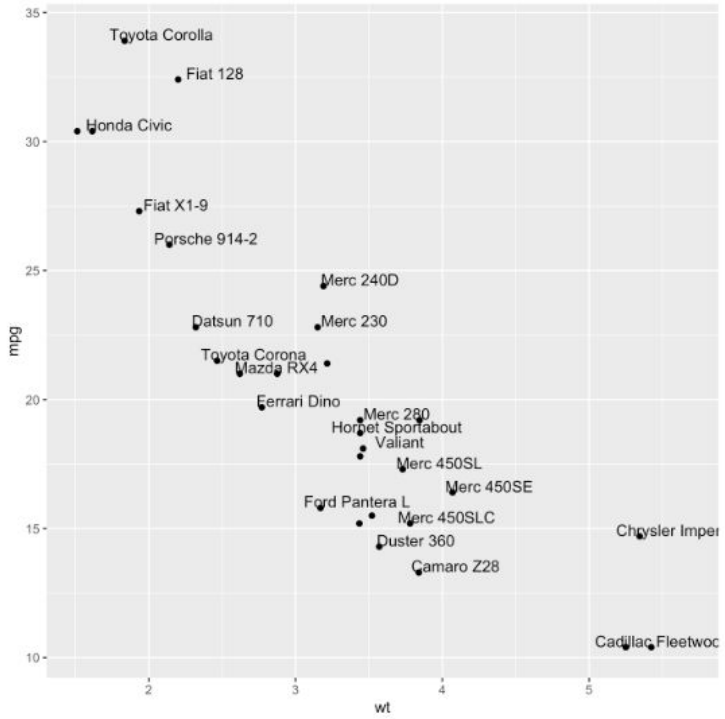




Adding text with `geom_text()`

This example demonstrates how to use `geom_text()` to add text as markers. It works pretty much the same as `geom_point()`, but add text instead of circles. A few arguments must be provided:

- `label`: what text you want to display
- `nudge_x` and `nudge_y`: shifts the text along X and Y axis
- `check_overlap` tries to avoid text overlap. Note that a package called `ggrepel` extends this concept further

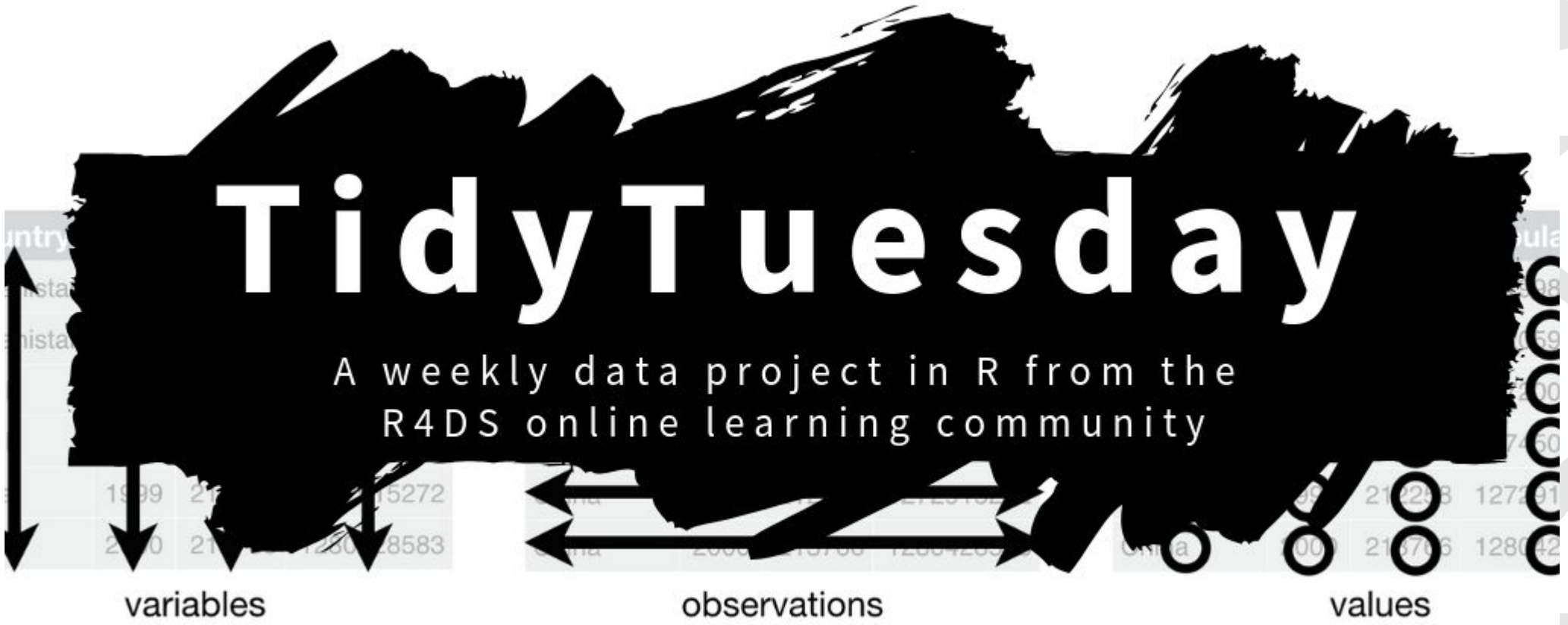



```
 # library
library(ggplot2)

# Keep 30 first rows in the mtcars natively available dataset
data=head(mtcars, 30)

# 1/ add text with geom_text, use nudge to nudge the text
ggplot(data, aes(x=wt, y=mpg)) +
  geom_point() + # Show dots
  geom_text(
```

More


TidyTuesday
A weekly data project in R from the
R4DS online learning community
variables
observations
values

New York Times Hardcover Fiction Bestsellers, 1931-2020

An interactive table for #TidyTuesday week 19 (2022)

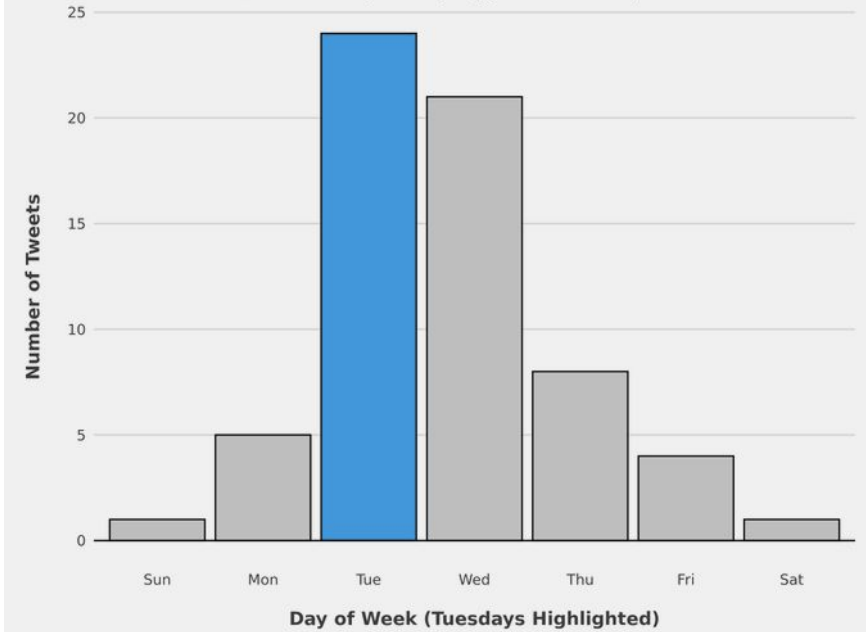
Search

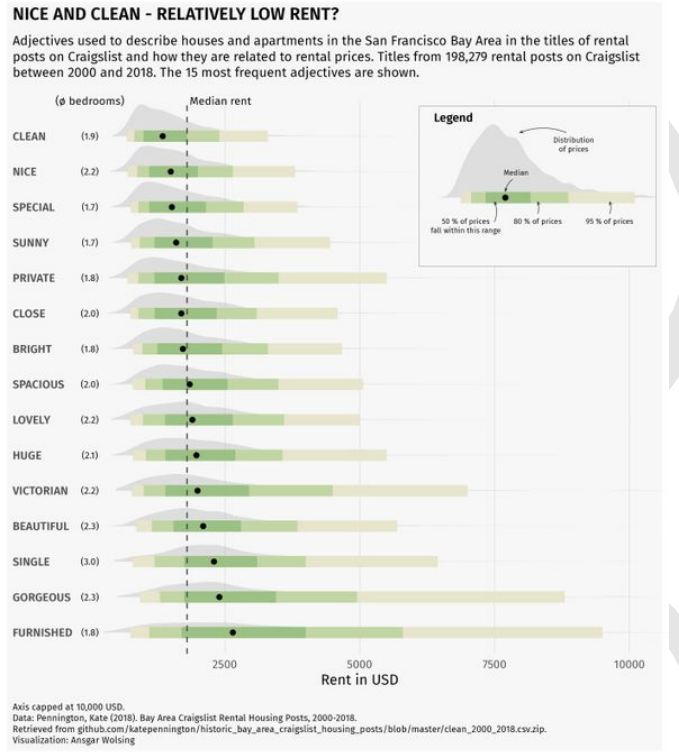
DEBUT RANK ON THE LIST	BEST RANK	YEAR	TITLE	AUTHOR	TOTAL WEEKS ON BEST SELLERS LIST
14	1	1990	OH, THE PLACES YOU'LL GO!	Dr. Seuss	178
10	1	1994	THE CELESTINE PROPHECY	James Redfield	165
9	1	2003	THE DA VINCI CODE	Dan Brown	165
7	1	1992	THE BRIDGES OF MADISON COUNTY	Robert James Waller	164
2	1	2014	ALL THE LIGHT WE CANNOT SEE	Anthony Doerr	132
13	1	1951	THE CAINE MUTINY	Herman Wouk	123
9	1	2018	WHERE THE CRAWDADS SING	Delia Owens	114
16	1	1955	AUNTIE MAME	Patrick Dennis	112
5	1	1942	THE ROBE	Lloyd C. Douglas	111
15	1	2009	THE HELP	Kathryn Stockett	108

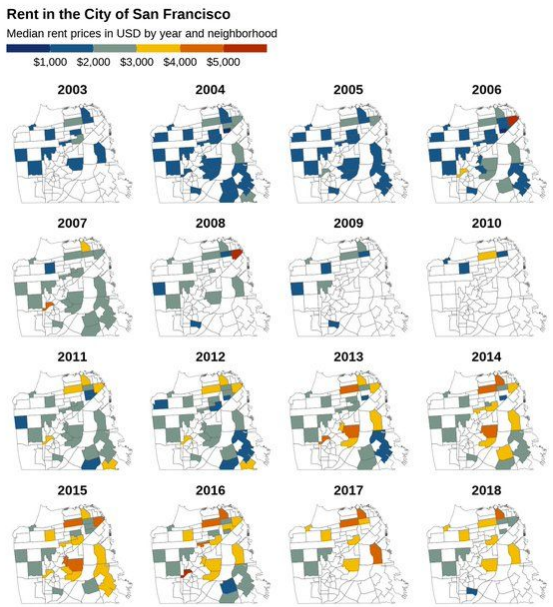
1–10 of 7431 rows

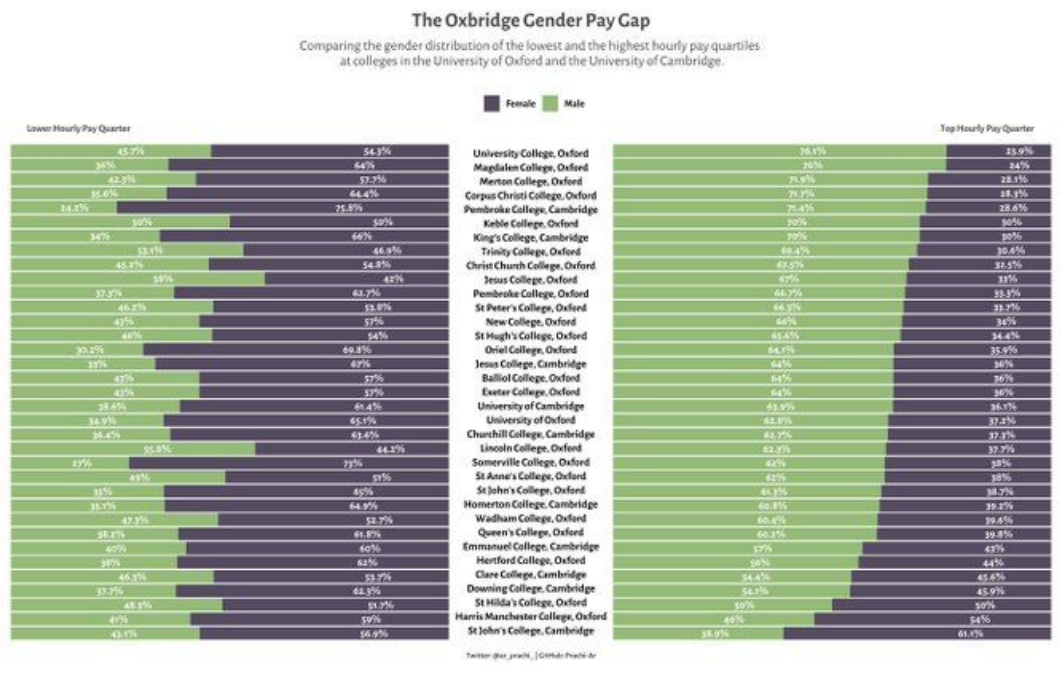
Previous 1 2 3 4 5 ... 744 Next

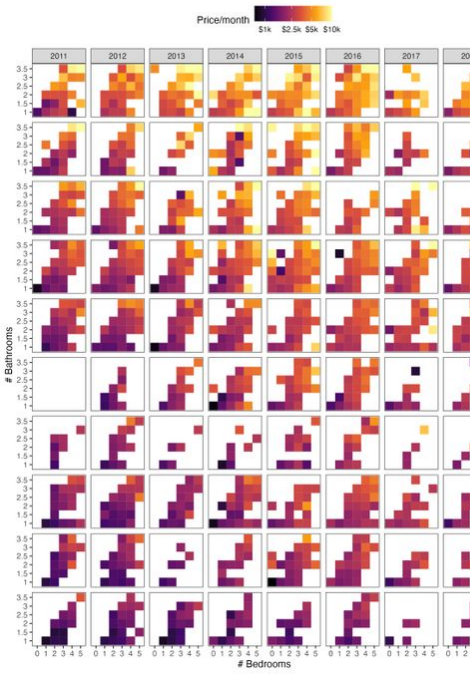
Data by: Kelly, Nicholas; White, Nicole, Glass, Loren, 03/01/2021, "The Program Era Project," DOI: https://doi.org/10.18737/CNJV1733p4520210415, Post45 Data Collective, V1.












Google
how to [INSERT HERE YOUR QUESTION] ggplot2
Buscar con Google
Voy a tener suerte
Ofrecido por Google en: English latviešu lietuvių русский

RSD

The greatest value of a picture is when it forces us to notice what we never expected to see.
—John Tukey

The goal is to turn *data* into *information*, and information into *insight*.
—Carly Fiorina

sergiouri.**be**

@sergio**uribe**

sergio.**uribe**@rsu.**lv**