

Marketing & Data Analysis

Visualizations + EDA

Santiago Alonso-Díaz

Tecnológico de Monterrey
EGADE, Business School

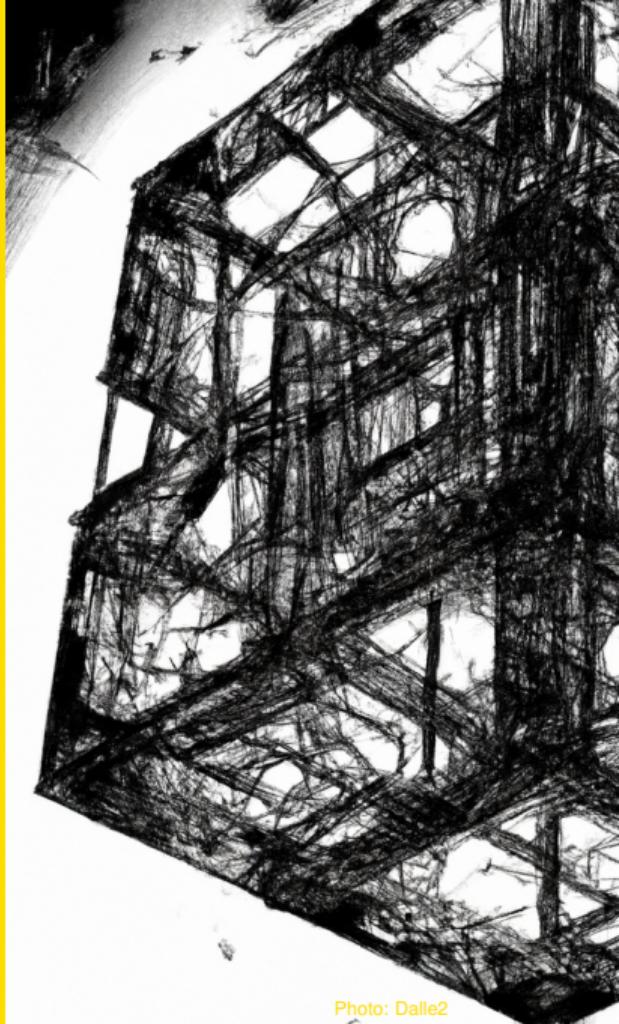


Photo: Dalle2

What to visualize/explore?

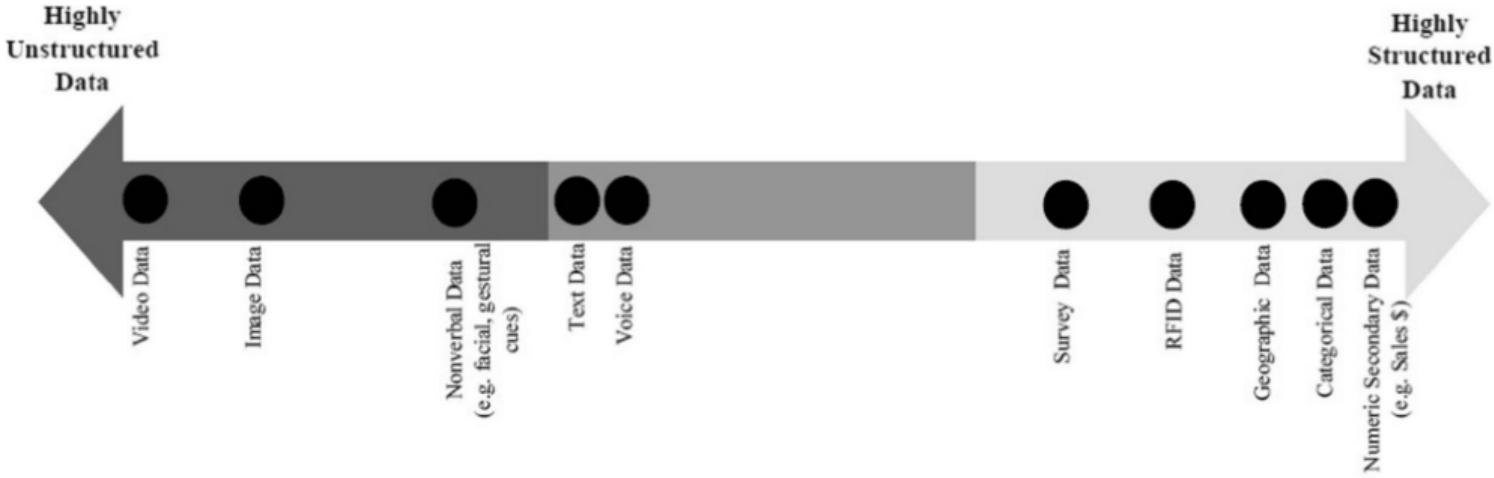


Figure: Balducci and Marinova, 2018

Table of contents

1 Exploratory Data Analysis

- Setup and cleaning
- Know your data (plots + stats)

2 Visualizations

3 References

Exploratory Data Analysis

EDA involves knowing your data, cleaning it, statistics and plotting techniques i.e.
this course

But let's see basic tips in [EDA](#)

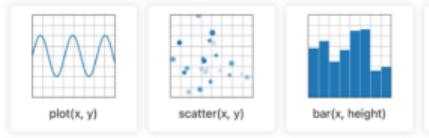
Visualizations

How to visualize?

Many types of charts.

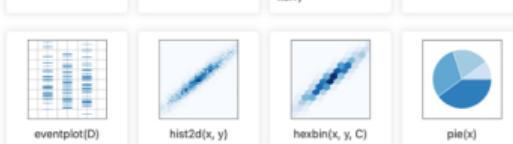
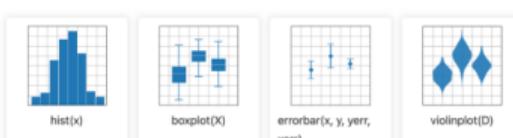
Pairwise data

Plots of pairwise (x, y) , tabular ($\text{var_0}, \dots, \text{var_n}$), and functional $f(x) =$



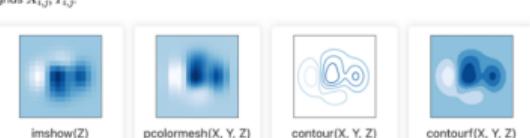
Statistical distributions

Plots of the distribution of at least one variable in a dataset. Some of these methods also compute the distributions.



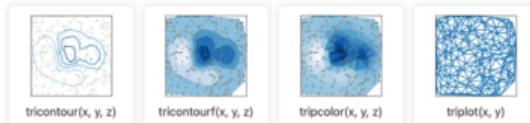
Gridded data:

Plots of arrays and images $Z_{i,j}$ and fields $U_{i,j}, V_{i,j}$ on regular grids and corresponding coordinate grids $X_{i,j}, Y_{i,j}$.



Irregularly gridded data

Plots of data $Z_{x,y}$ on unstructured grids, unstructured coordinate grids (x, y) , and 2D functions $f(x, y) = z$.



3D and volumetric data

Plots of three-dimensional (x, y, z) , surface $f(x, y) = z$, and volumetric $V_{x,y,z}$ data using the `mpl_toolkits.mplot3d` library.

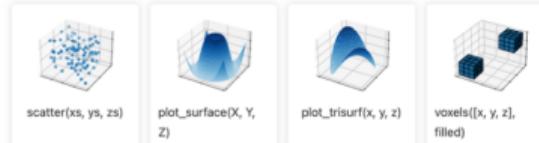


Figure: Matplotlib.org

Principles of visualization Wilke, 2019

Visualization needs to be proportional to the data values they represent.

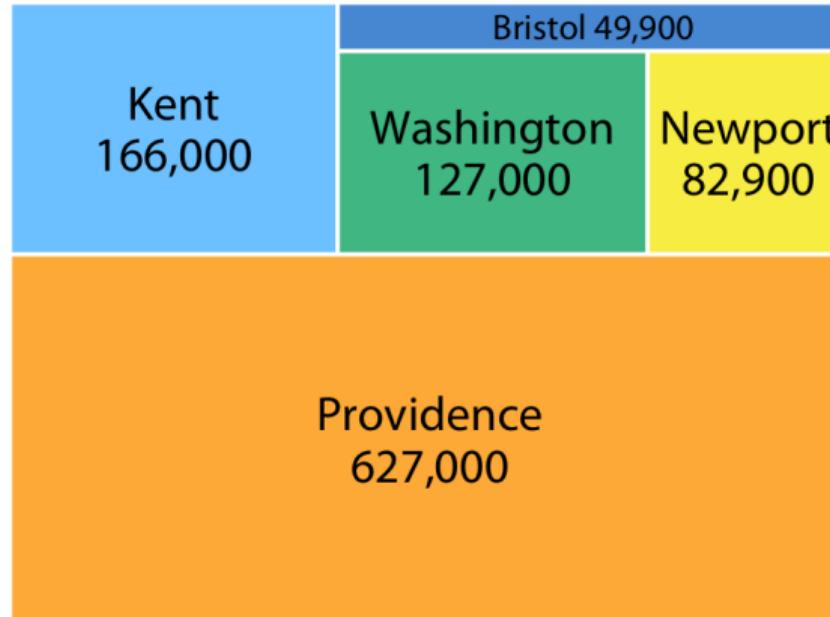


Figure: Wilke, 2019

Principles of visualization

Avoid overlaps e.g. with transparency and jittering

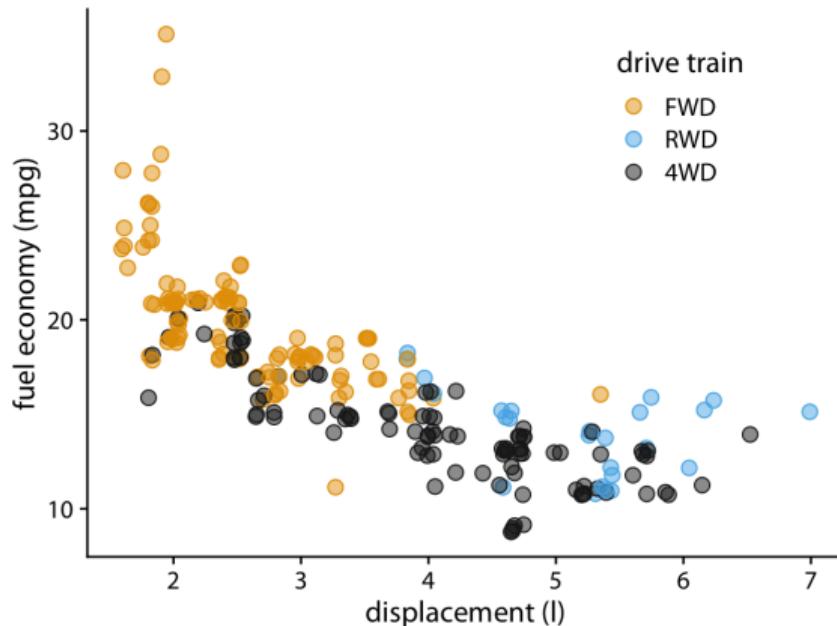


Figure: Wilke, 2019

Principles of visualization

Simple color scales. The rainbow scale looks bad and it is hard to read in this map.

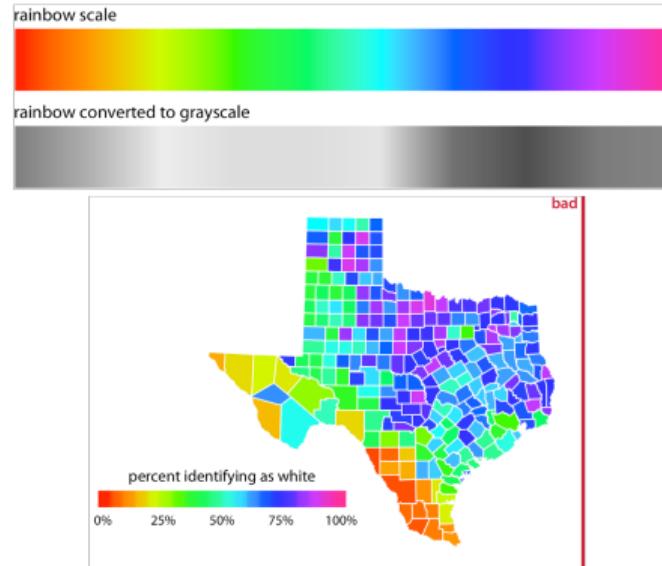


Figure: Wilke, 2019

Principles of visualization

Colorblind friendly schemes

original



deuteranomaly



protanomaly



tritanomaly



Figure: Wilke, 2019

Principles of visualization

Redundancy can be good. Here, color and shape enhance the difference in categories

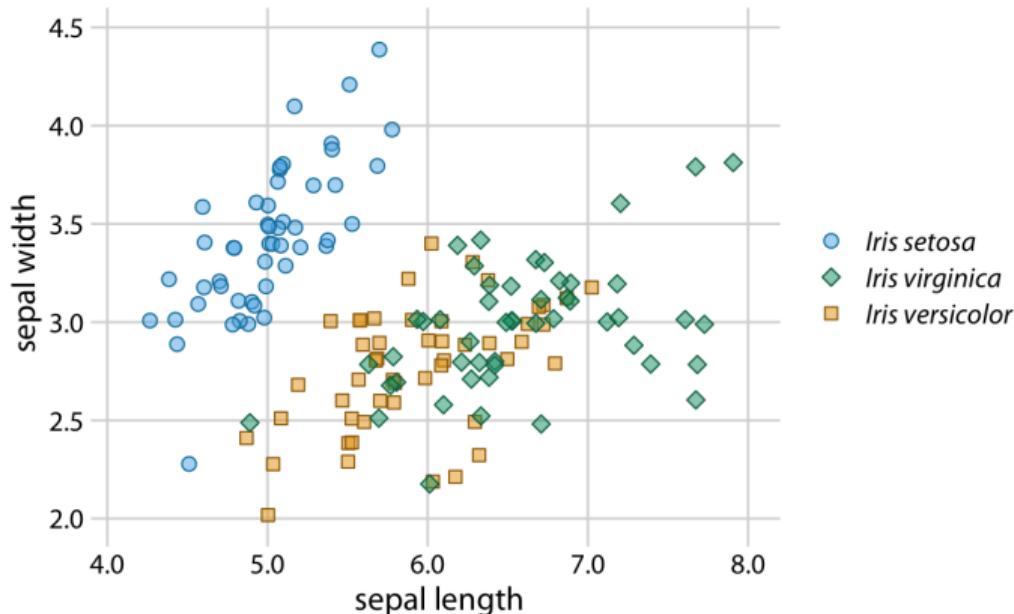


Figure: Wilke, 2019

Principles of visualization

Use consistent codes for the same categories across plots. Here the same color for each gender.

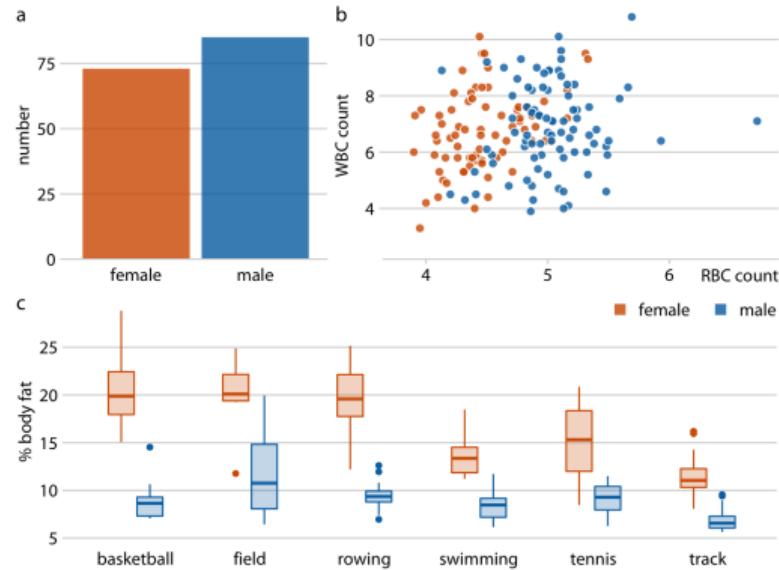


Figure: Wilke, 2019

Principles of visualization

Legends are not always necessary. Labels on the graphs are a good alternative.

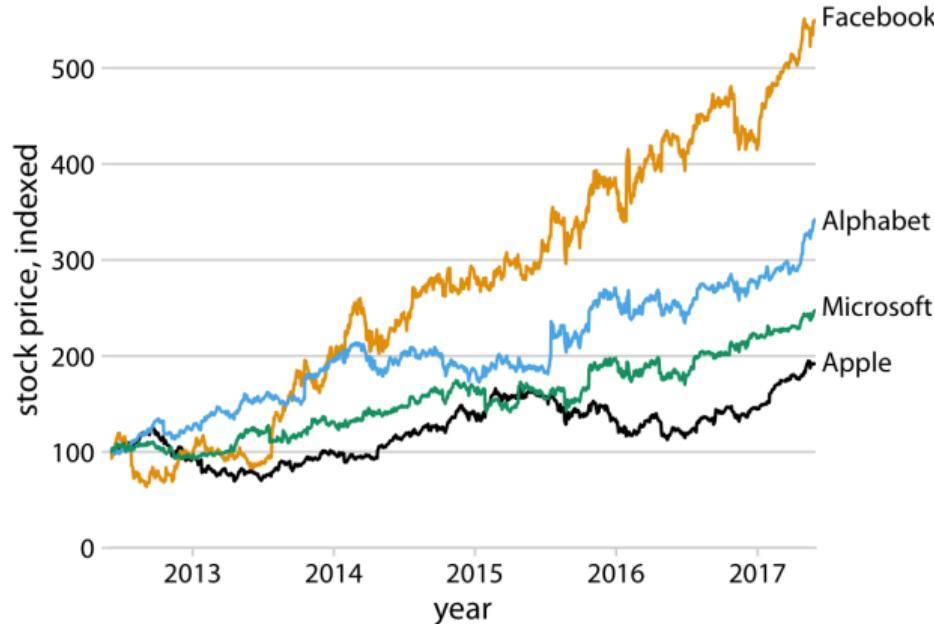


Figure: Wilke, 2019

Principles of visualization

Present heterogeneous results with multiple plots, with the same y-scale

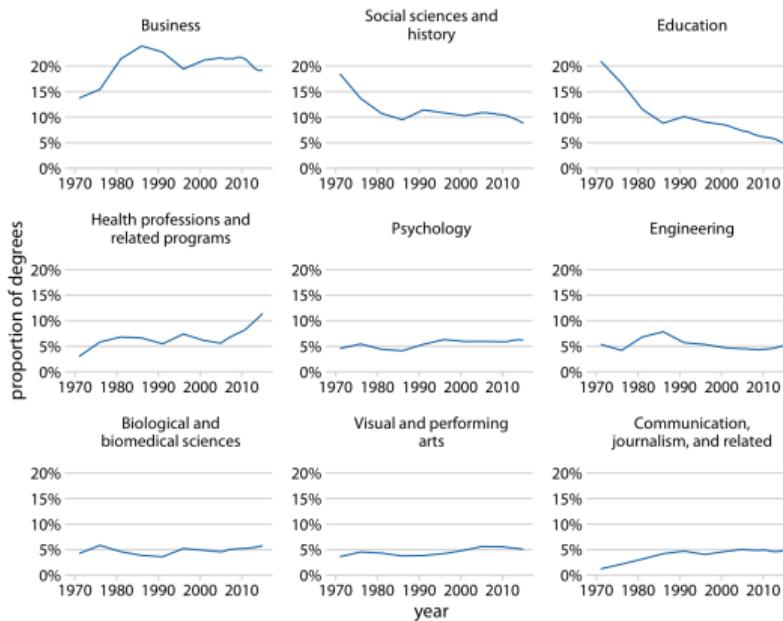


Figure: Wilke, 2019

Principles of visualization

Put title and axes labels

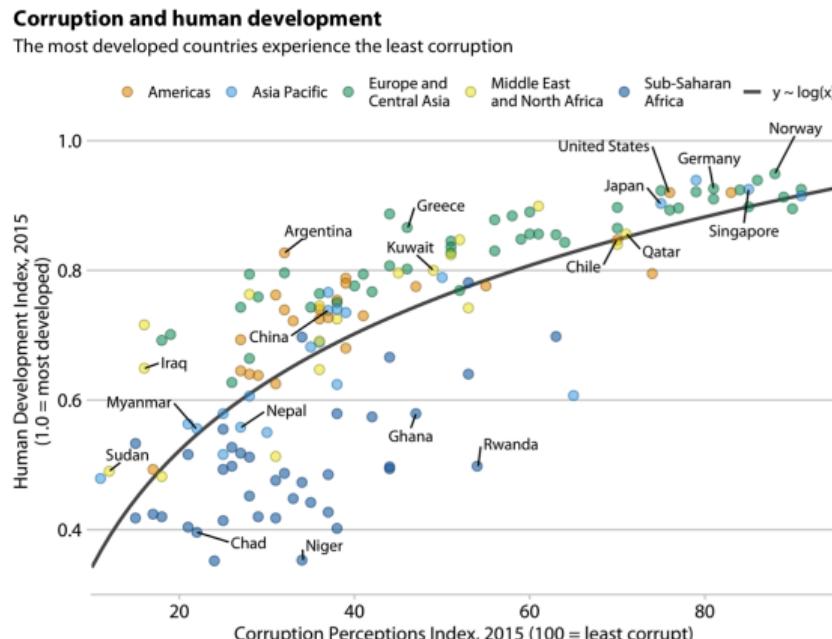


Figure: Wilke, 2019

Principles of visualization

Be generous with axes labels font size.

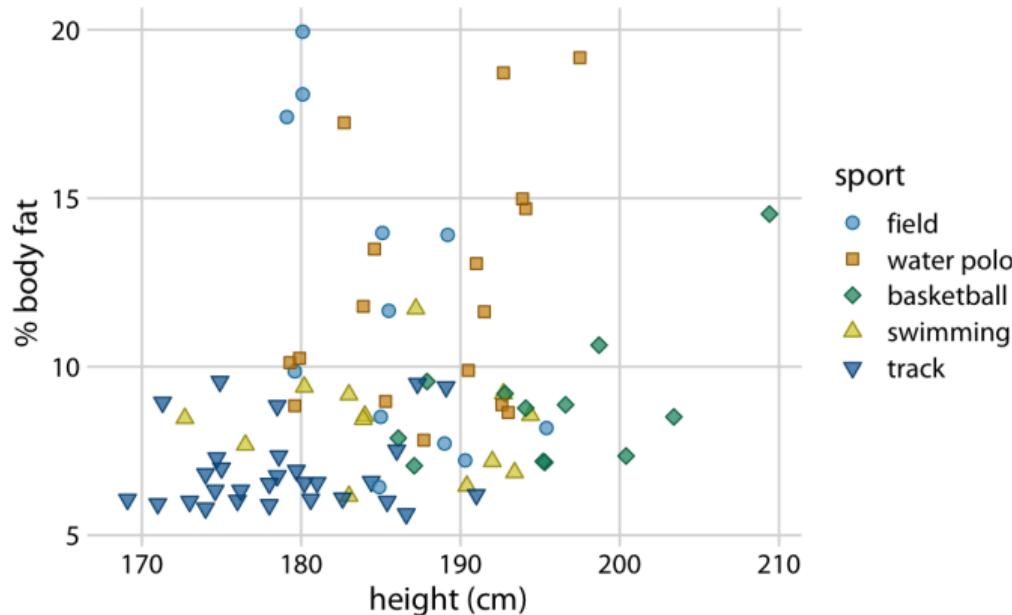


Figure: Wilke, 2019

Principles of visualization

Minimize ink devoted to non-data. The bottom plot respects more this principle than the top one. Drop grid lines?

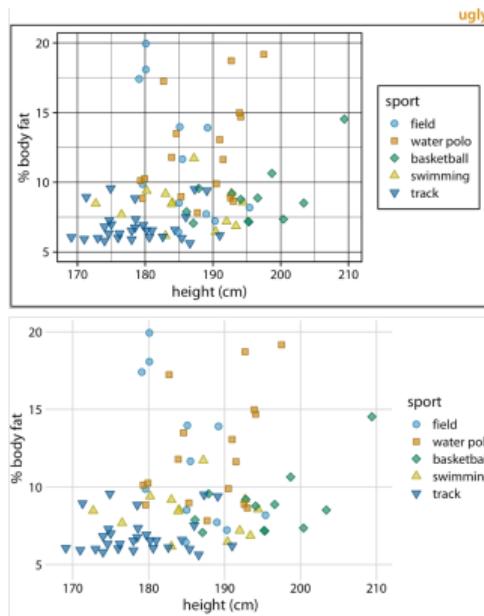
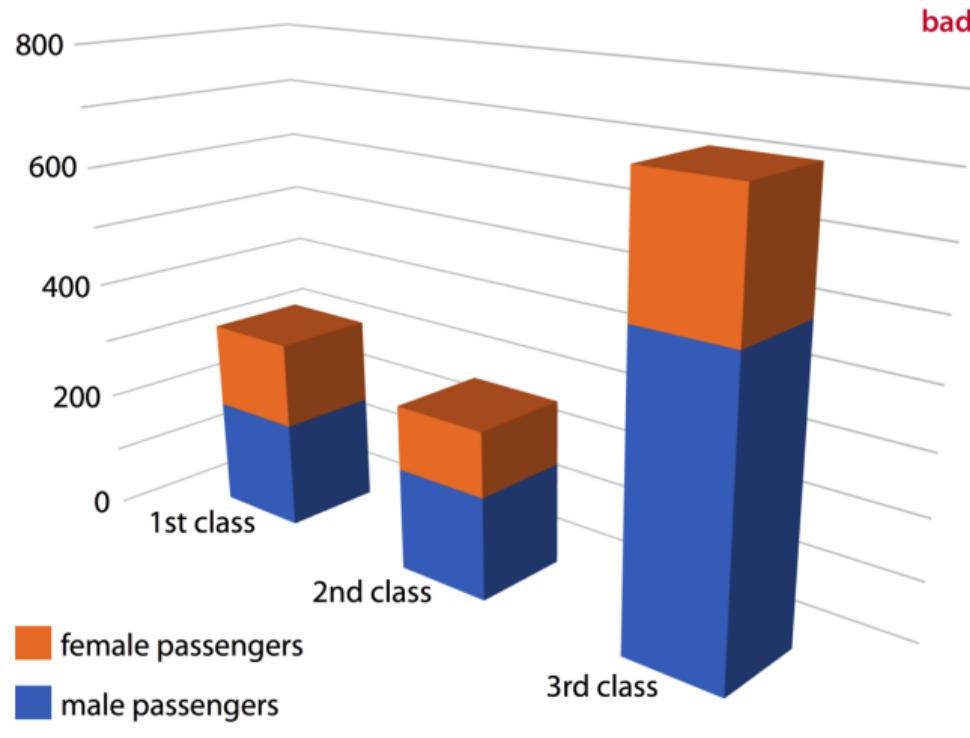


Figure: Wilke, 2019

Principles of visualization

Avoid 3D if unnecessary



Principles of visualization

If possible, tell a story. The left shows us that preprints stopped. The right why: another repository appeared.

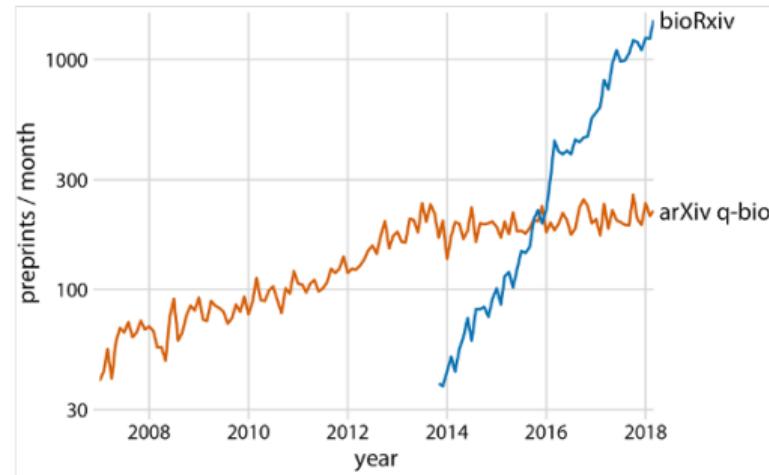
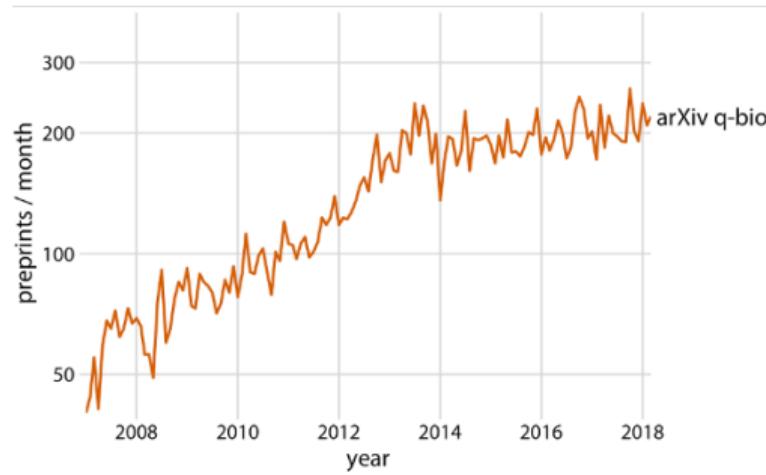


Figure: Wilke, 2019

Principles of visualization

When telling a story, prepare people for complex plots. This is simple (next slide the complex chart)

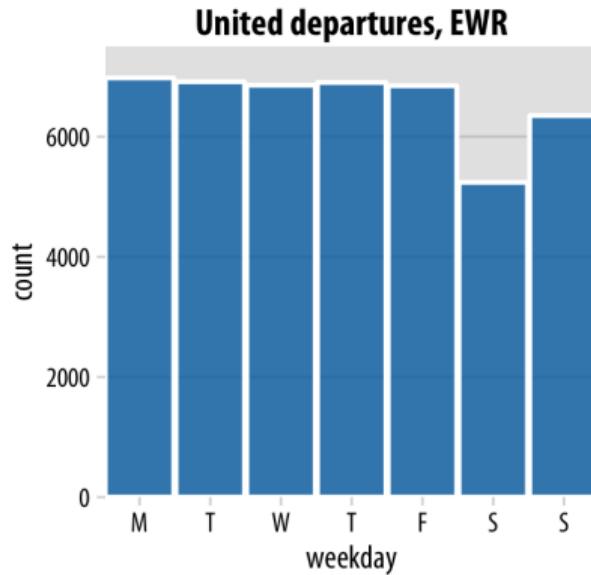


Figure: Wilke, 2019

Principles of visualization

This is more elaborated but similar to the simple one.

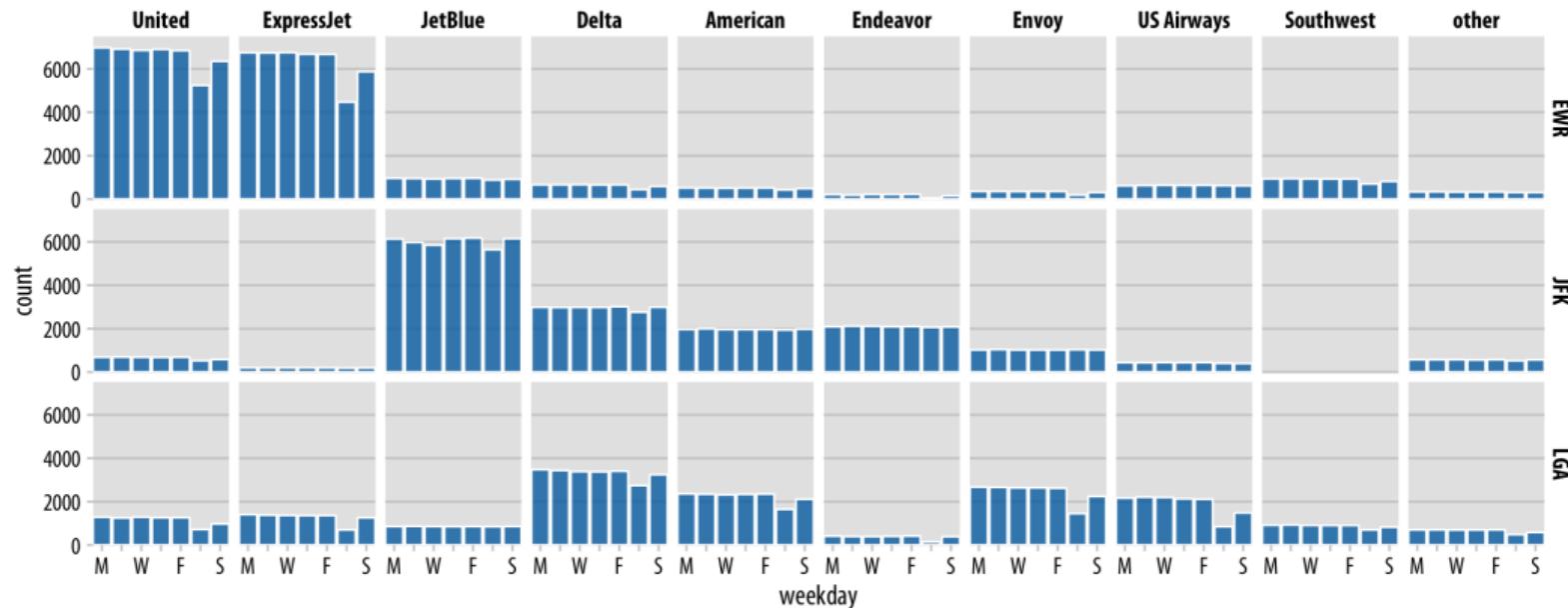


Figure: Wilke, 2019

Visualizations

Book: Wilke, 2019

Github repository: Wilke, 2019

References

-  **Balducci, B., & Marinova, D. (2018).** Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46, 557–590.
-  **Wilke, C. O. (2019).** *Fundamentals of data visualization: A primer on making informative and compelling figures*. O'Reilly Media.