

YOU WILL READ THIS FIRST.

And then you will read this line next.

You will go back to read this body copy if you want to know more. It takes the most effort to read because it has a lot of text in a small font in a light weight with tight line spacing. Many people will skip paragraphs like this unless if they aren't engaged right away. This is why it's important to draw attention to your message using visual hierarchy.

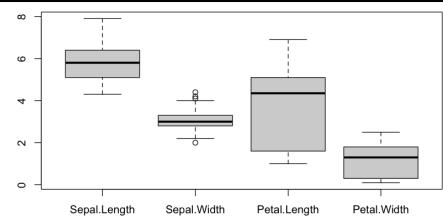
You'll probably
read this before
the paragraph.

1

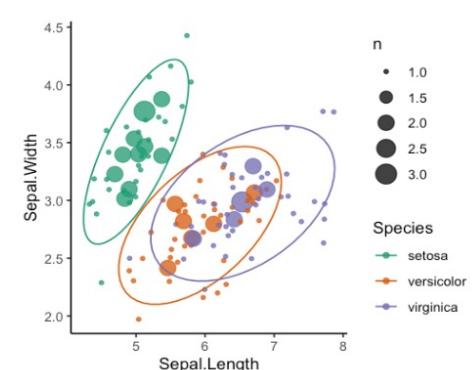
Beyond bar charts

Data visualisation
workshop

Laura Dijkhuizen
Bioinformatics & Theoretical
Biology @UU



Welch Two Sample t-test
data: iris\$Sepal.Length[iris\$Species == "setosa"] and iris\$Sepal.Length[iris\$Species == "versicolor"]
t = -10.521, df = 86.538, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to zero



2

1

Challenger example

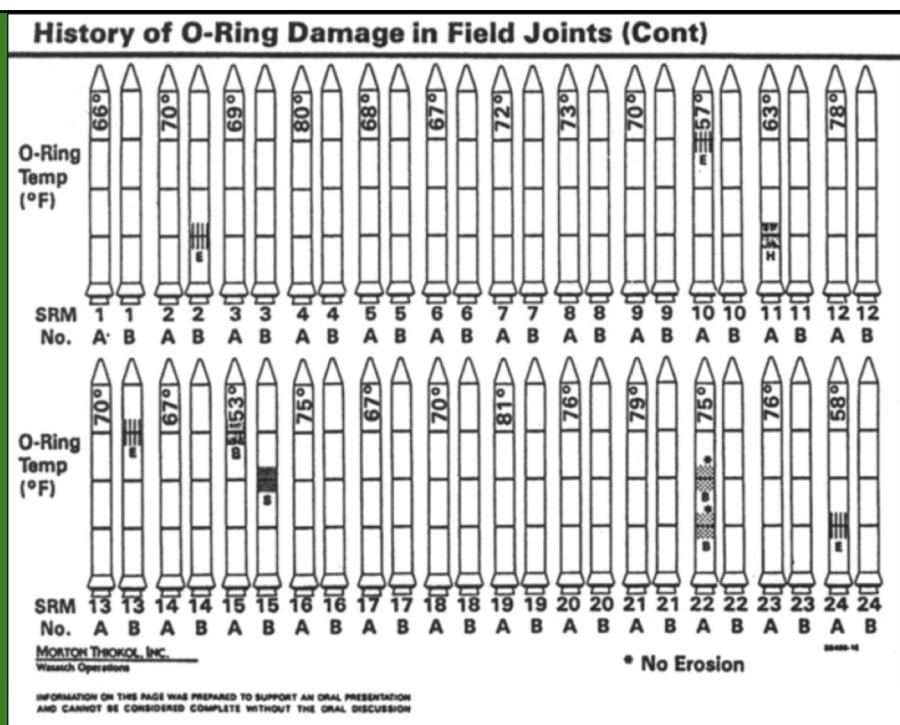
Challenger Space Shuttle accident in 1986.



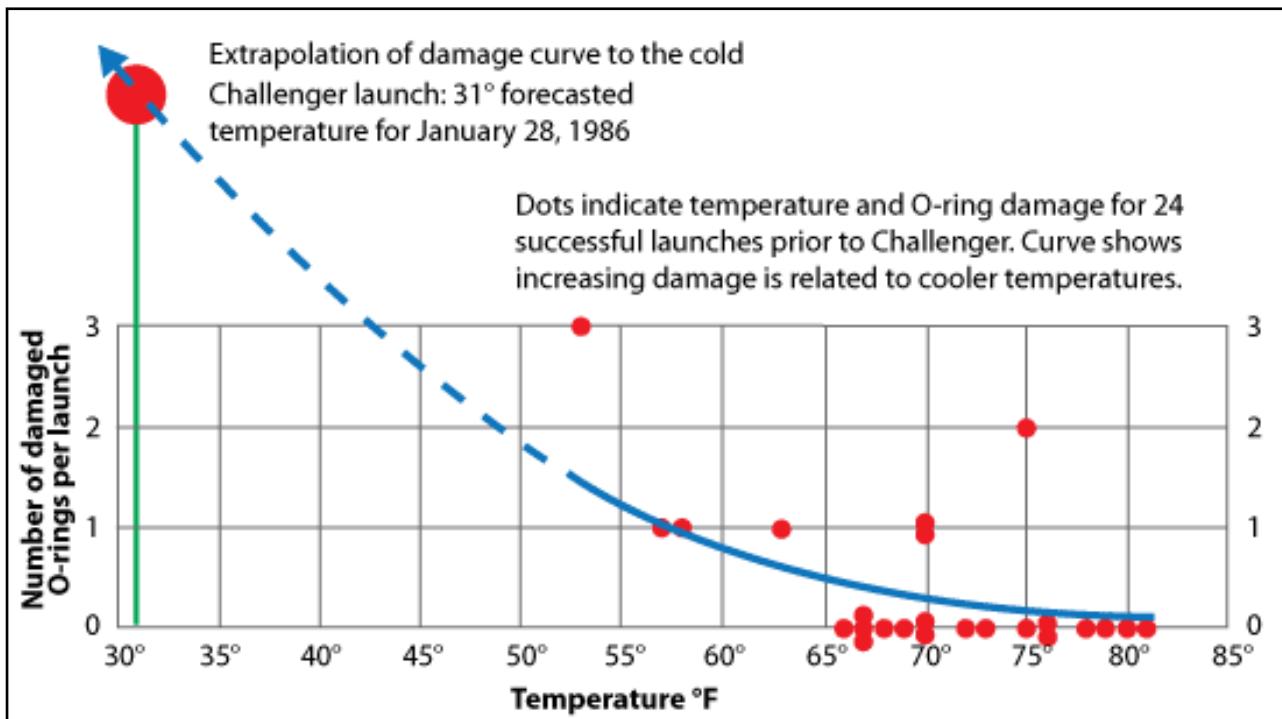
3

Challenger example:

Forcasted
temperature:
31 °F



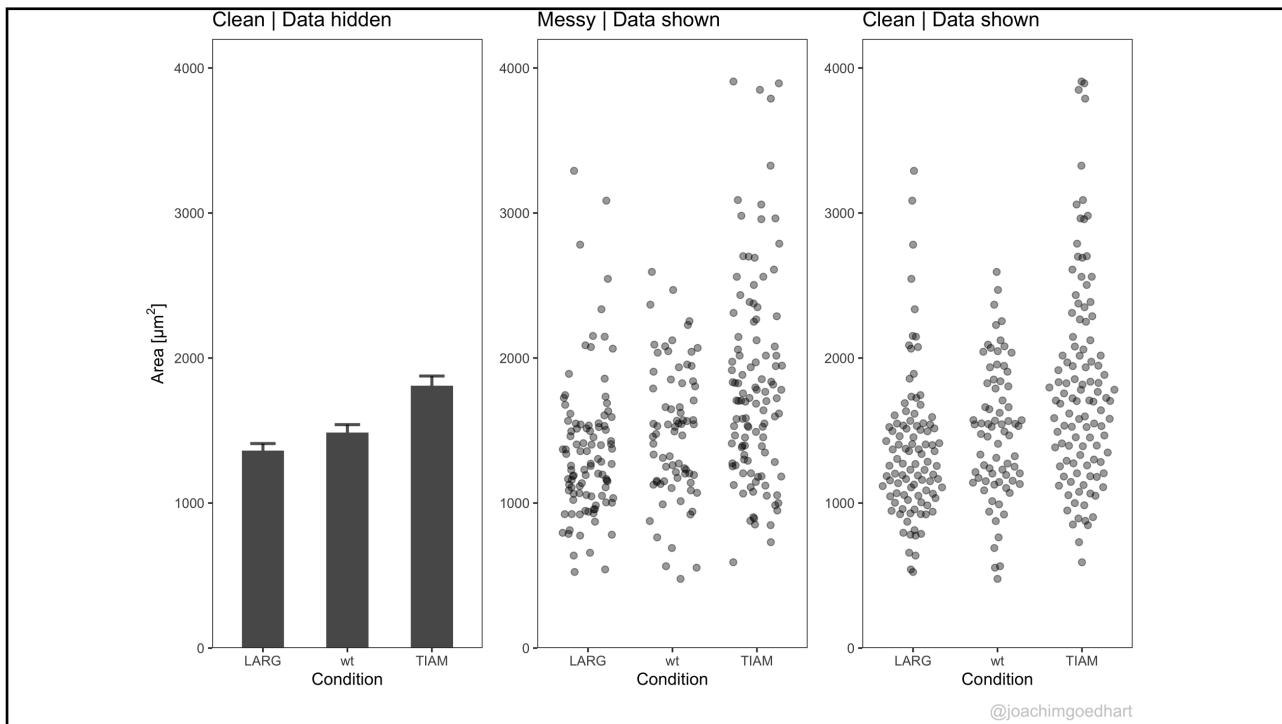
4



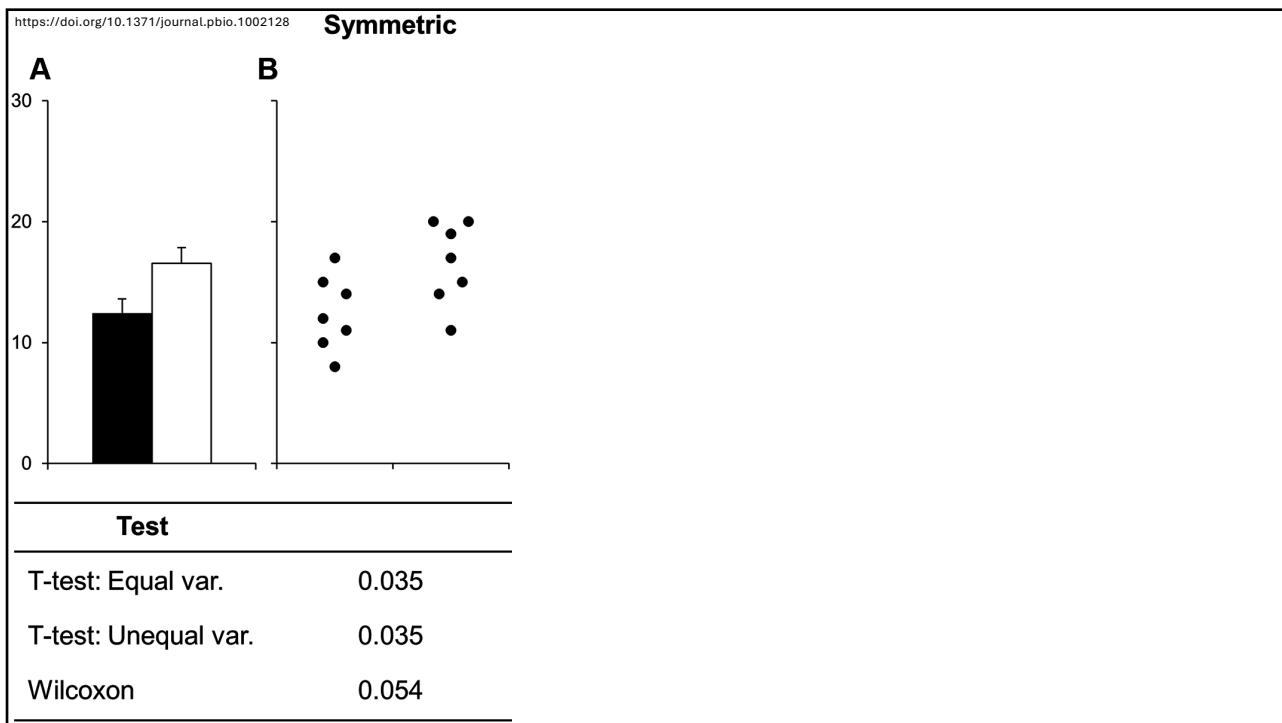
5

Let's look at some figures!

6

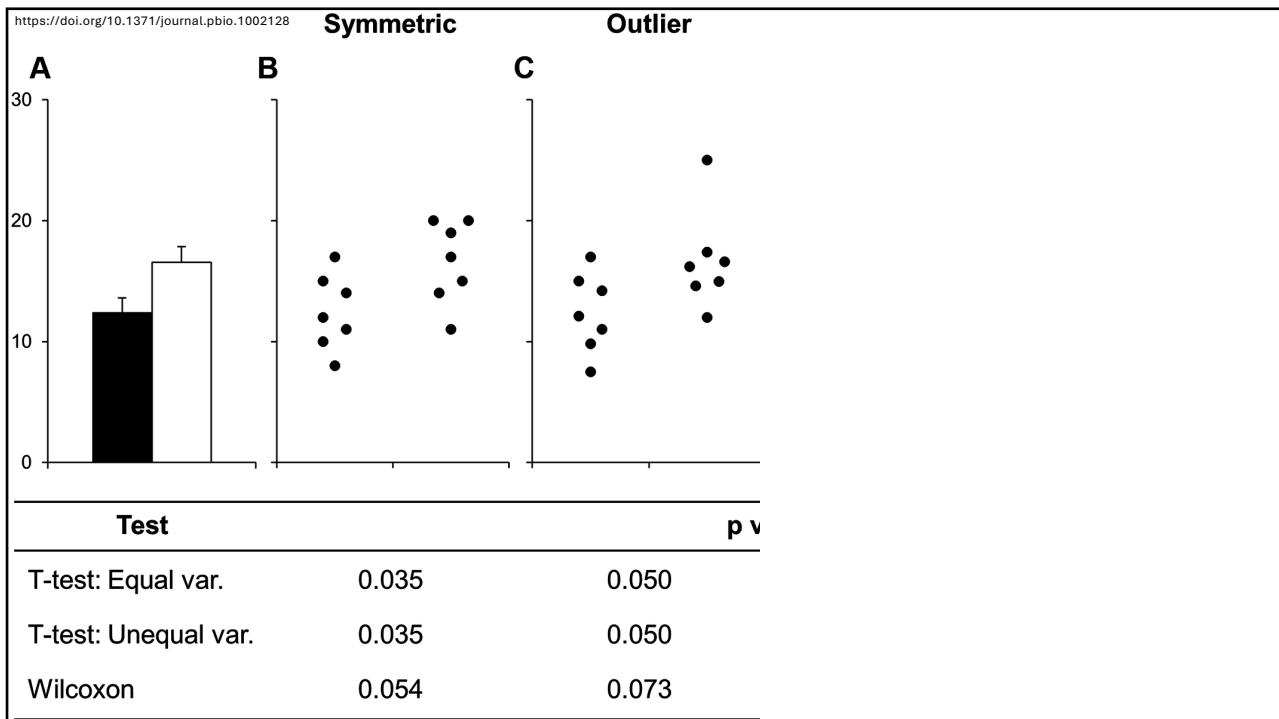


7

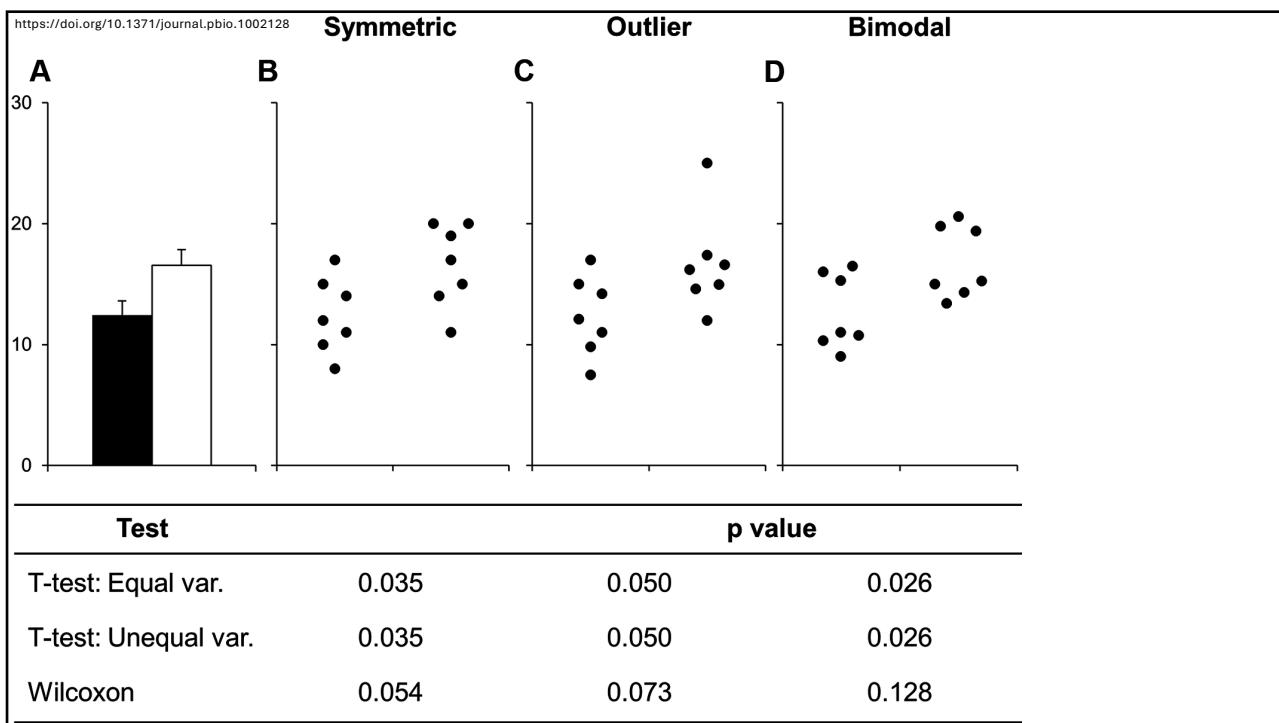


8

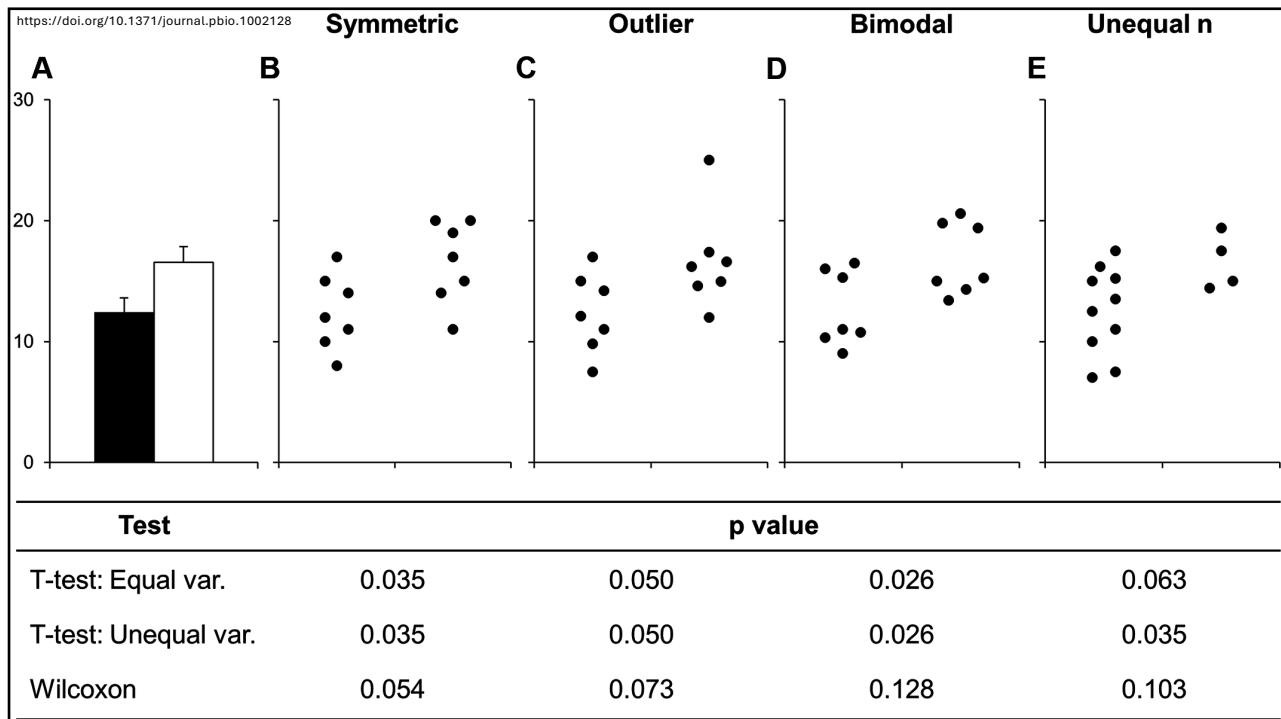
4



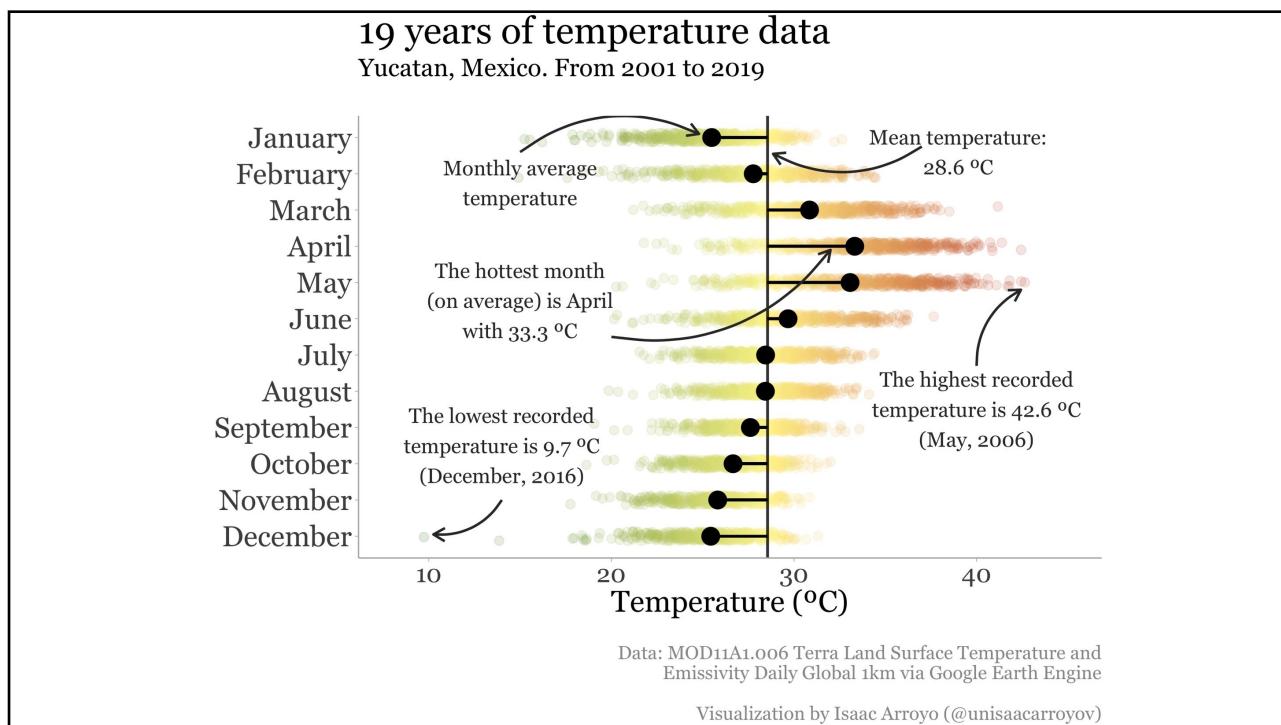
9



10



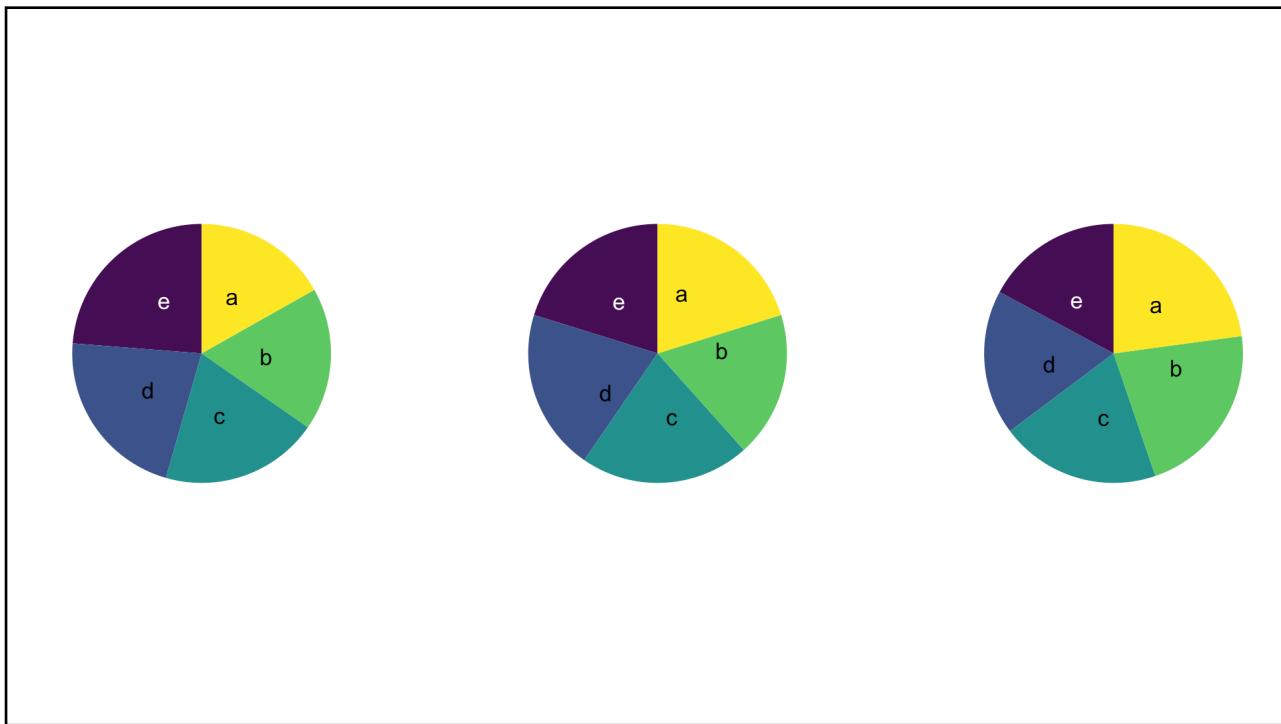
11



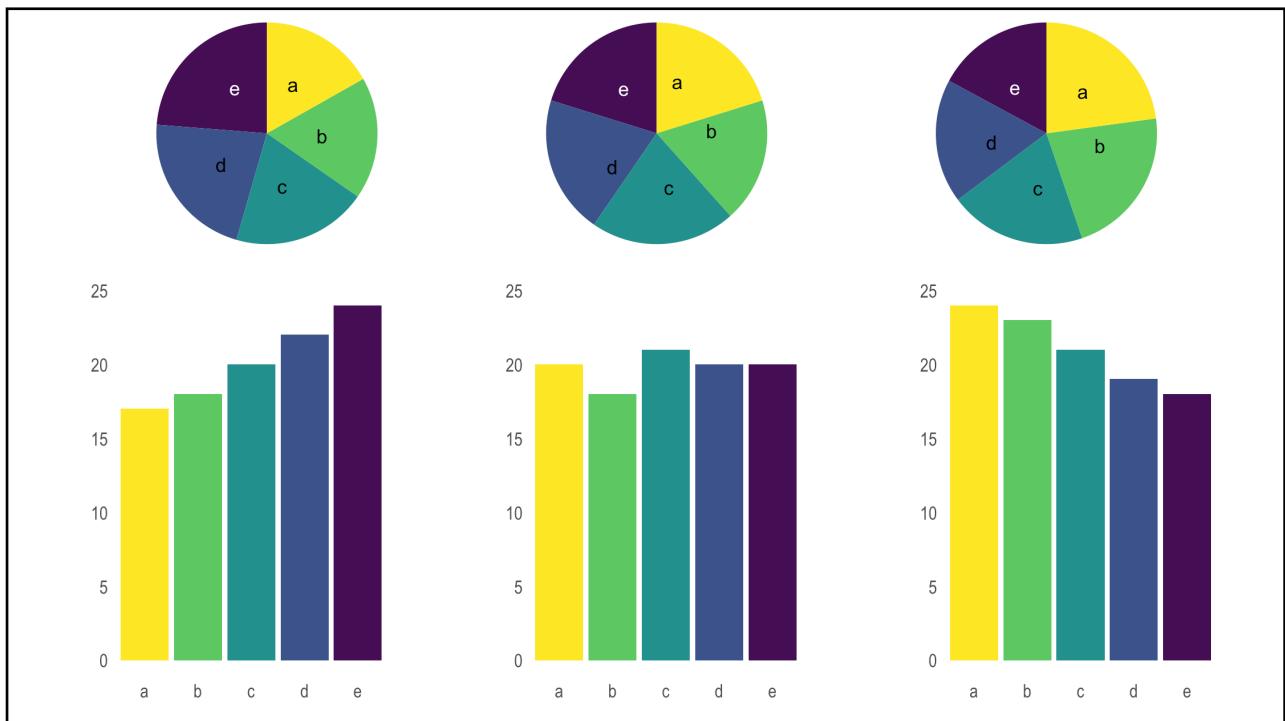
12



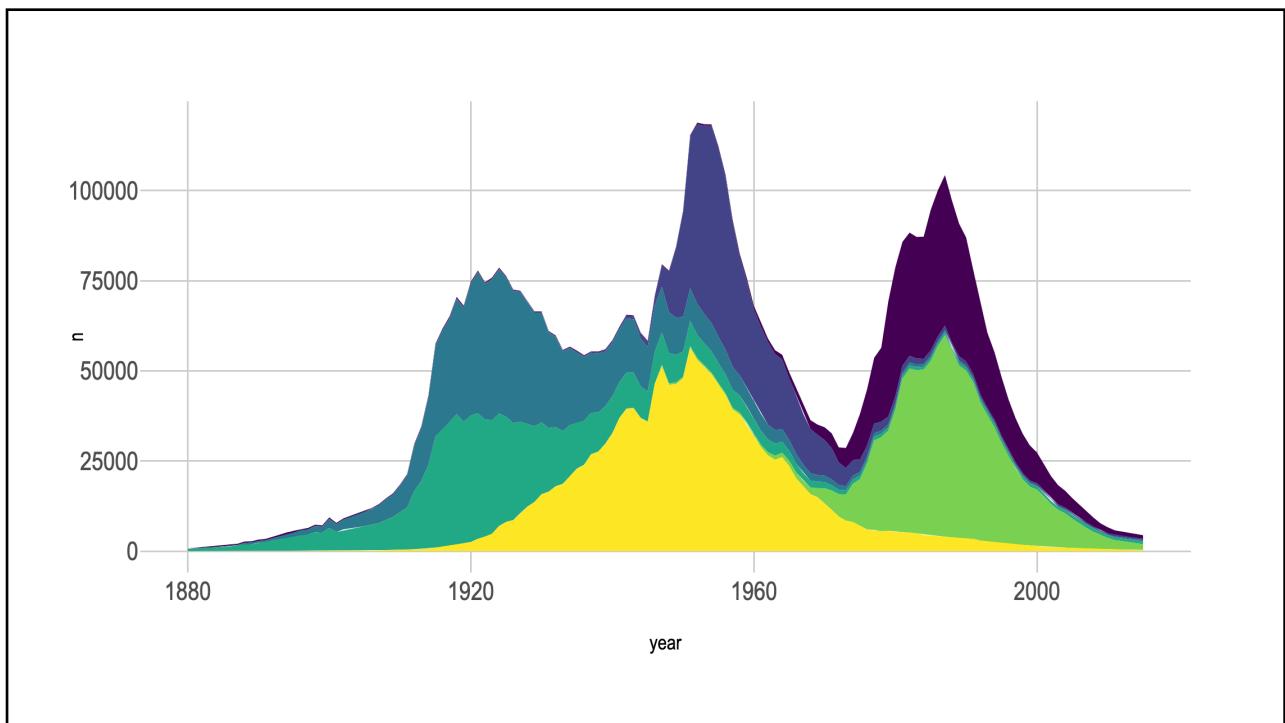
13



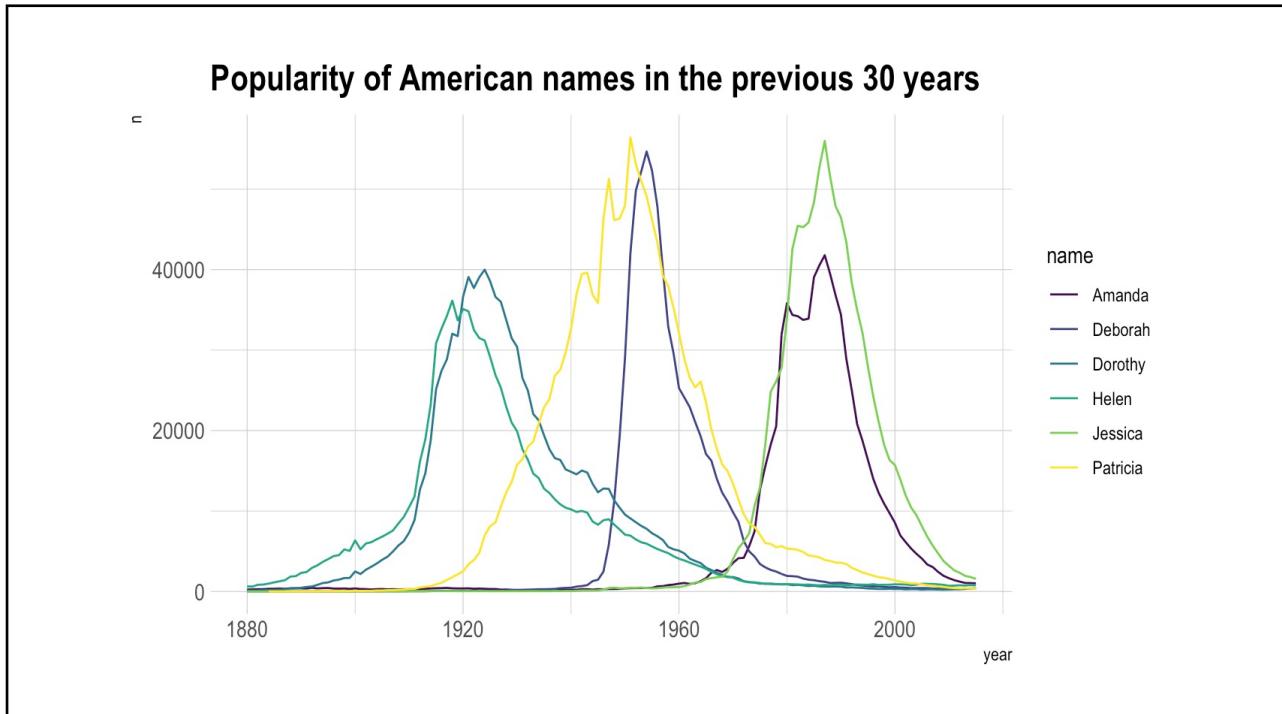
14



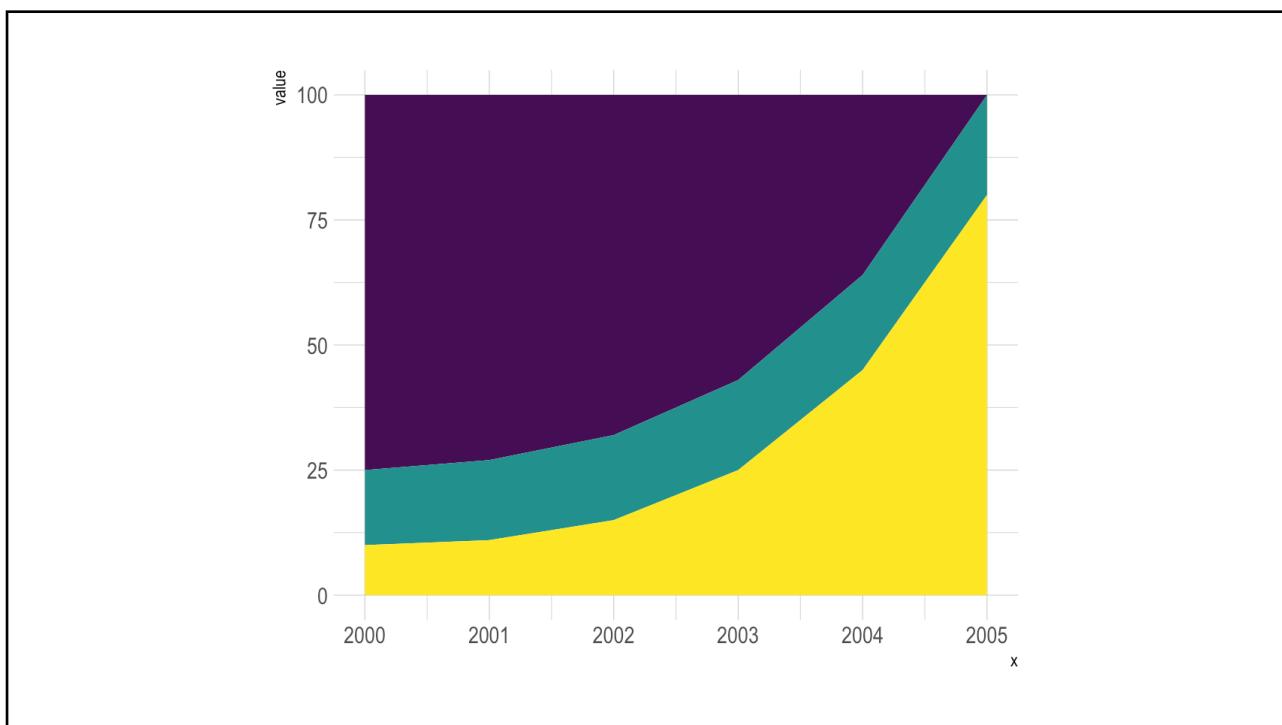
15



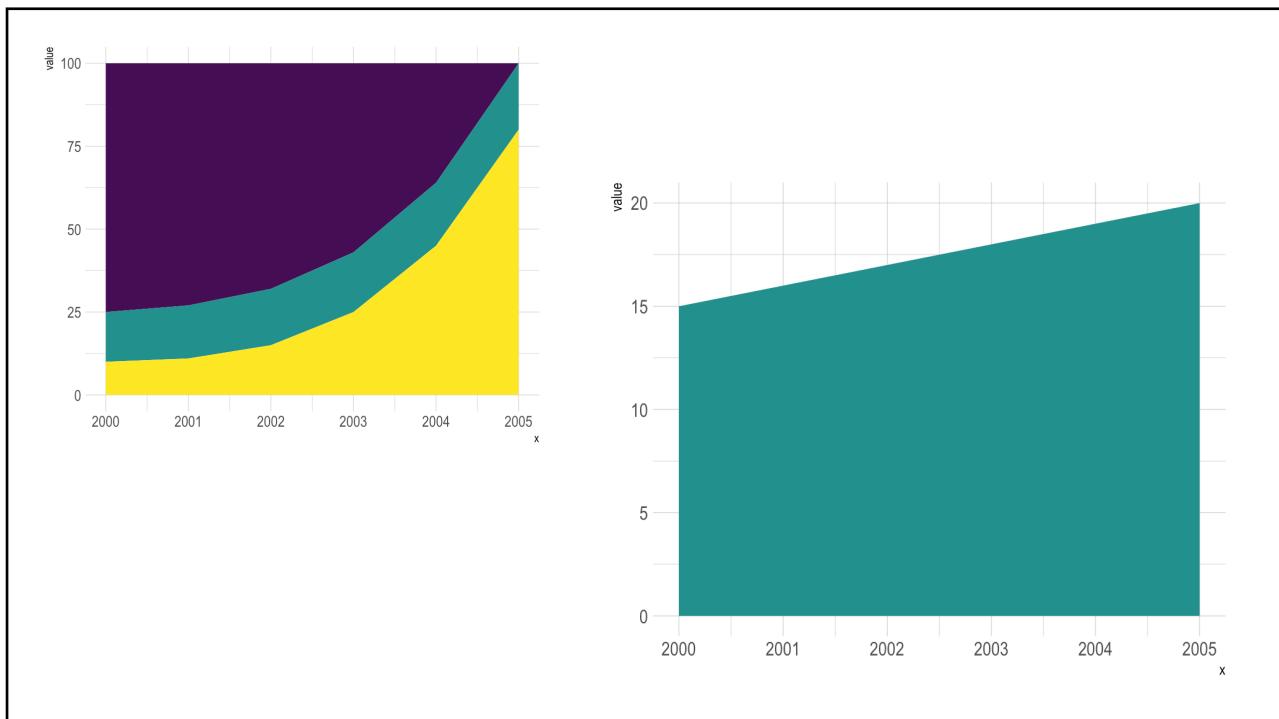
16



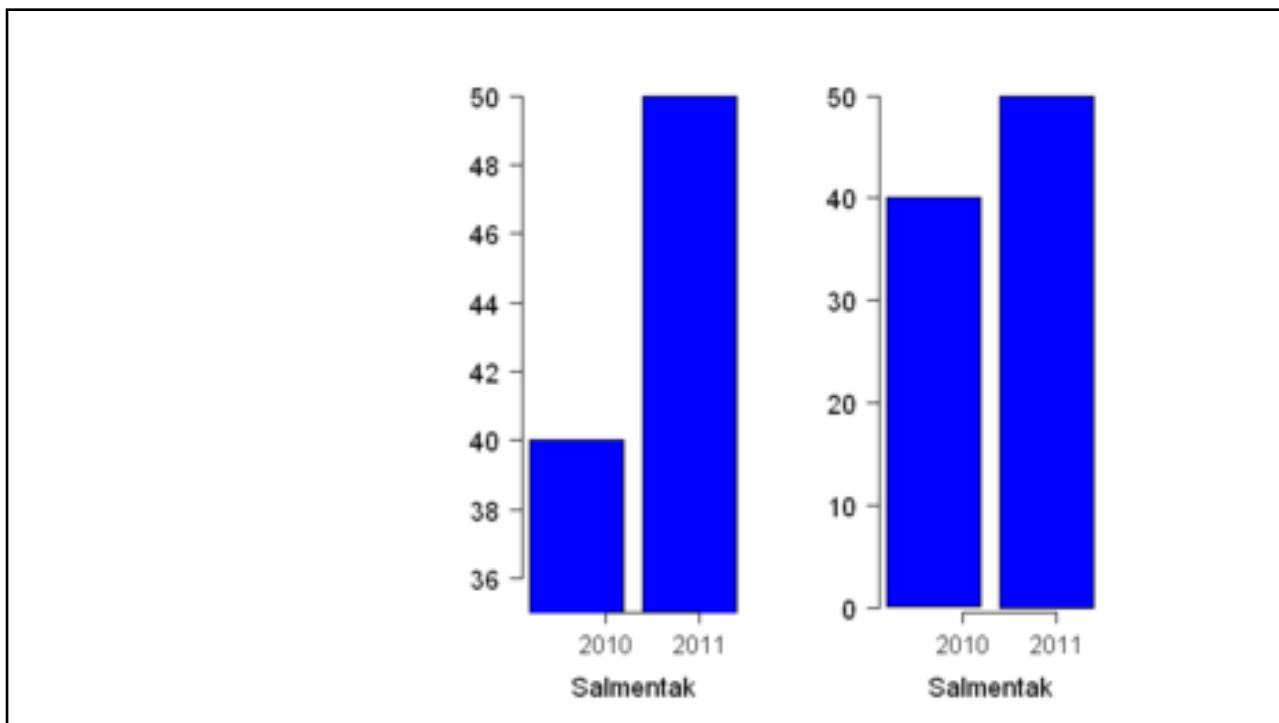
17



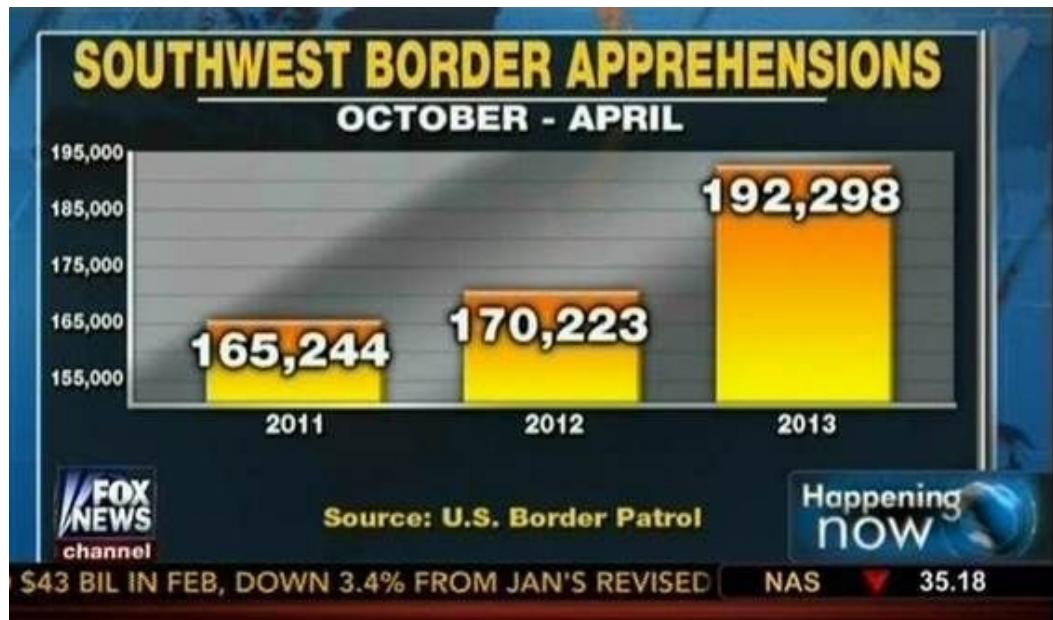
18



19



20

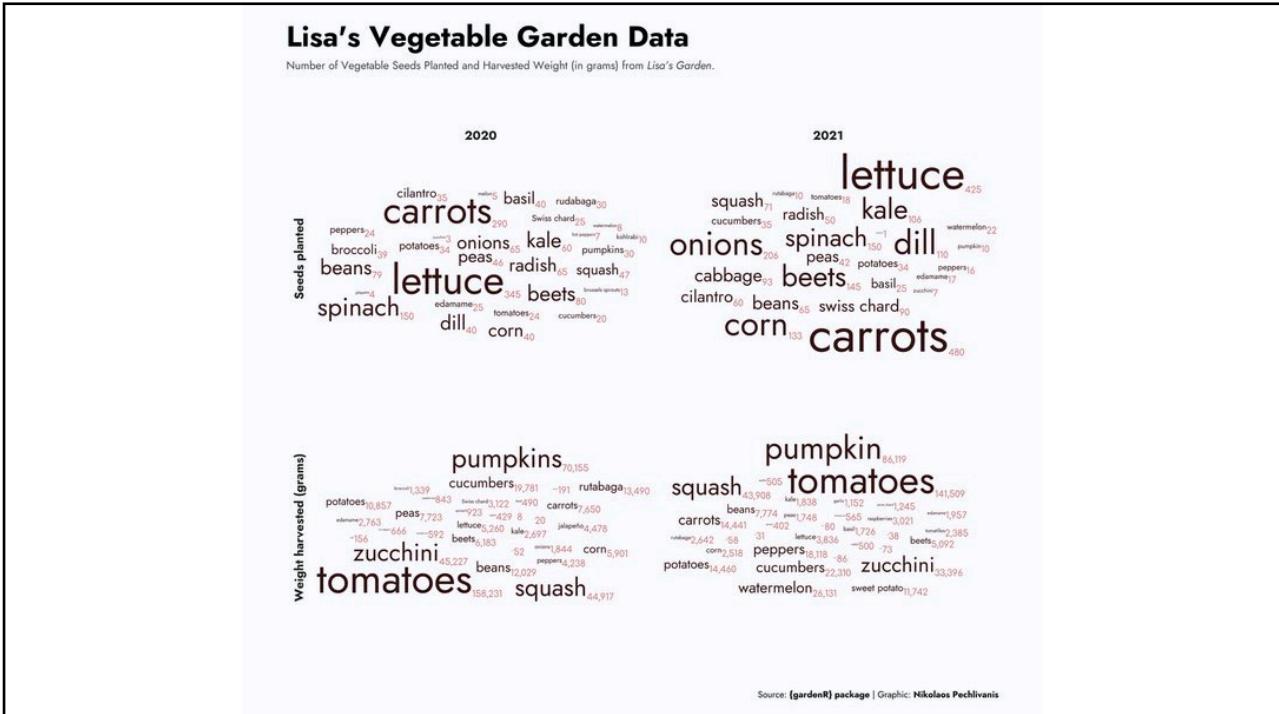


21

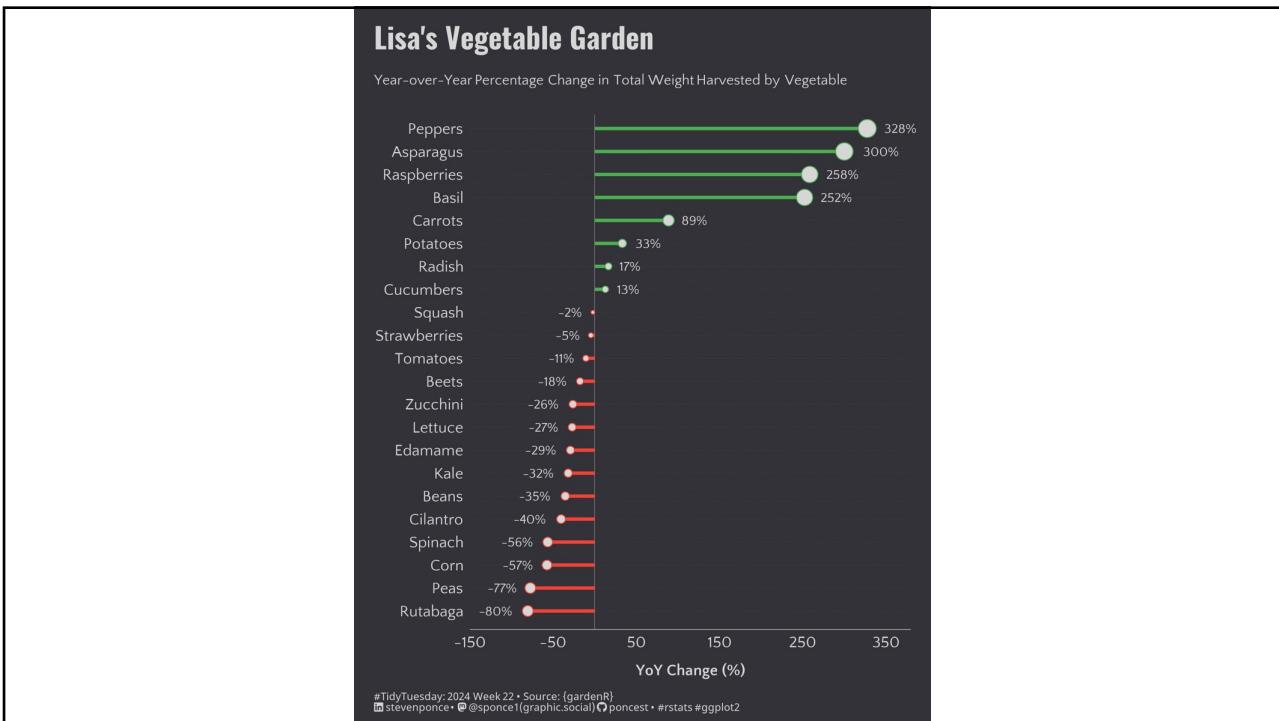
Take home 1:

There are many factually correct ways to visualise your data, but few truly effective ones.

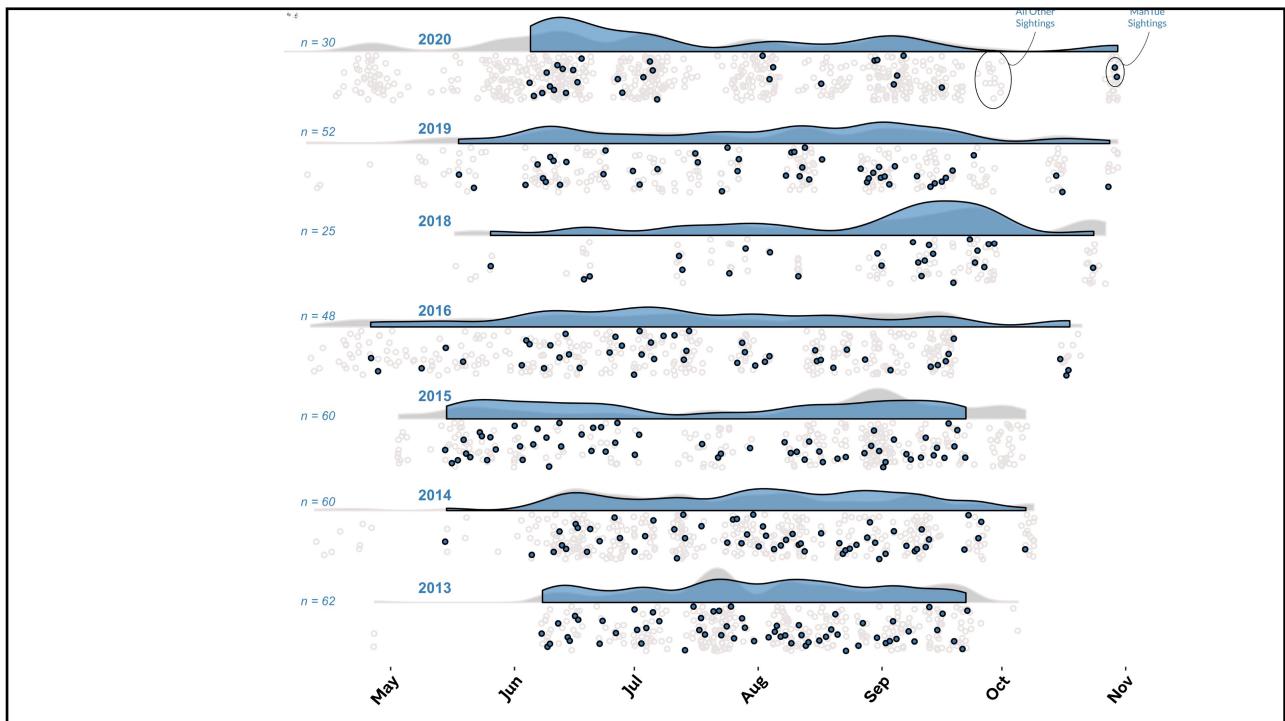
22



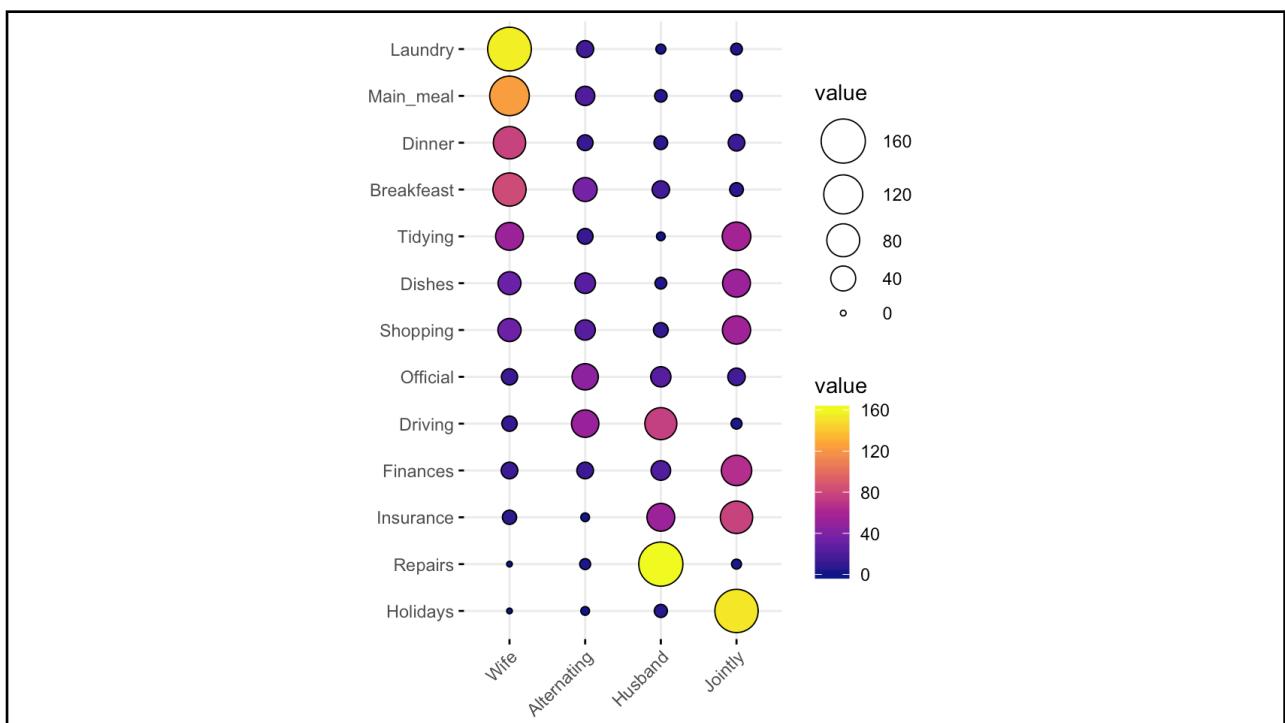
23



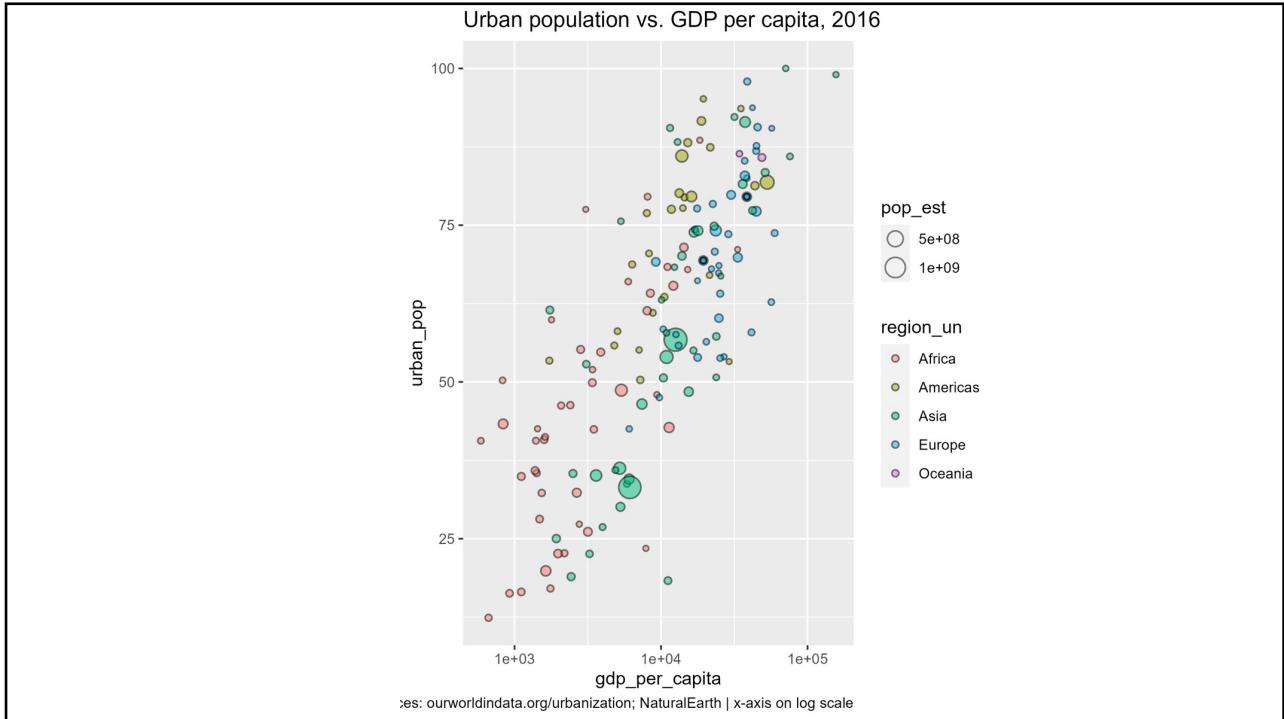
24



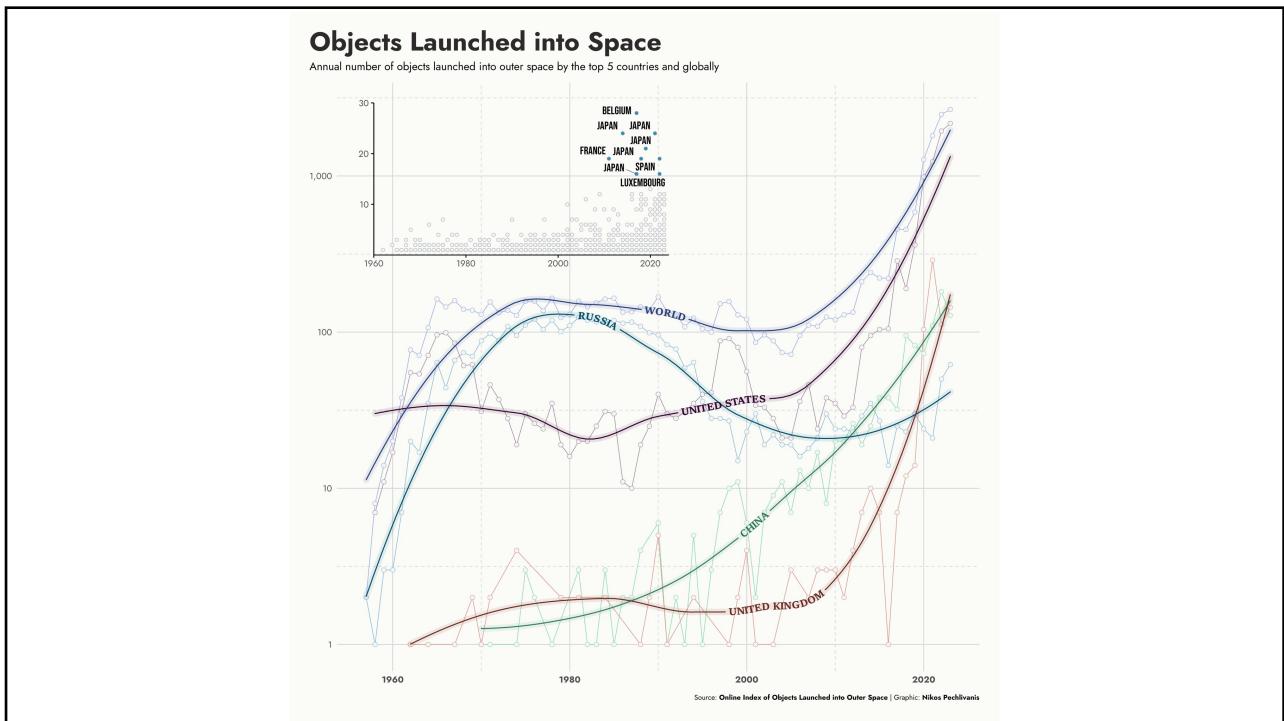
25



26

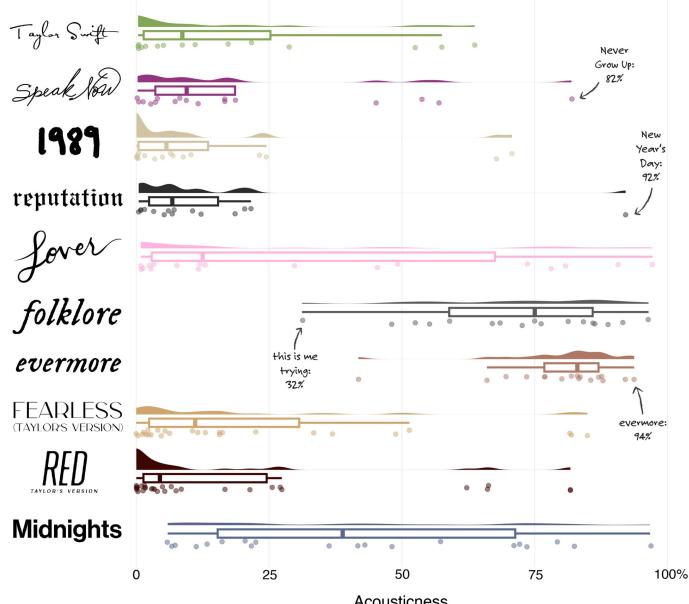


27



28

How does acousticness vary across Taylor Swift albums?

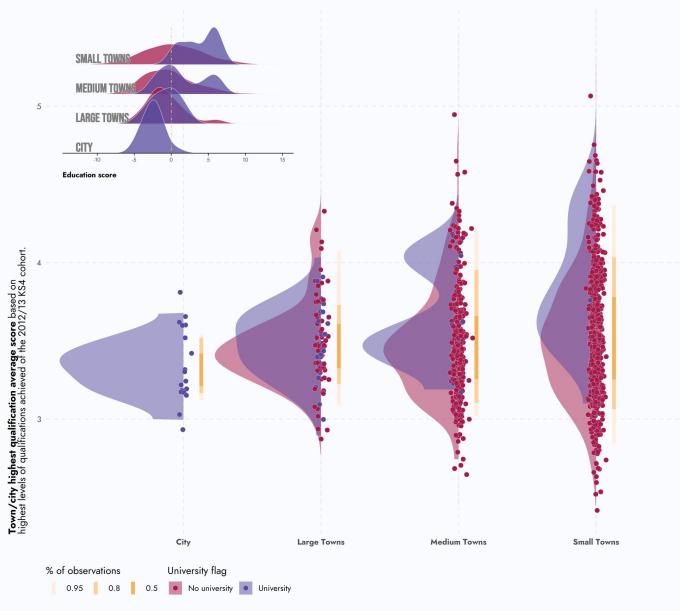


Elmera Azadpour | Data: (taylor) R package

29

Educational attainment of young people in English towns

Elevated town/city education score and higher average qualification attainment among young residents in English small towns hosting a university



30

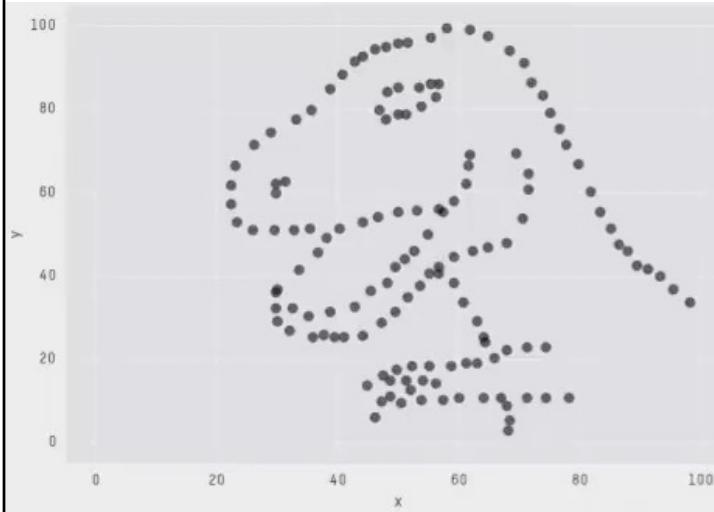
Take home 2:

**A good figure reveals patterns in data
that you as a researcher want to share.**

**A reader shouldn't need to look for that pattern.
Ideally, it jumps off the screen.**

31

Explore Plot your data

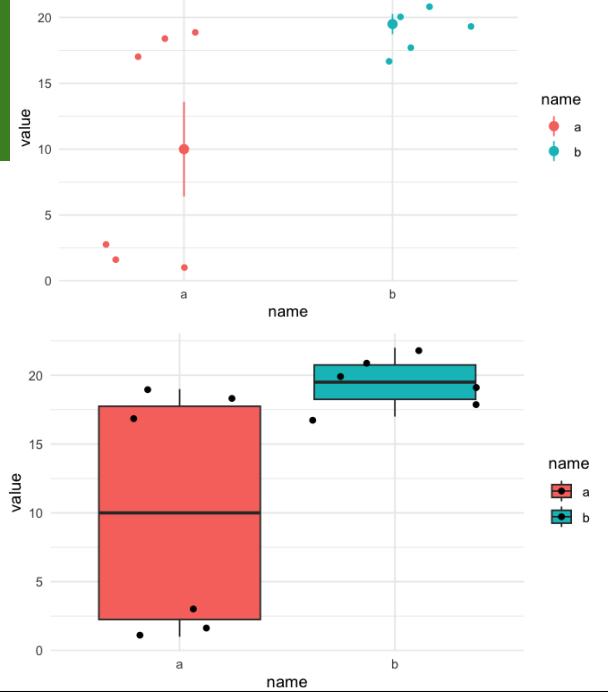
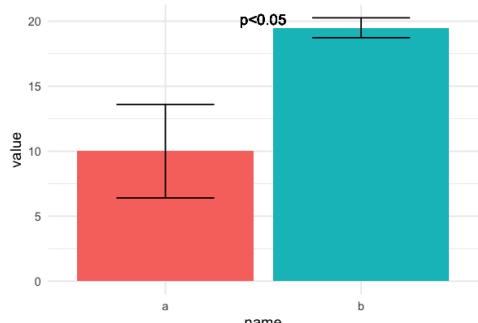


X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

32

How to fraud with barcharts

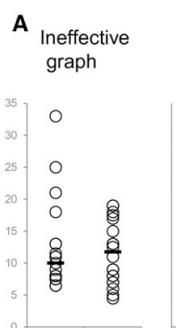
Would you believe this barchart in a paper?



33

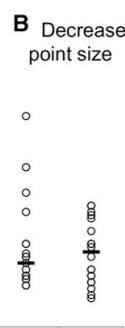
Ineffective → Effective

A Ineffective graph

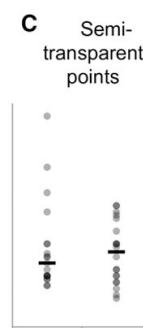


Step 1:
Make all
points
visible

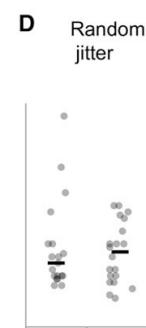
B Decrease point size



C Semi-transparent points



D Random jitter



E Symmetric jitter



F Ineffective graph

G Increase width

H Emphasize summary statistics

34

17

VISUALIZING DATA

Error bars hide:

- Outliers
- Distribution shape
- Bi- or multi-modal distributions
- Asymmetric distributions

<https://twitter.com/sketchscience>



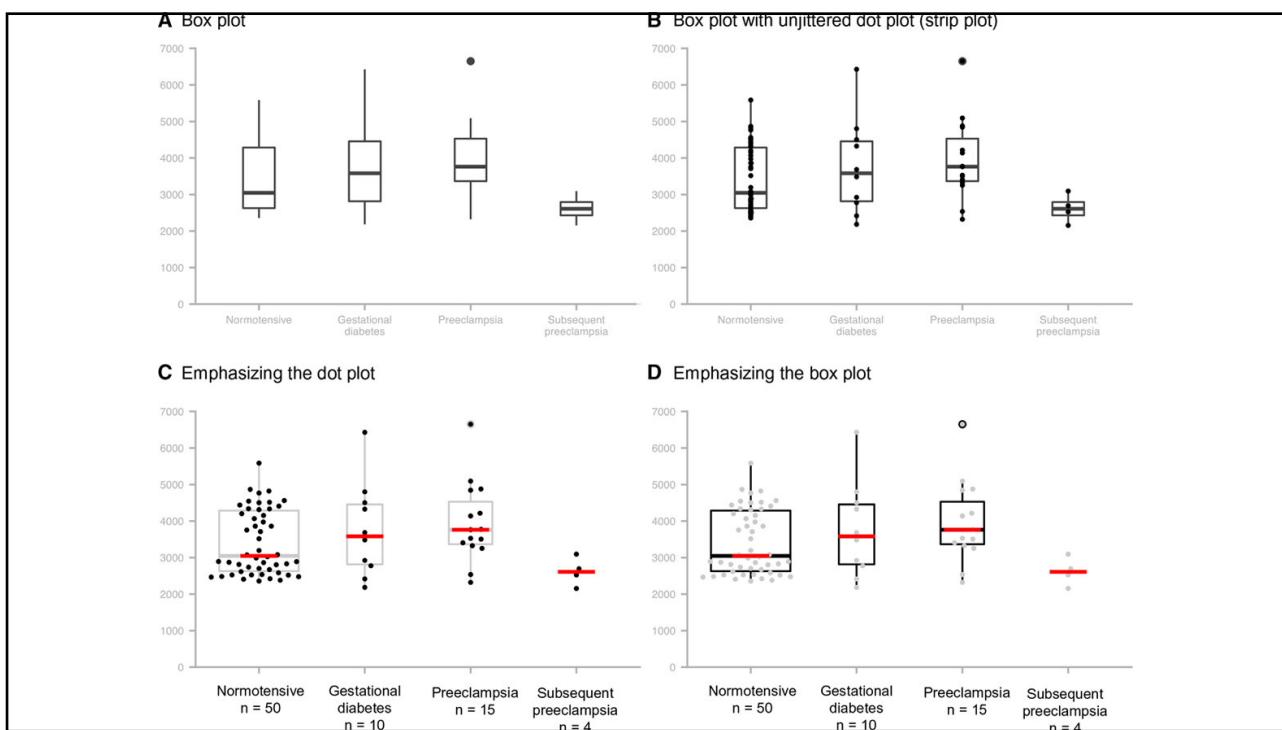
A bar chart showing 'Response' on the y-axis (0.0 to 10.0) and 'Treatment' on the x-axis (Control, Experiment 1, Experiment 2). The bars are colored red, green, and blue respectively. Error bars are present on each bar.



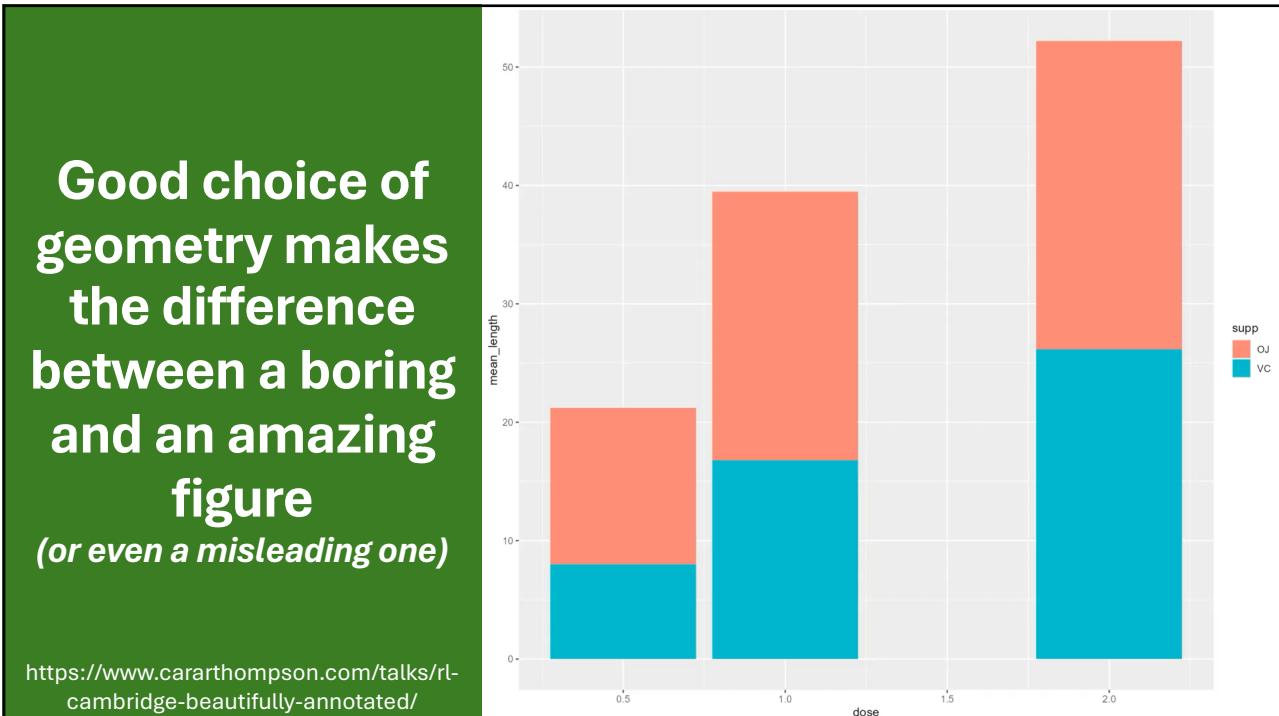
A box plot showing 'Response' on the y-axis (0.0 to 10.0) and 'Treatment' on the x-axis (Control, Experiment 1, Experiment 2). The boxes are colored red, green, and blue respectively. The box plot includes whiskers, a median line, and individual data points (dots).

Sketching Science

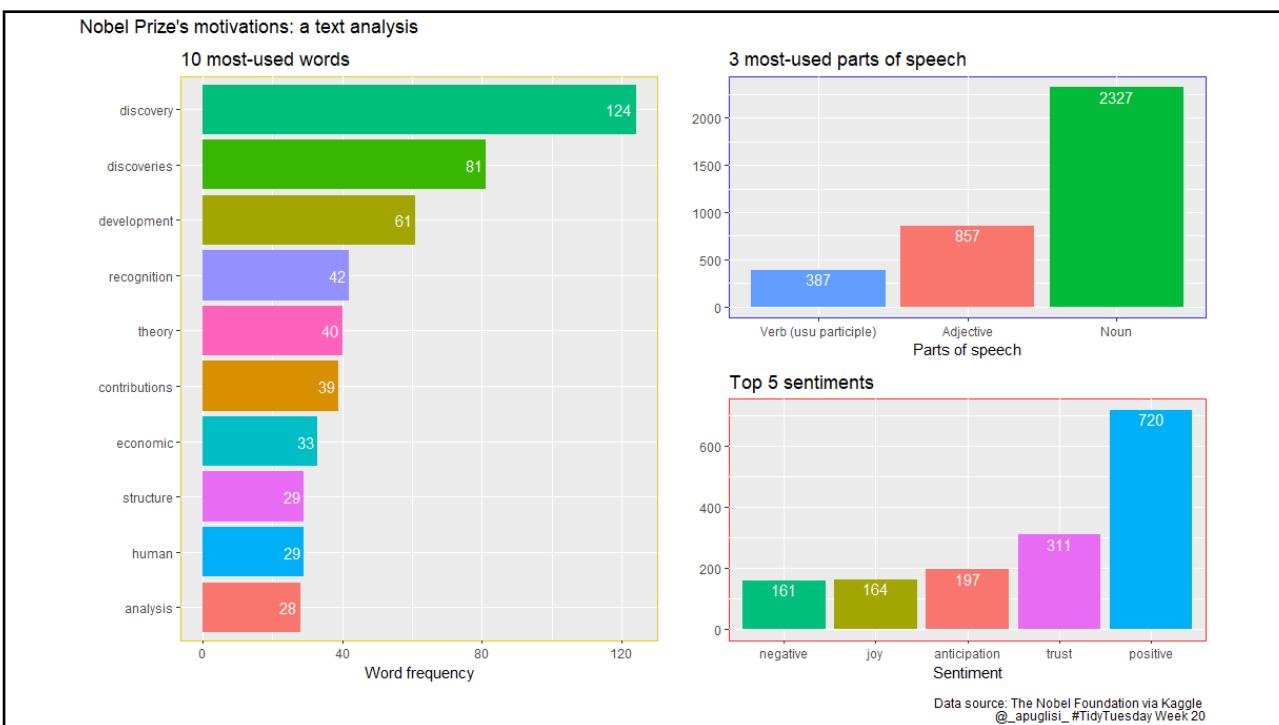
35



36



37



38

19

Take home 3:

Find a visualisation that shows:

Firstly, your main pattern effectively

Secondly, as much of the raw data as the image can handle

39

**GGplot
in a nutshell**

Guides & Legends

Themes

Facets

**Coordinate system
& zoom**

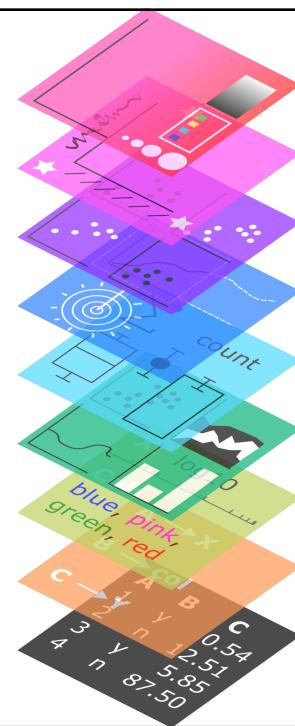
Statistics

Geometries

Scales

**Mapping /
aesthetics**

Data



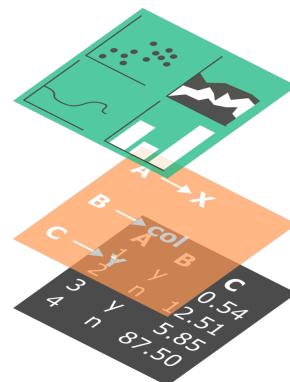
40

GGplot in a nutshell

Geometries

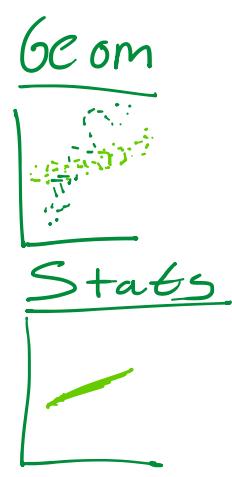
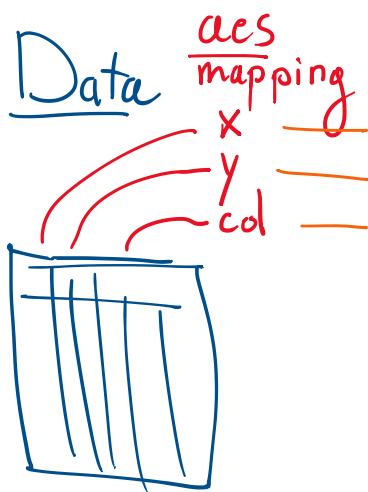
**Mapping /
aesthetics**

data



41

Aesthetic mapping links your data to your plot

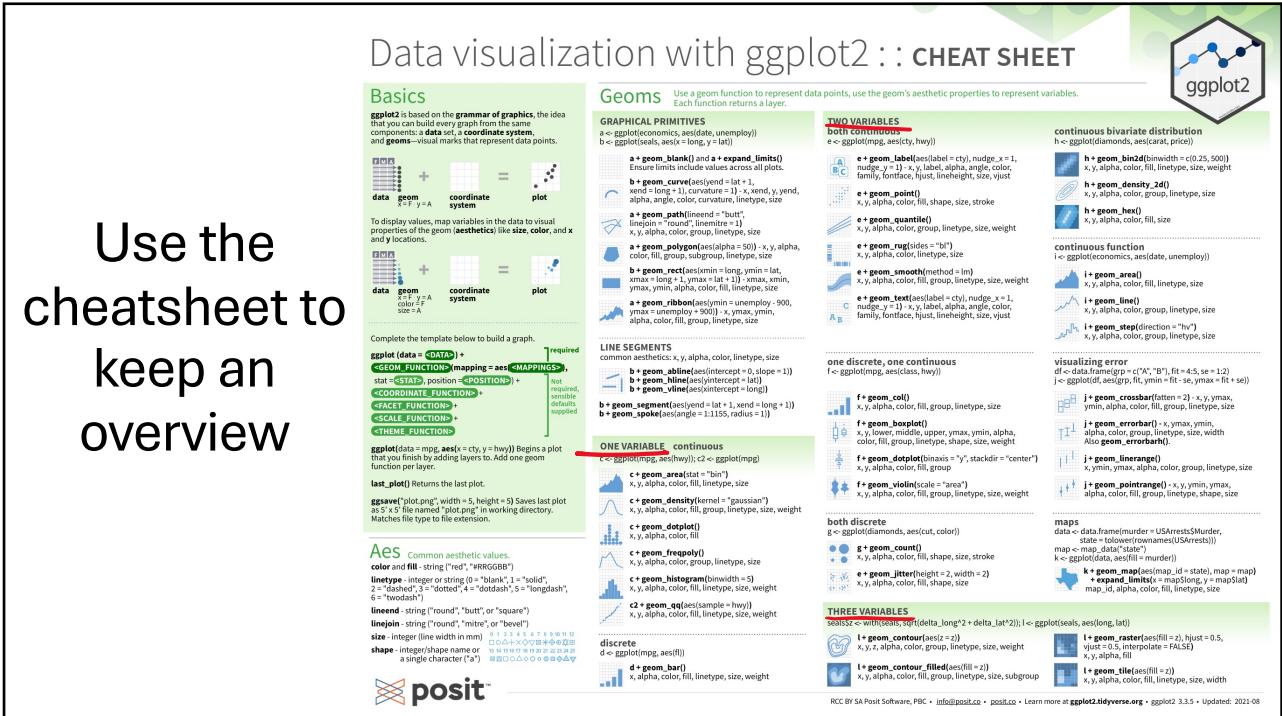


A → x
B → col
C → y

42

21

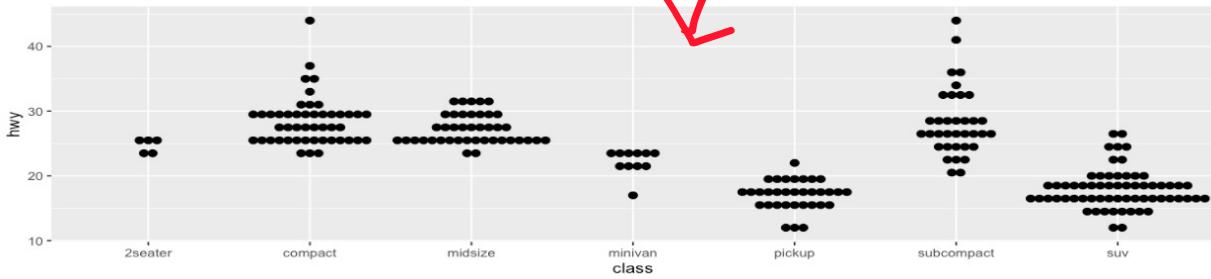
Use the
cheatsheet to
keep an
overview



43

Cheat-sheet has example-
code to get started

```
```{r}
f <- ggplot(mpg,aes(class,hwy))
f + geom_dotplot(binaxis = 'y',stackdir = 'center')
```
```



44

To R studio!

45

When making a figure

Ask yourself, and your colleagues, often: **What do I want my reader to take from this figure**

And can you plot that in a more direct way?

46

What's not in this workshop

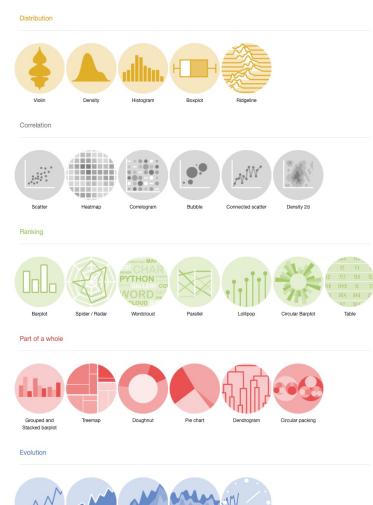
- clear labels and legends
 - colour usage
 - scales
 - themes
 - exporting specifically for publication
- and much more

--> Next ggplot course for beginners and novices in November!

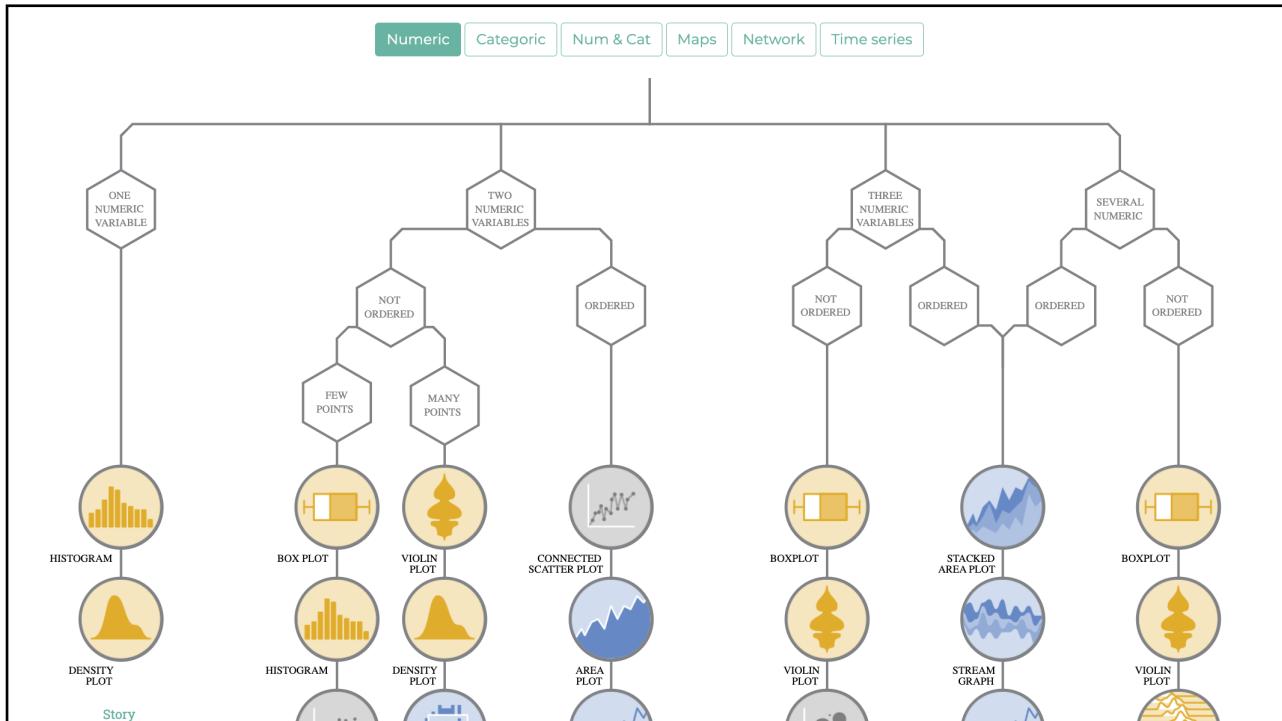
47

Choosing a geometry or statistic

- <https://r-graph-gallery.com/>
- <https://ggplot2.tidyverse.org/>
- <https://www.data-to-viz.com>
- <https://datavizcatalogue.com>



48



49



BOXPLOT
Summarize the distribution of numeric variables

About

A **boxplot** gives a nice summary of one or several numeric variables. The line that divides the box into 2 parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers.

Common Mistakes

- Boxplot hides the sample size of each group, [show it with annotation or box width](#).
- Boxplot [hides the underlying distribution](#). Use jitter if low number of data points, or use violin with bigger data.
- [Order your boxplot by median](#) can make it more insightful.

Code

R graph gallery | Python | D3.js | React



NOT ORDERED



VIOLIN PLOT

BUBBLE PLOT

3D SCATTER OR SURFACE

ORDERED



STREAM GRAPH

LINE PLOT

AREA (SM)

NOT ORDERED



VIOLIN PLOT

RIDGE LINE

PCA

50