

Data Visualization 5: Effective Charts

Stat 133 by Gaston Sanchez

Creative Commons Attribution Share-Alike 4.0 International CC BY-SA

“The ability to create effective charts and graphs has become almost as indispensable as good writing.”

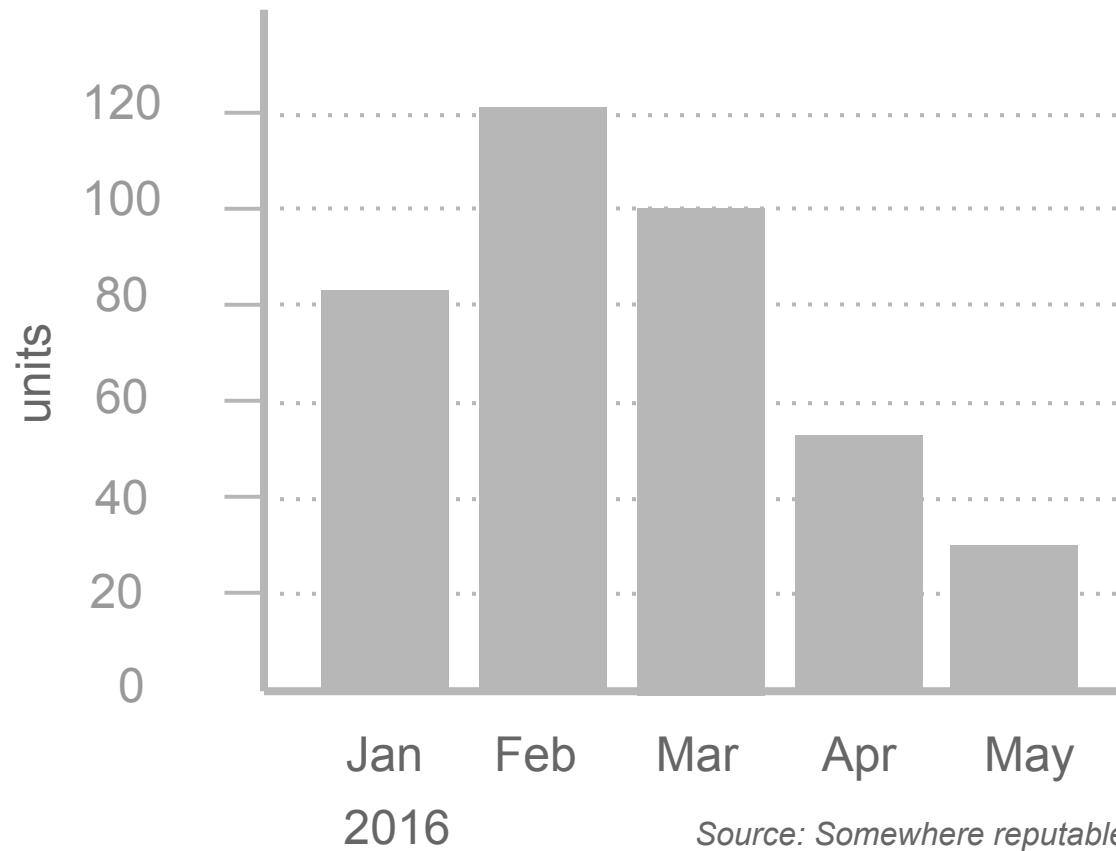
Donna Wong

Not yet :(

Graphical Elements and Some Vocabulary

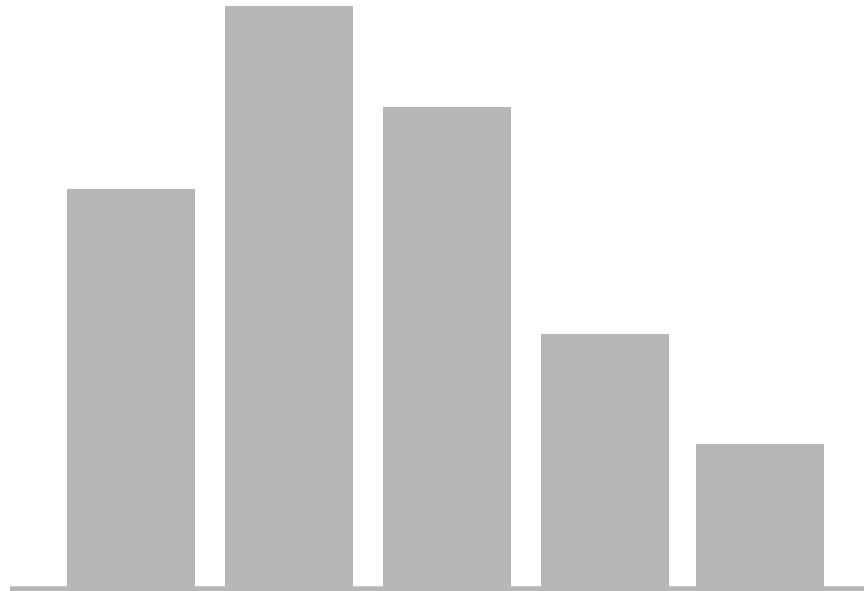
Title of this Graph

A description of the data or something worth highlighting to set the stage



Visual Cues

Encoding data with shapes, colors, and sizes. Which cues you choose depends on your data and your goals



Coordinate System

Mapping data requires a system of coordinates: cartesian, polar, etc



Scale

Increments that make sense can increase readability as well as shift focus



Title of this Graph

A description of the data or something worth highlighting to set the stage

units

Context

If your audience is unfamiliar with the data, it's your job to clarify what values represent and explain how people should read your plot

2016

Source: Somewhere reputable

Title of this Graph

A description of the data or something worth highlighting to set the stage

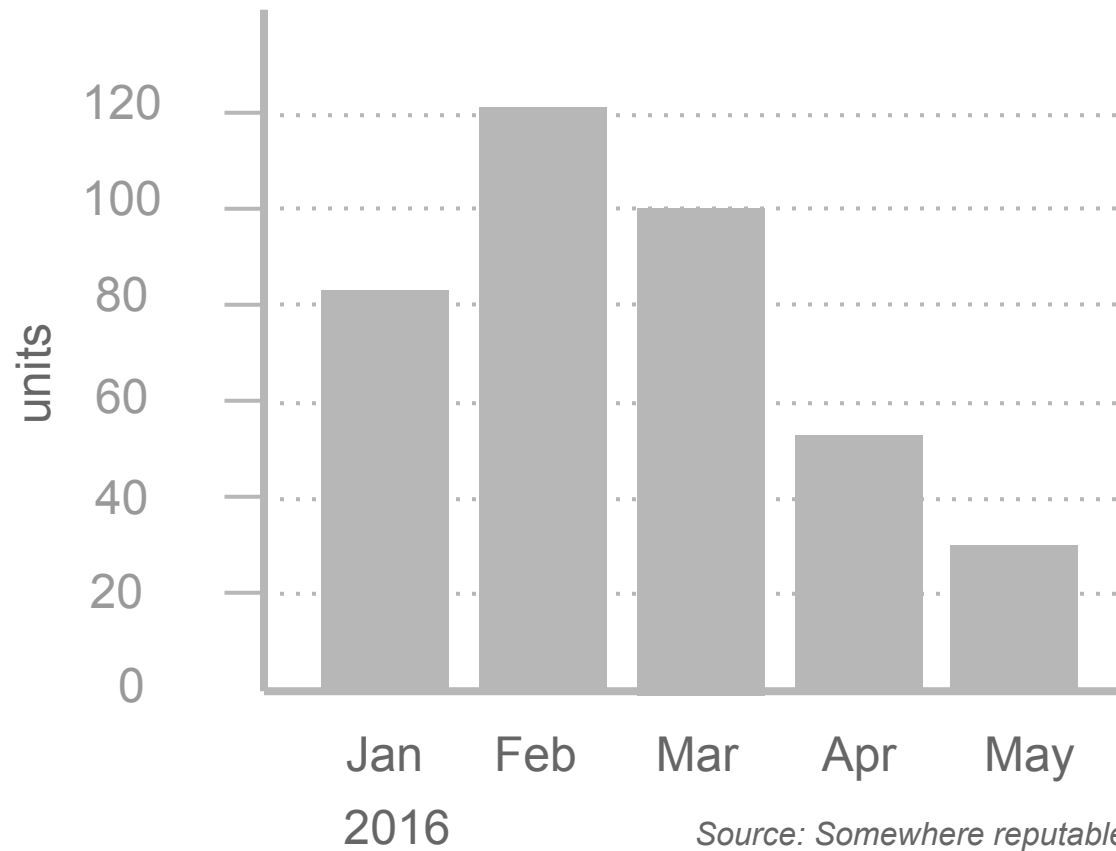
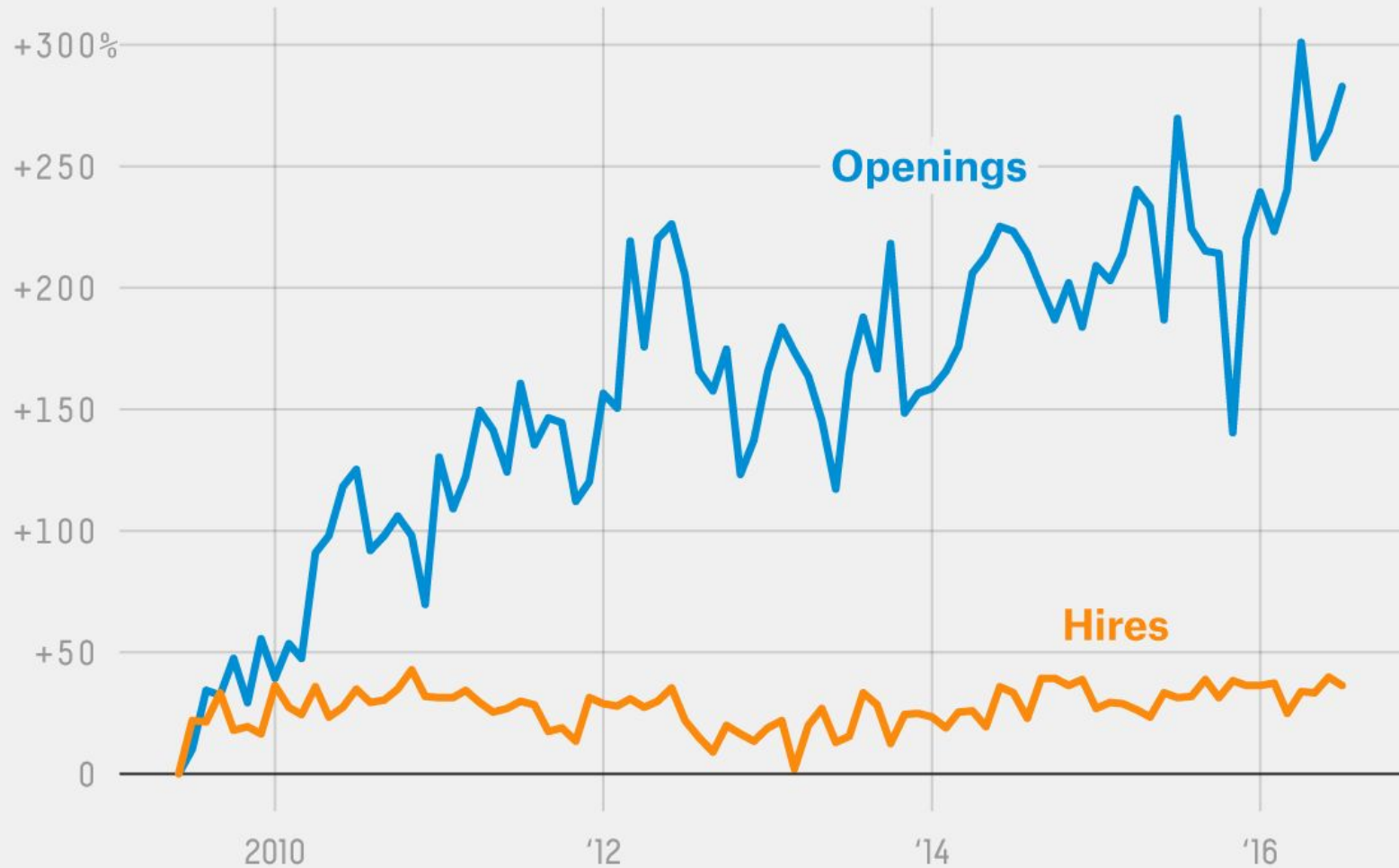


Chart Elements

Manufacturers are posting jobs, not filling them

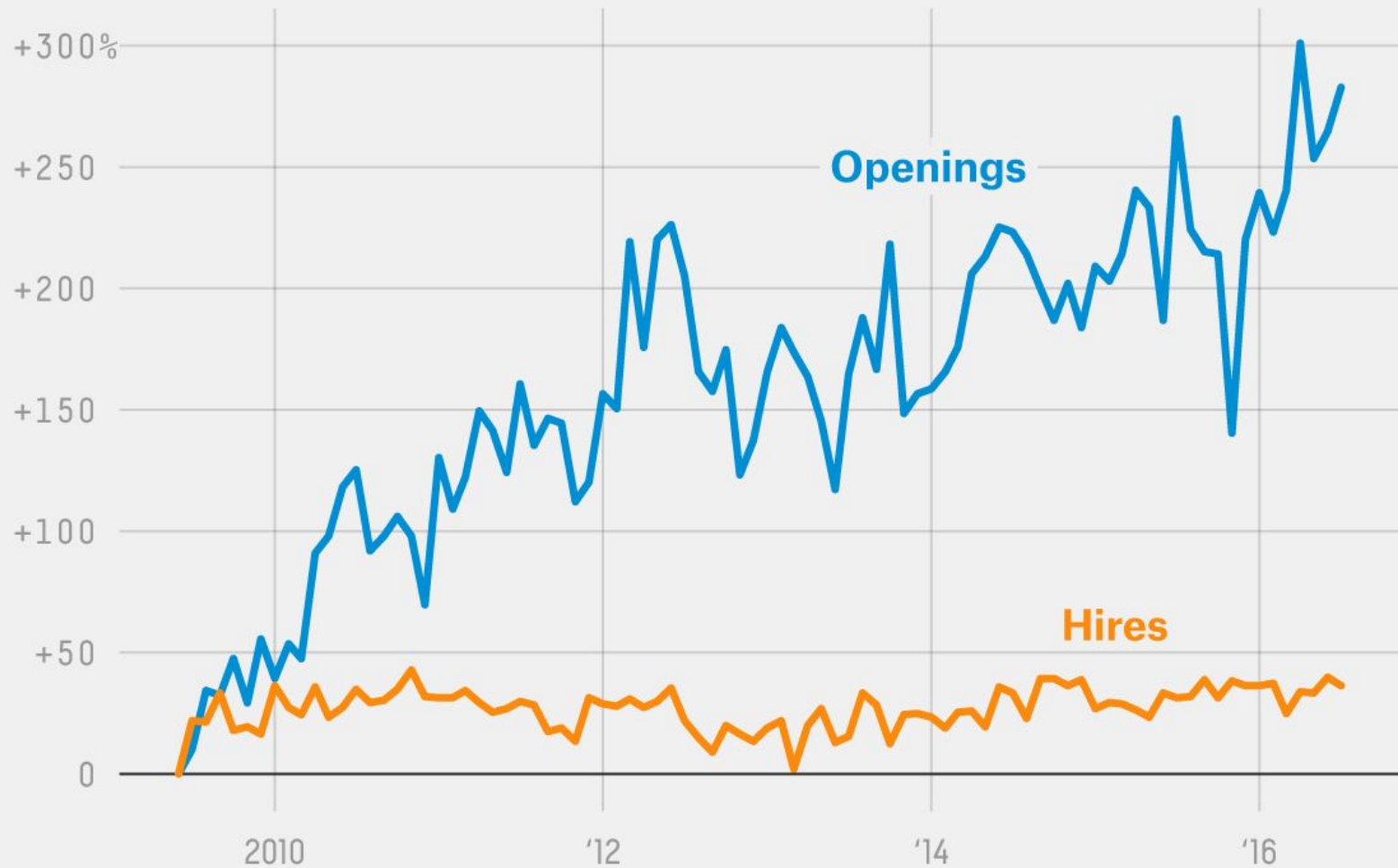
Change since June 2009, seasonally adjusted



Manufacturers are posting jobs, not filling them

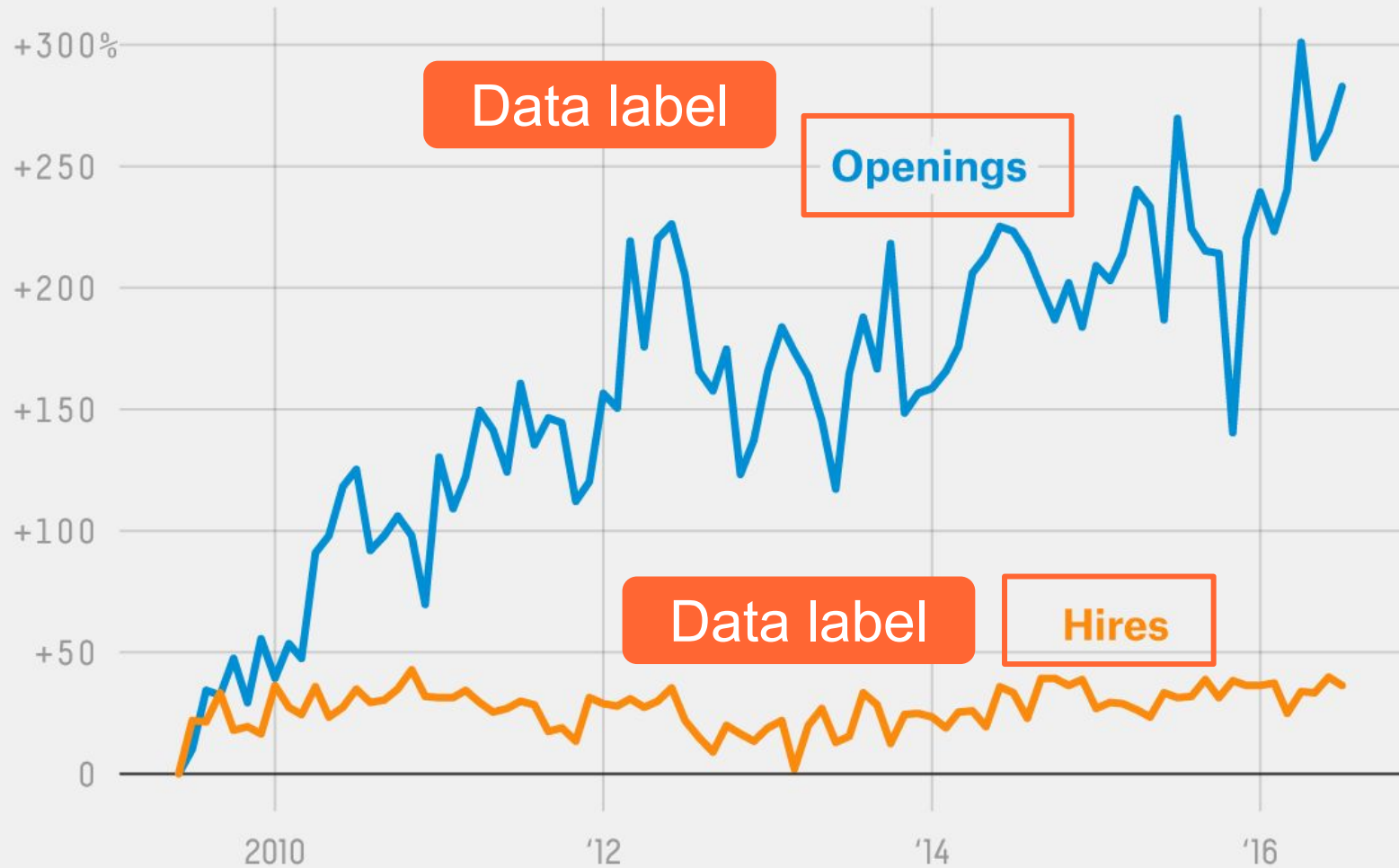
Change since June 2009, seasonally adjusted

title



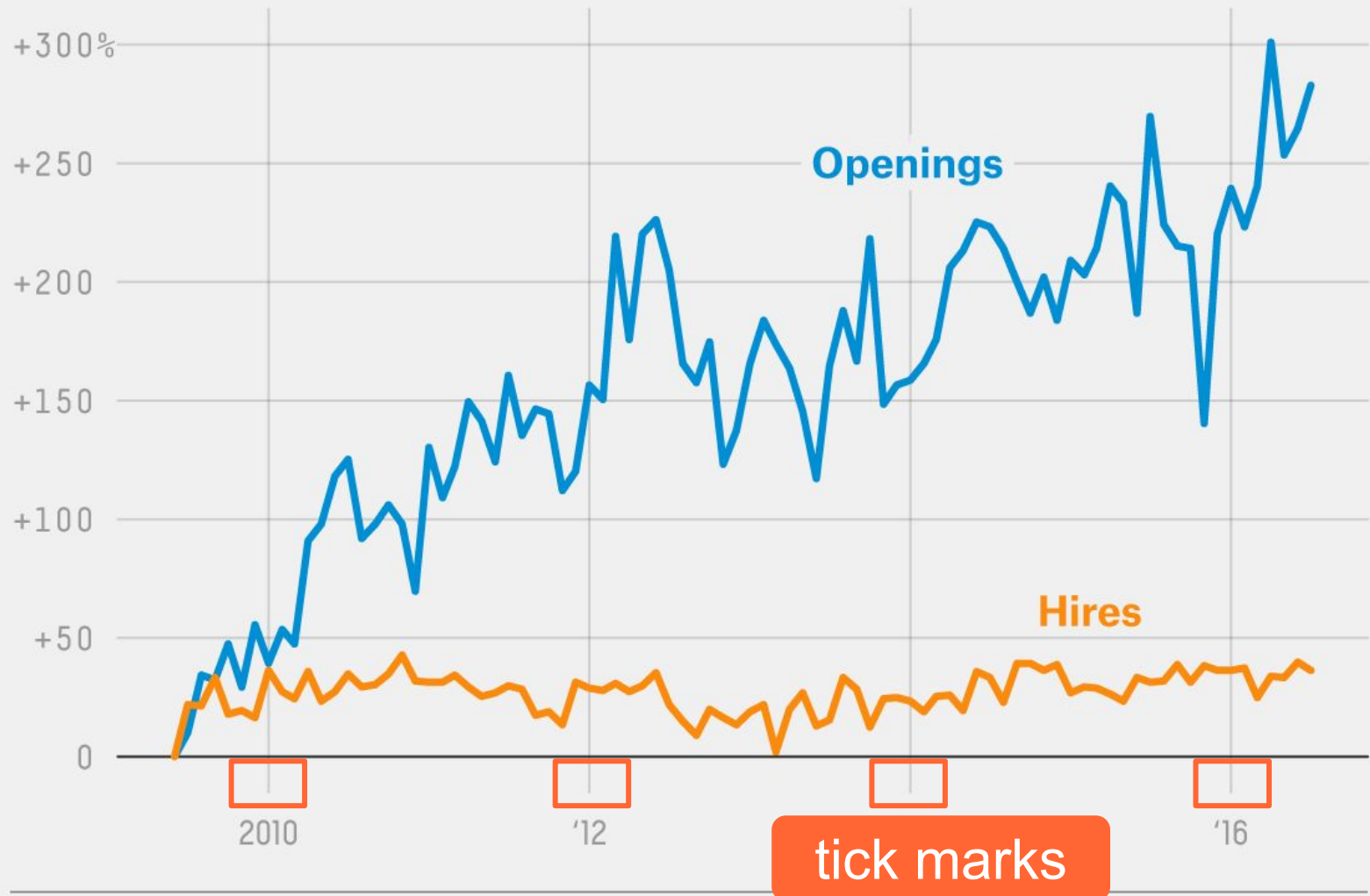
Manufacturers are posting jobs, not filling them

Change since June 2009, seasonally adjusted



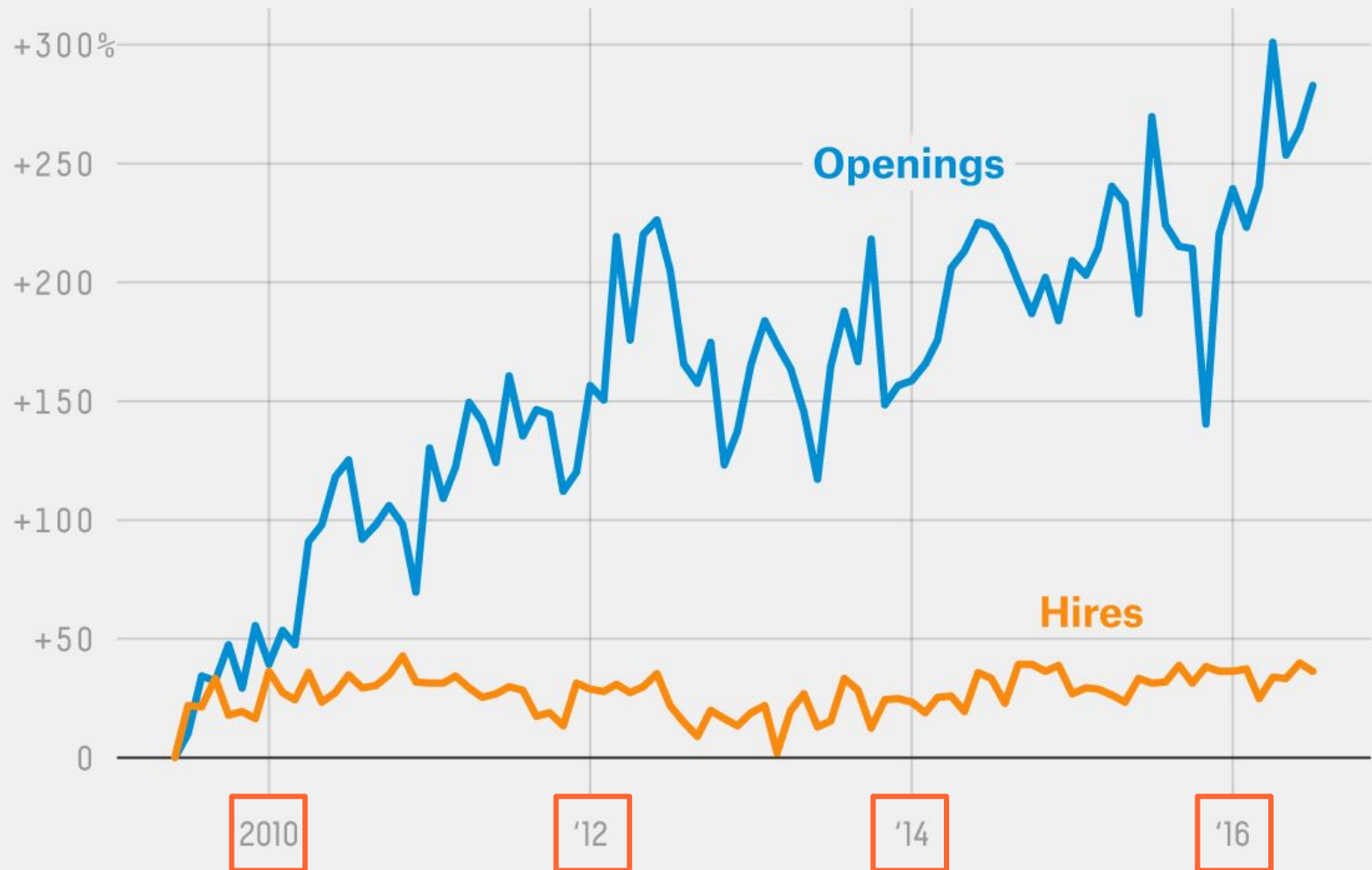
Manufacturers are posting jobs, not filling them

Change since June 2009, seasonally adjusted



Manufacturers are posting jobs, not filling them

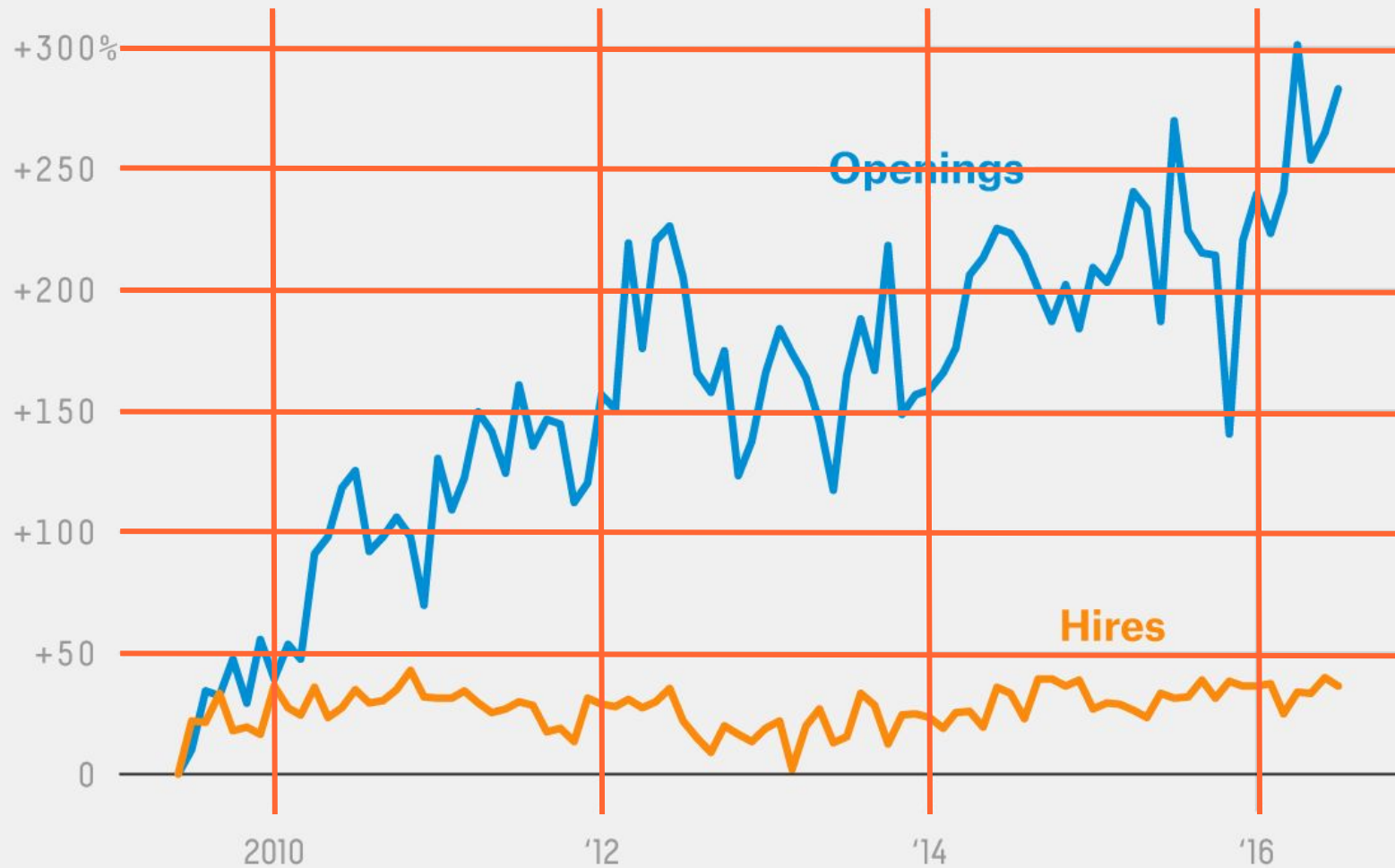
Change since June 2009, seasonally adjusted



Manufacturers are posting jobs, not filling them

Change since June 2009, seasonally adjusted

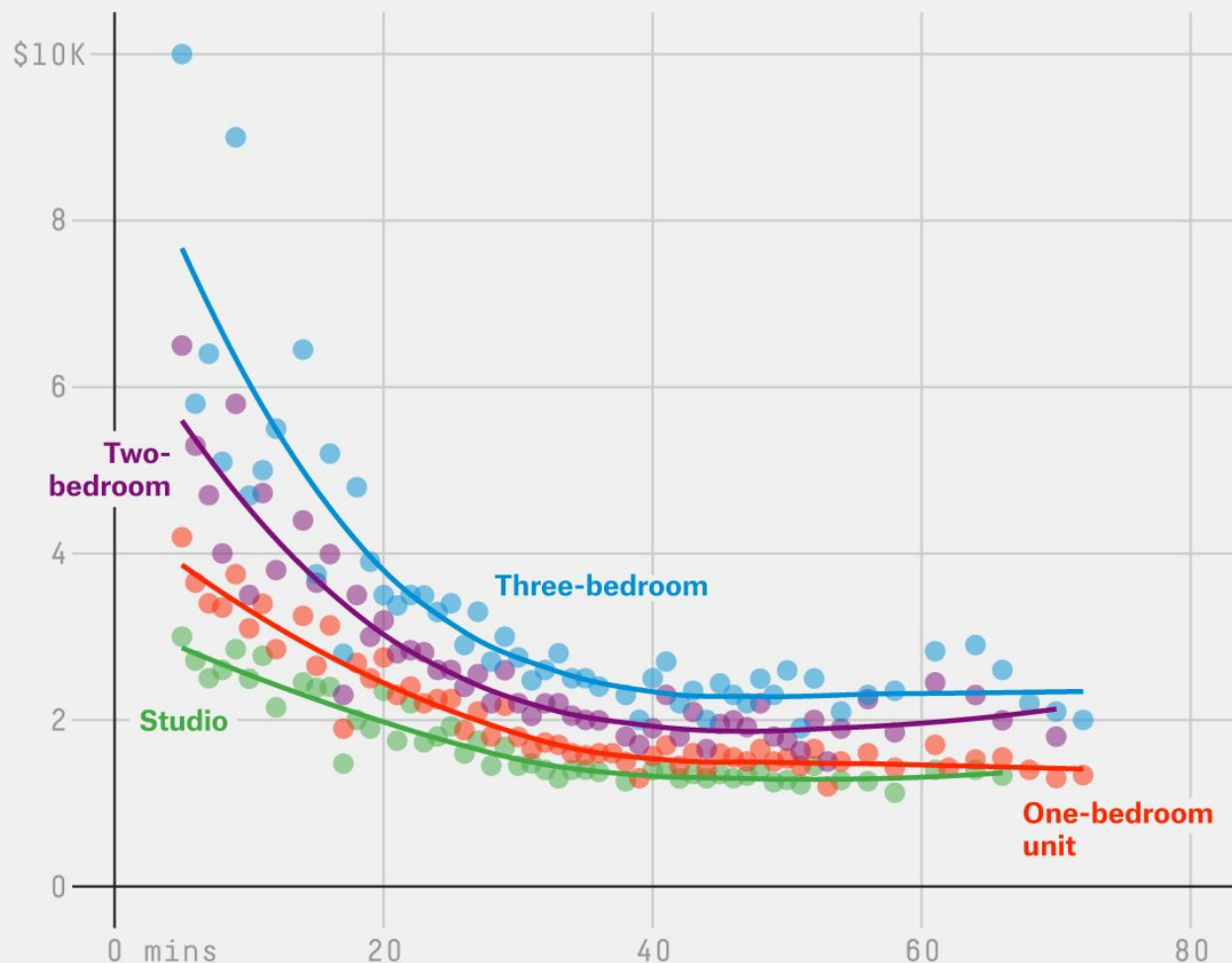
Grid lines



Another example

New Yorkers pay up for a shorter commute

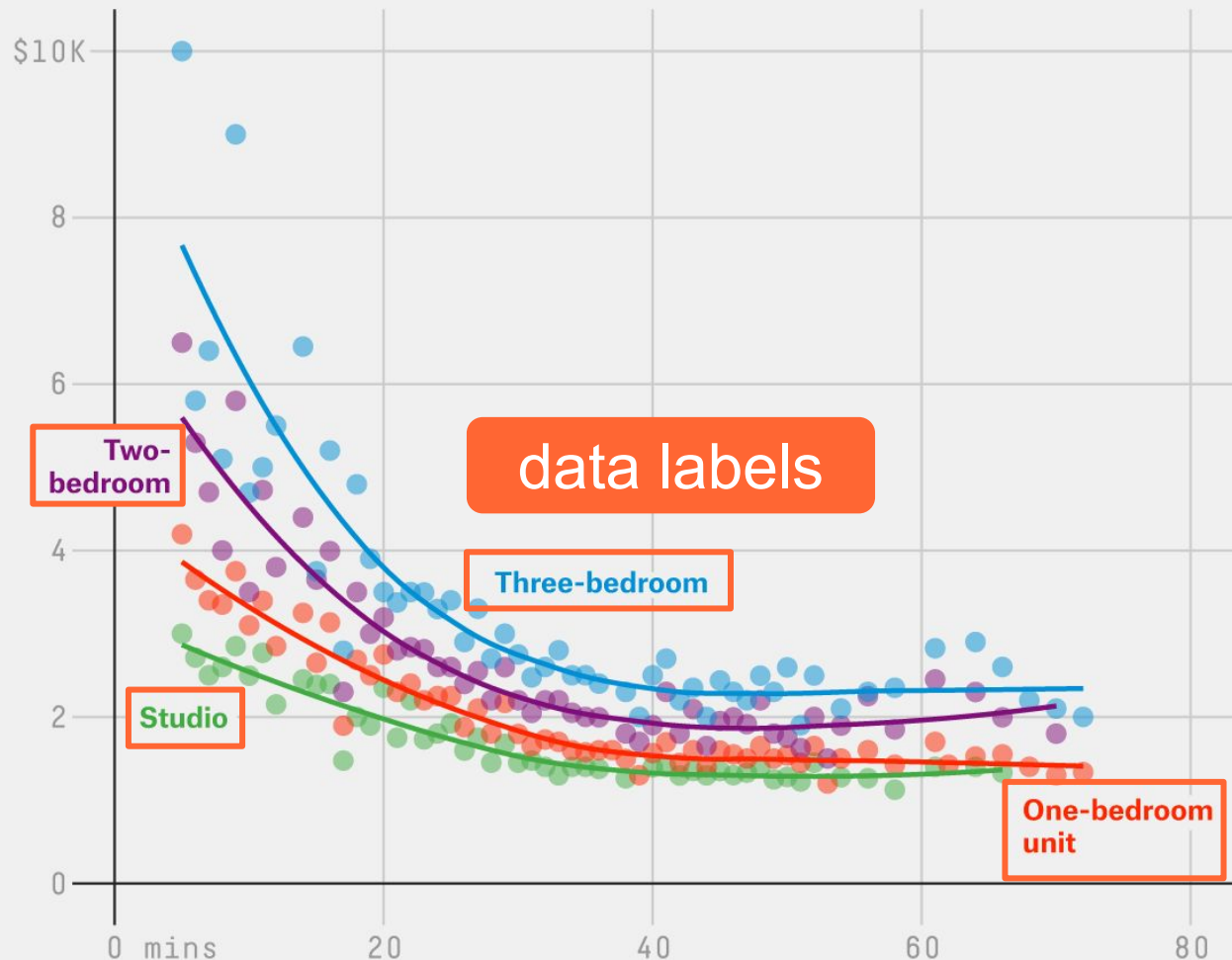
Median monthly NYC rent in 2015 vs. commute time by subway



Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

New Yorkers pay up for a shorter commute

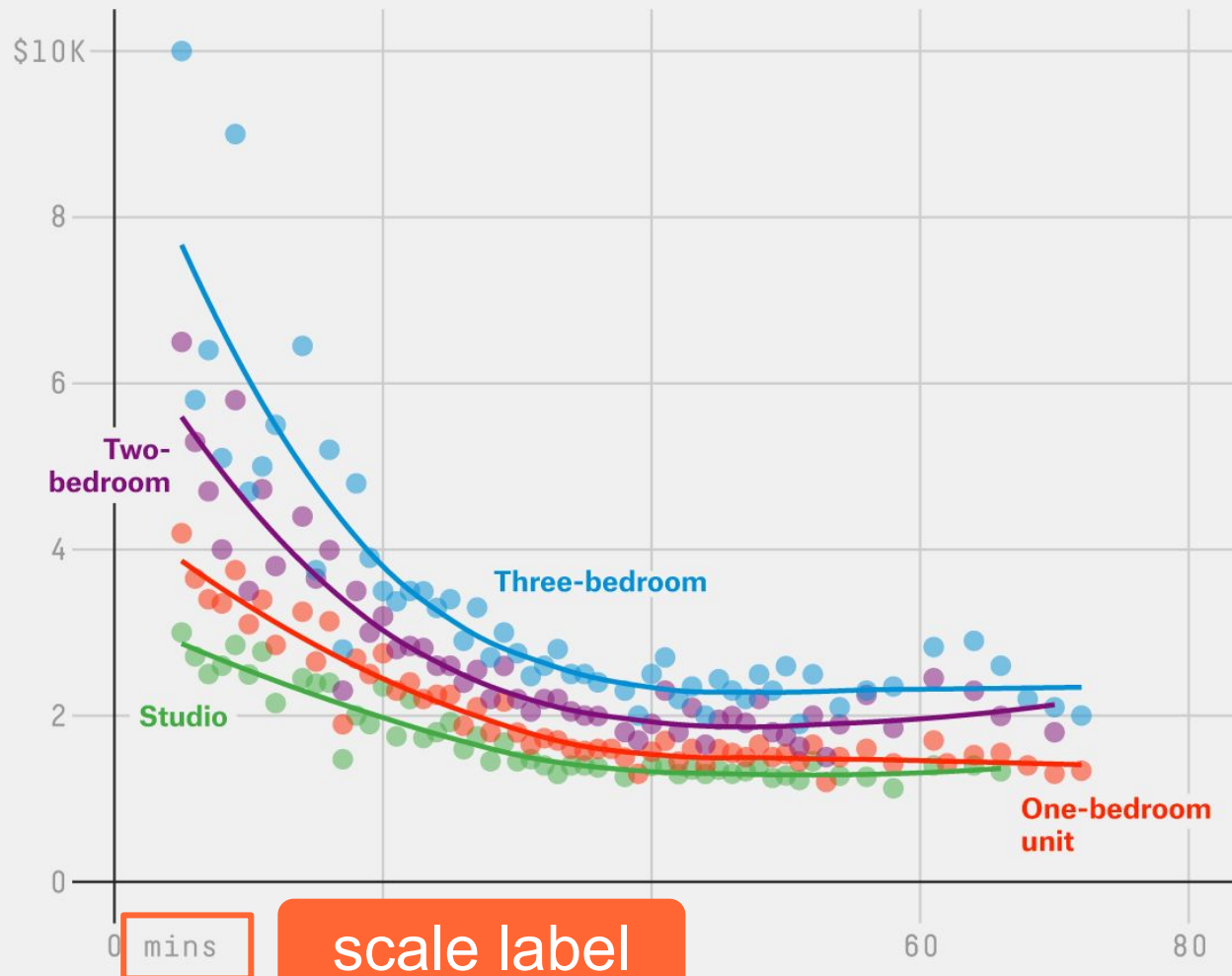
Median monthly NYC rent in 2015 vs. commute time by subway



Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

New Yorkers pay up for a shorter commute

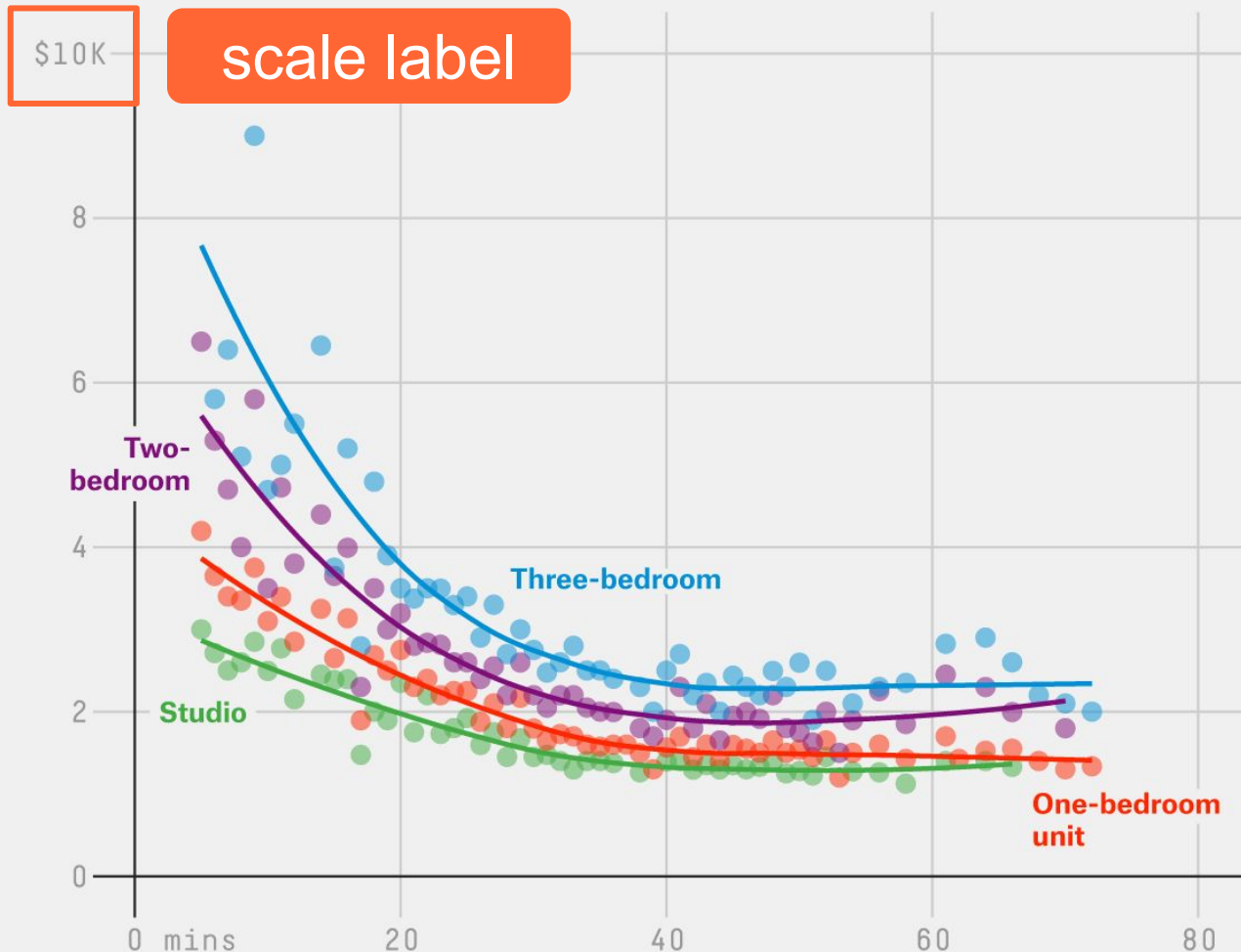
Median monthly NYC rent in 2015 vs. commute time by subway



Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

New Yorkers pay up for a shorter commute

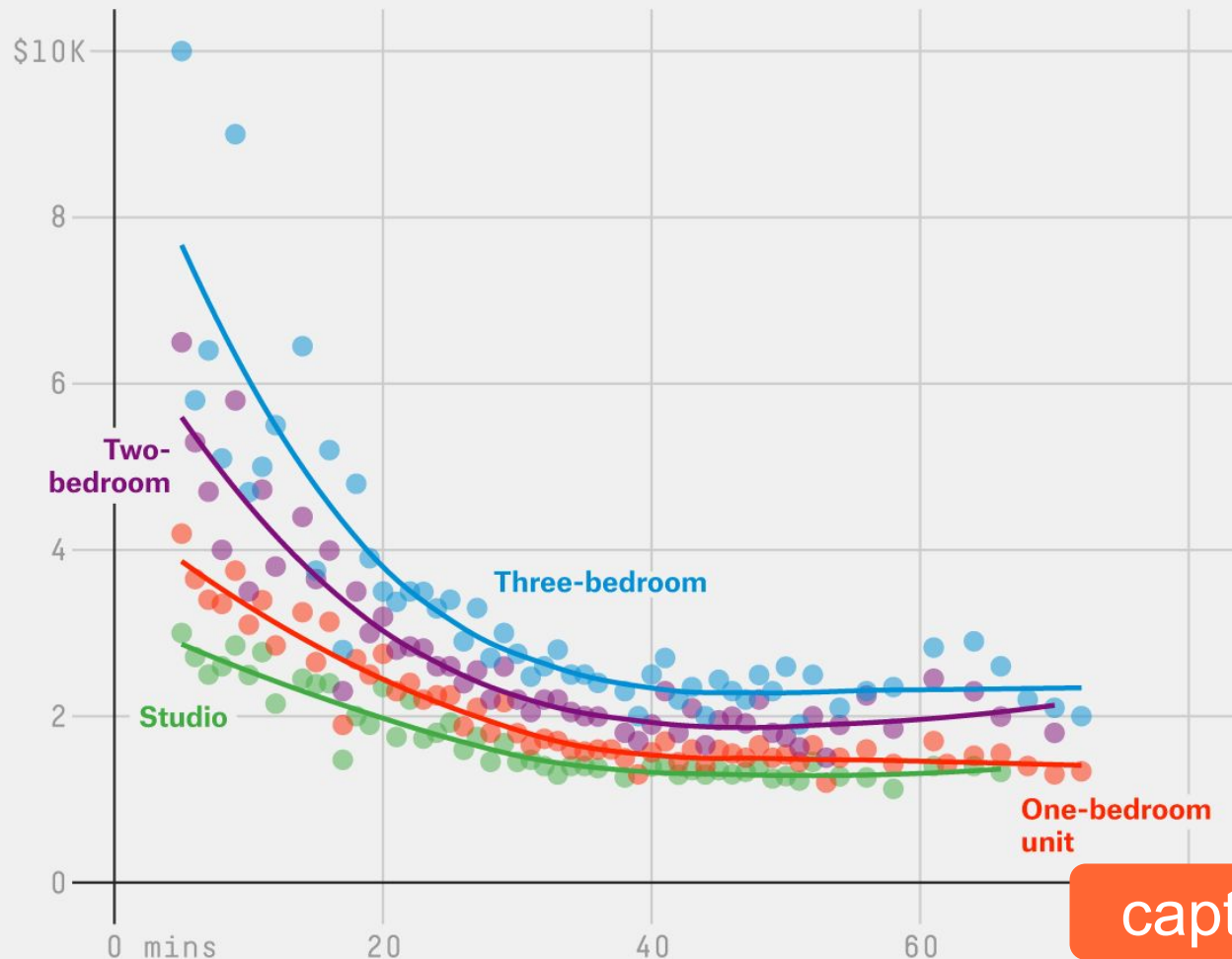
Median monthly NYC rent in 2015 vs. commute time by subway



Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

New Yorkers pay up for a shorter commute

Median monthly NYC rent in 2015 vs. commute time by subway




caption

Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

What graph should you use?

(We're going to focus on graphics for communication)



“With computer technology,
anyone can create graphics,
but few of us know how to do
it well.”

Donna Wong

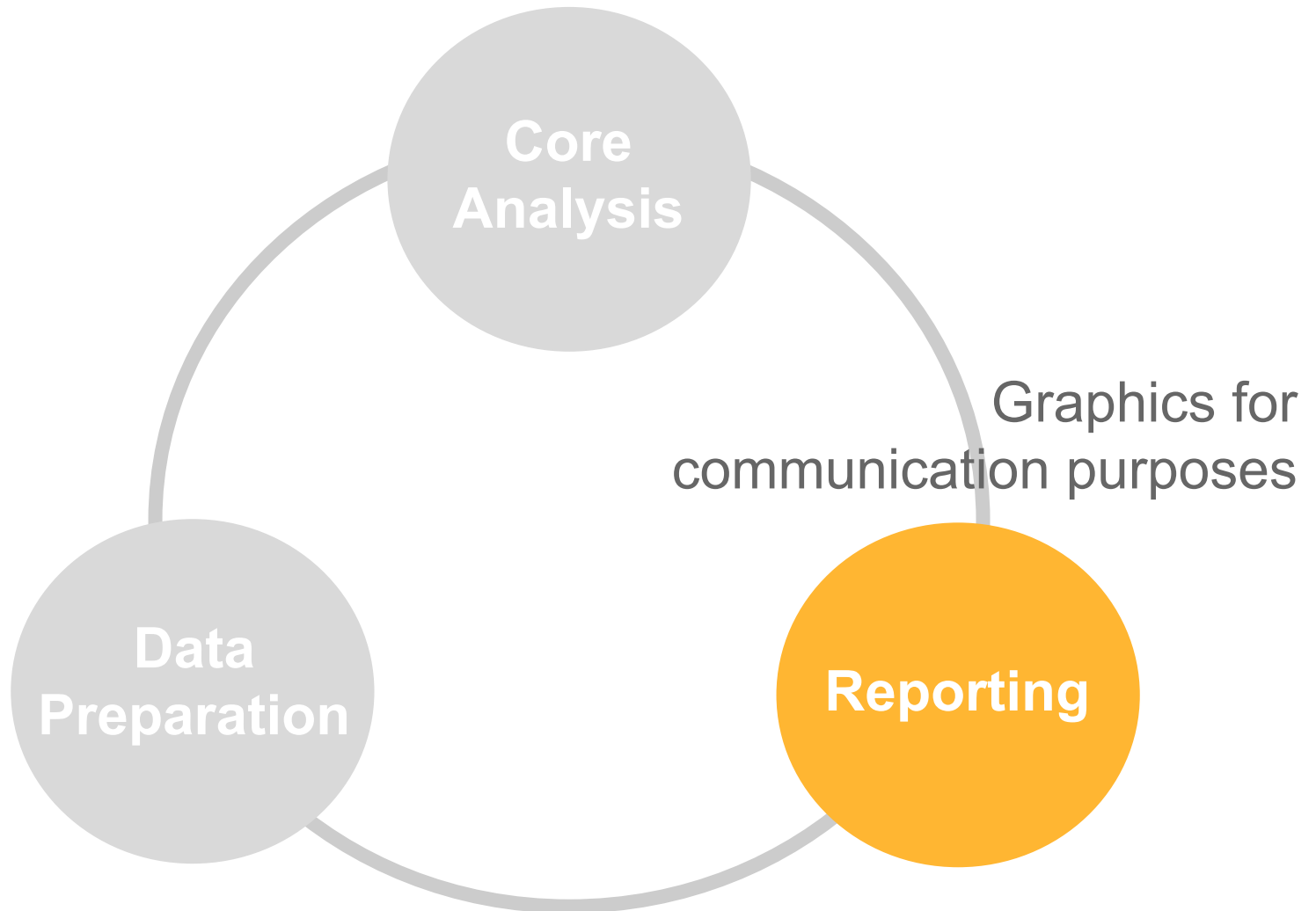
Charts for ...

exploration

vs

presentation

Data Analysis Cycle (DAC)




Recap: Graphics for Communication

Graphics for presenting data

To be consumed by others

Must care about visual appearance and design principles

Require a lot of iterations in order to get the final version



“The important criterion for a graph is not simply how fast we can see a result; rather it is whether through the use of the graph we can see something that would have been harder to see otherwise or that could not have been seen at all.”

William Cleveland

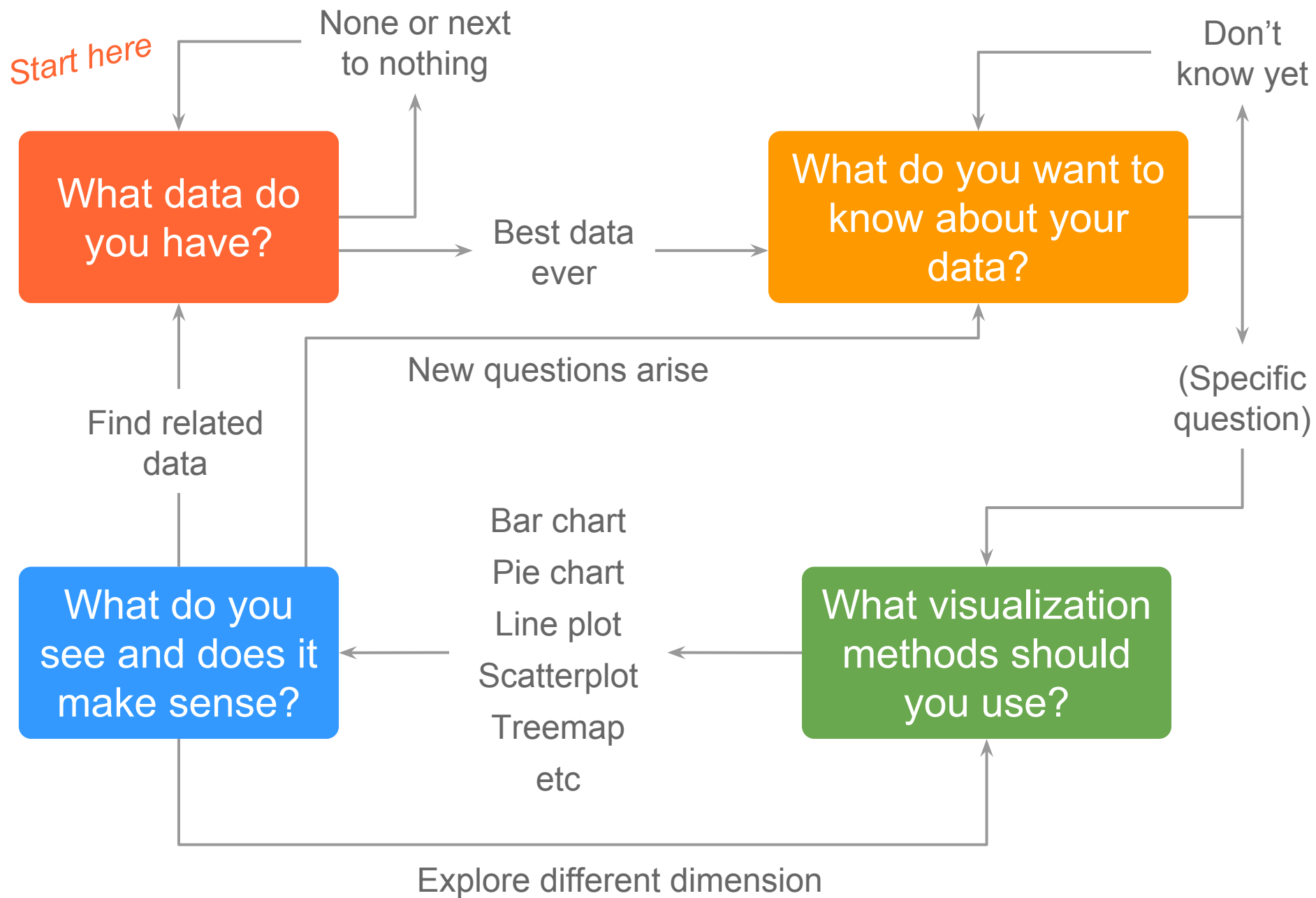
(The Elements of Graphig Data)

So you want to make a plot?

The plotting steps vary by dataset and project

But you should consider four things:

1. What data do you have?
2. What do you want to know about the data?
3. What visualization methods should you use?
4. What do you see and does it make sense?



Always ask

What is the data set of interest?

What variables (columns) will be used to make the plot?

What graphic shapes (geoms) will be used to display the data?

What features of the shapes will be used to represent the data values?

What data do you have?

How many variables?

1. One variable
2. Two variables
3. Three or more

What type of variables?

4. Quantitative, qualitative, time

What do you want to know about your data?

- Part-to-whole analysis
- Ranking analysis
- Deviation analysis
- Times series (trends in time)
- Distribution analysis
- Correlation analysis
- Multivariate analysis

What do you do see?

- Systematic variation
- Increasing patterns
- Decreasing patterns
- Atypical or outliers
- Noise?

Recommendations

What task and/or question is on the table?

Think about what combination of objects you can use to encode your data.

Start working in analogical mode: get away from the computer, and work with pen-and-pencil.

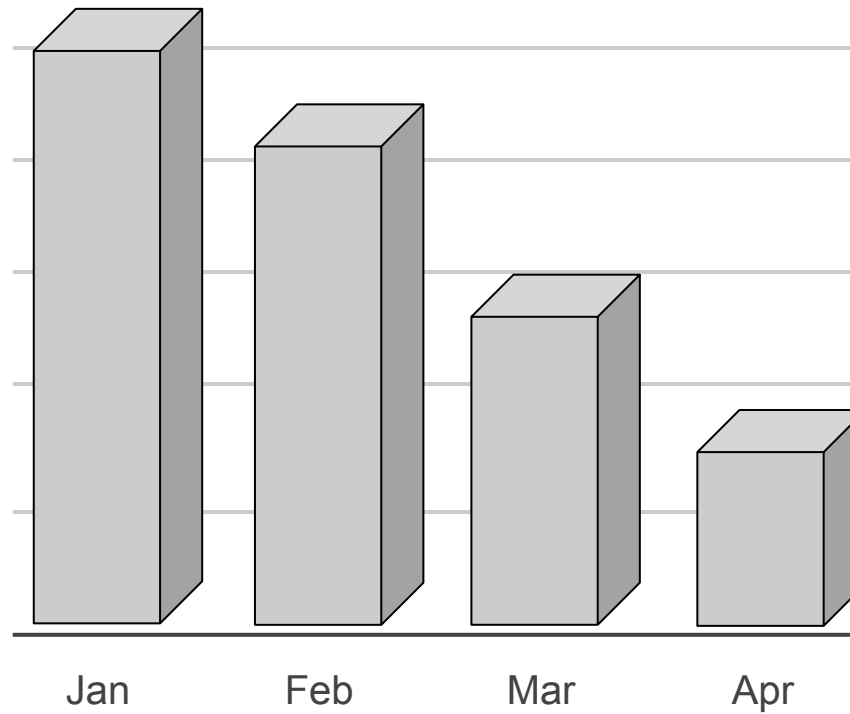
Once you have a good draft, then switch to the computer.

Polish your graph until you get a nice display.

Some Guidelines

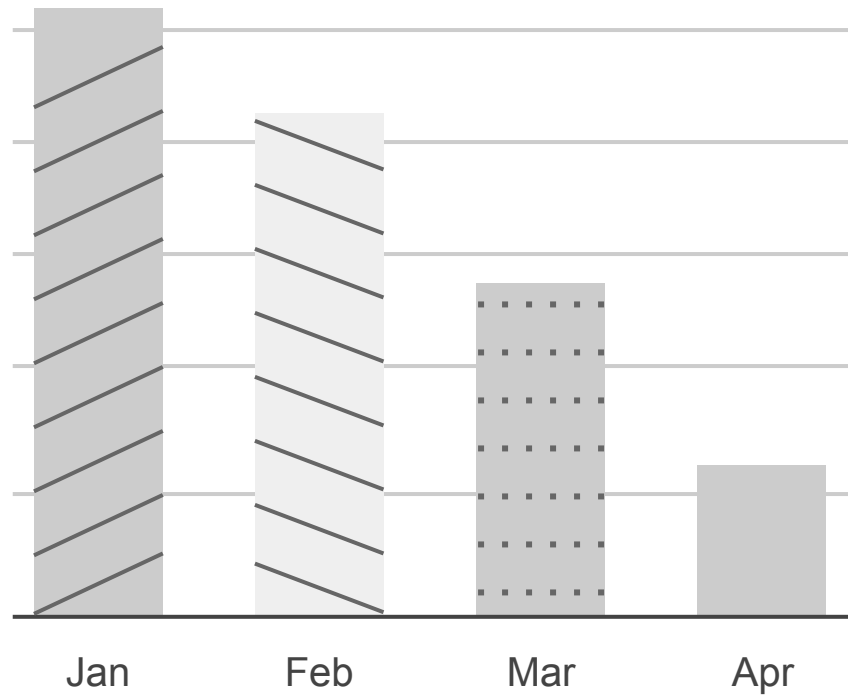


Don't use three-dimensional bars:
it adds no information
it is hard to guess where the top is



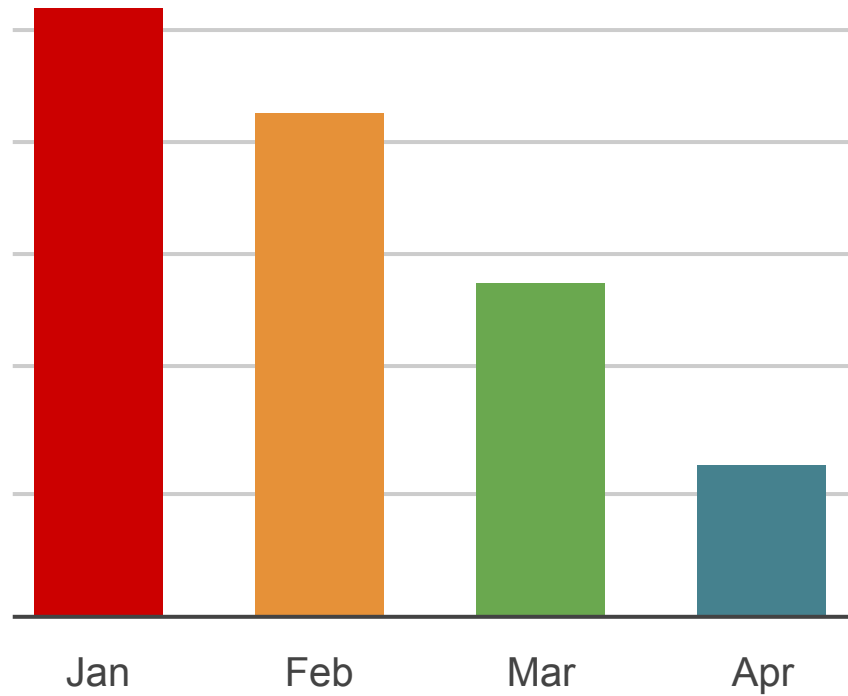


Don't use multiple distracting shades

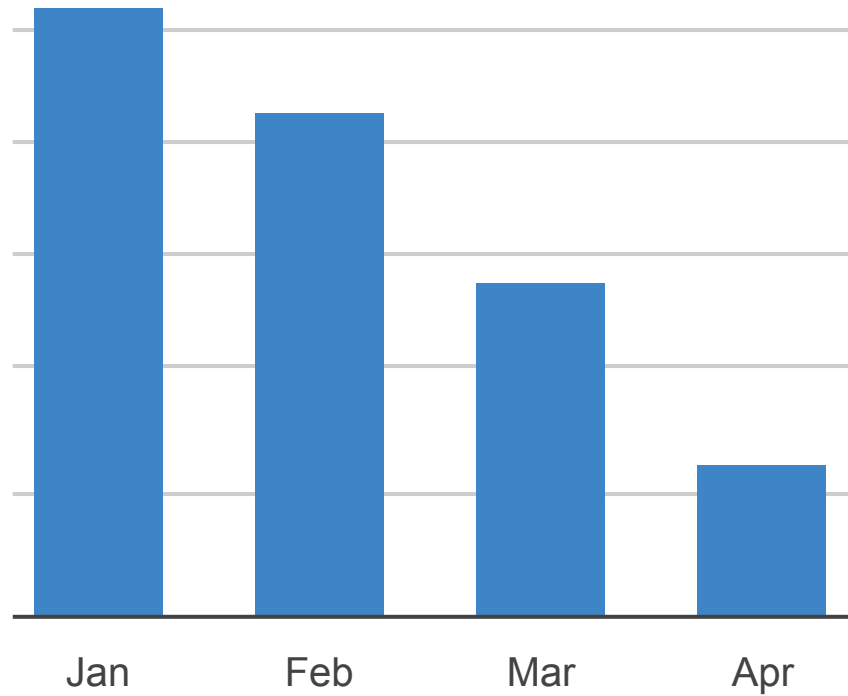




Don't use multiple colors to represent the same kind of data

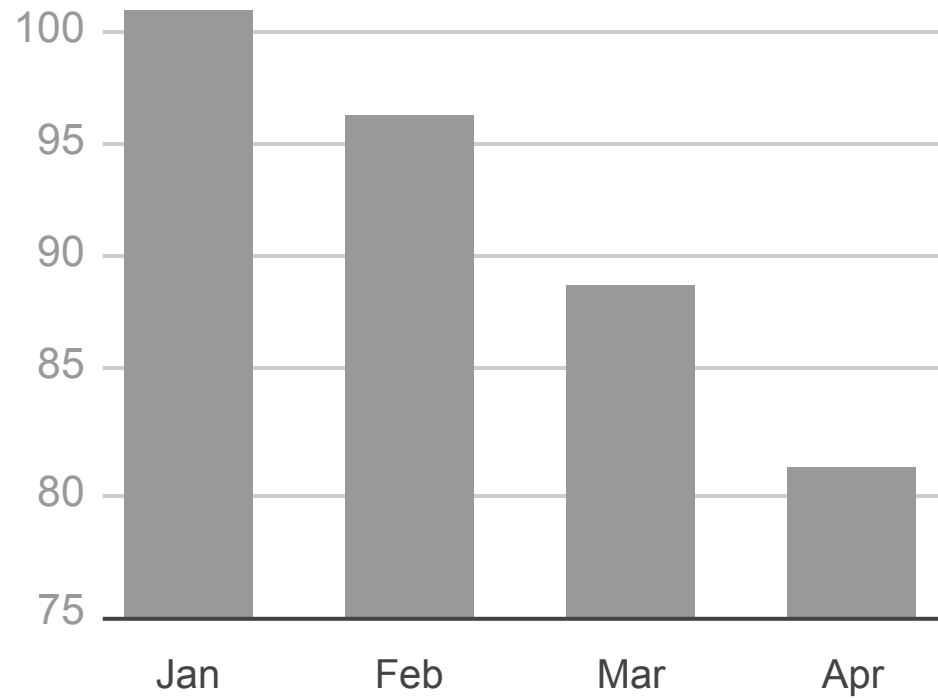


Use the same color to represent the same variable

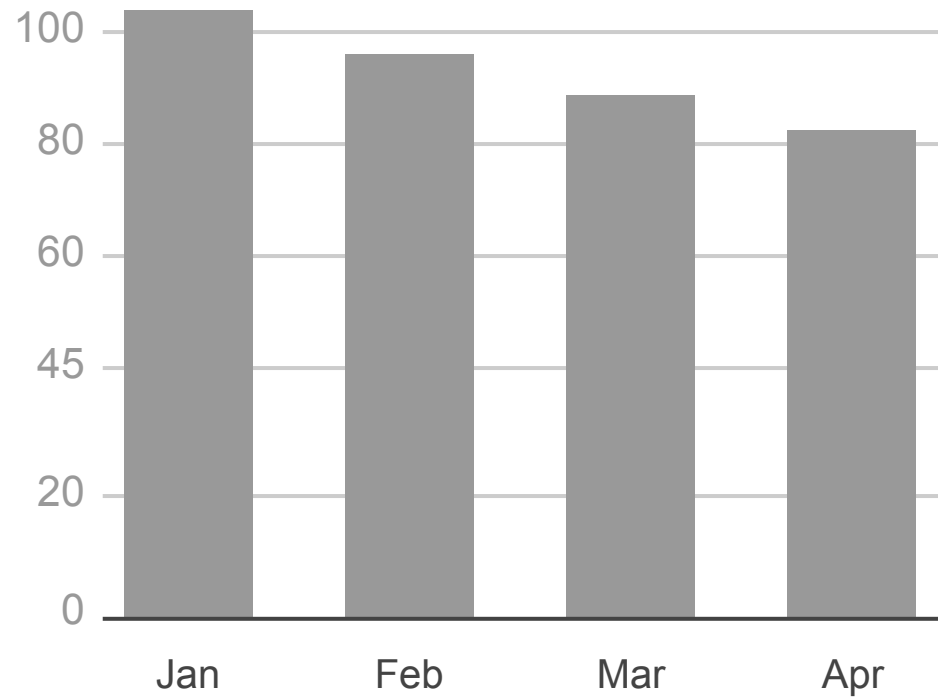




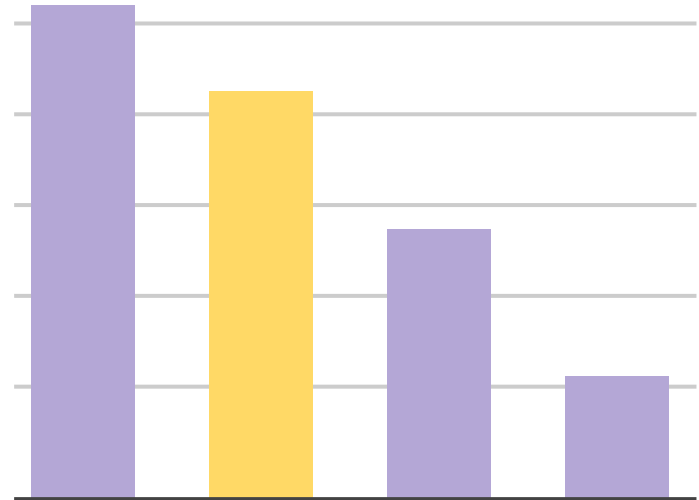
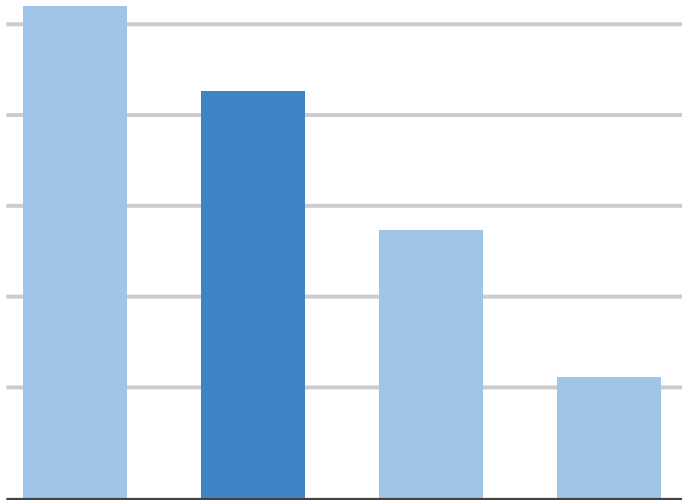
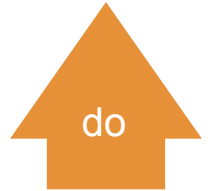
Don't truncate baseline of bar charts



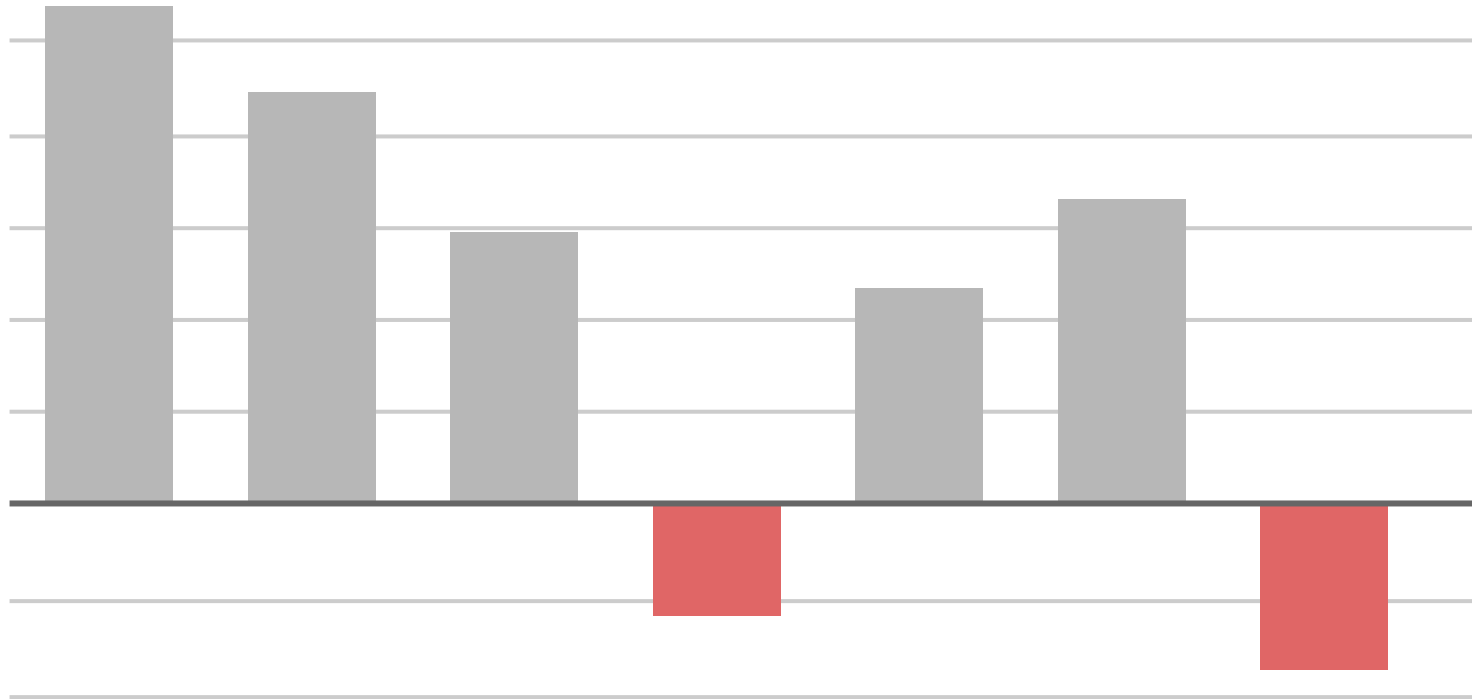
Always start at the zero baseline
No exceptions!



Use a darker shade or a different color
to highlight the focal point



You can use contrasting colors to depict negative values

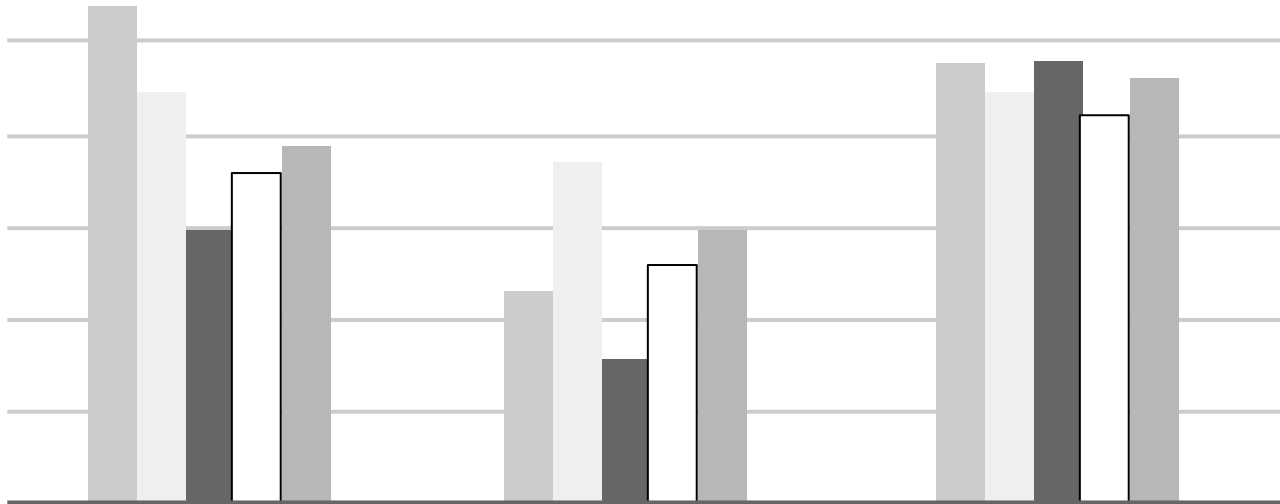


You can use a gray background to identify the negative zone

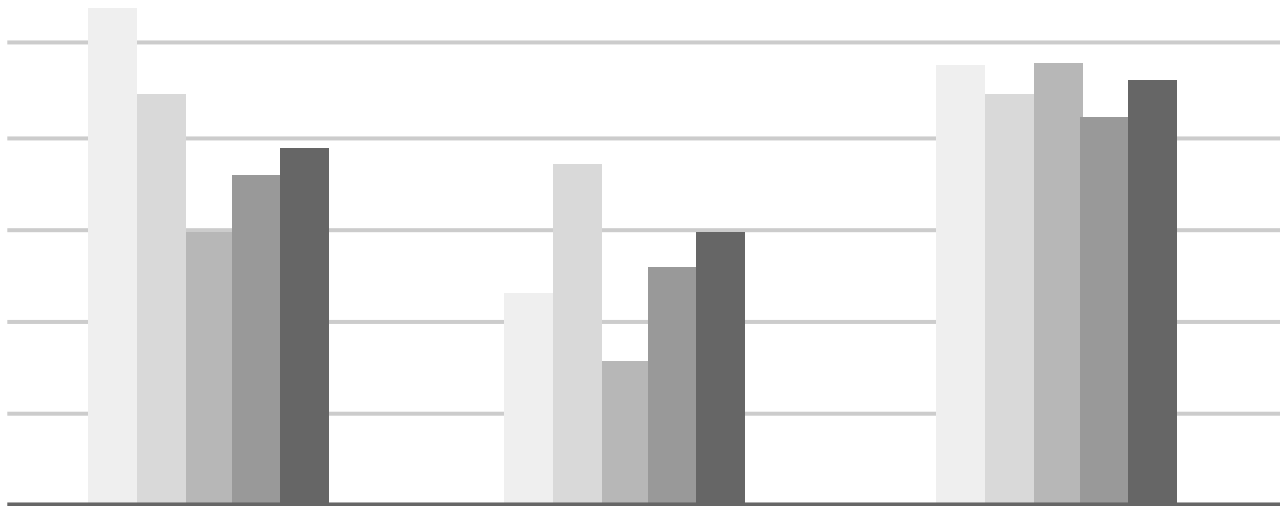
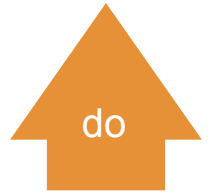




No zebra pattern

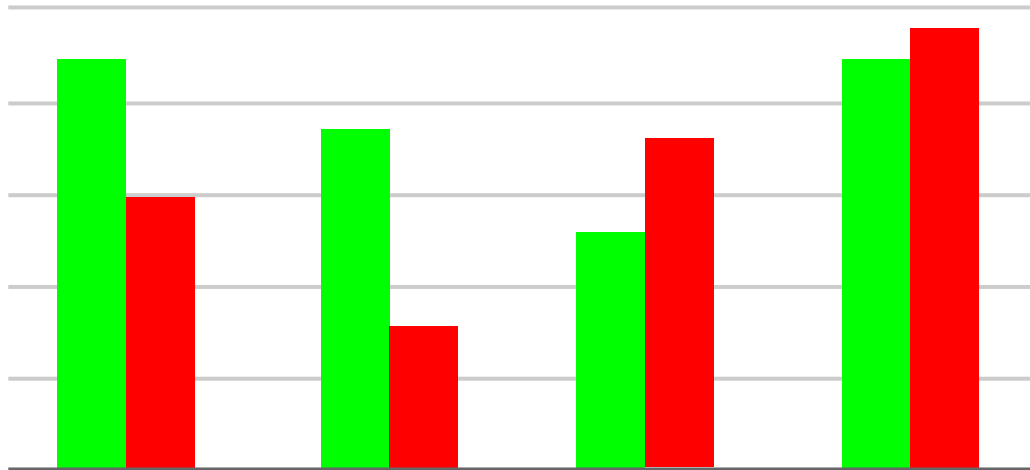


From lightest to darkest

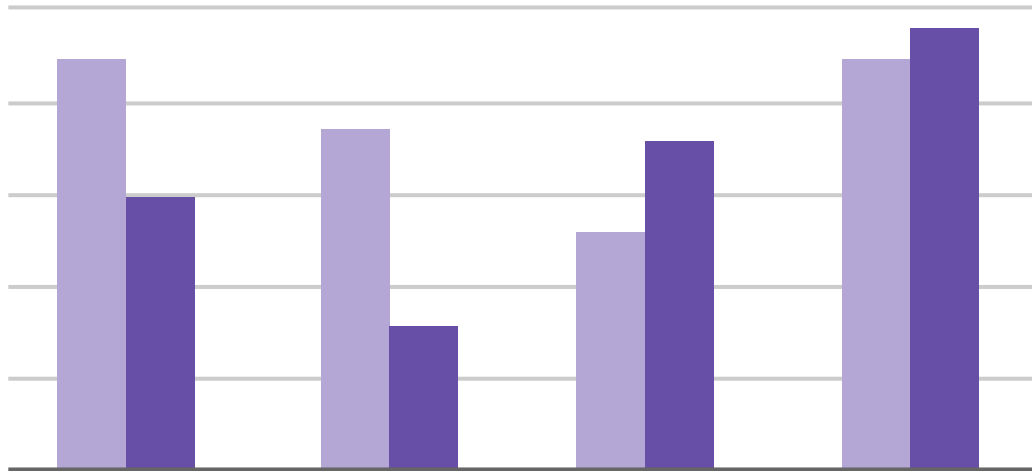




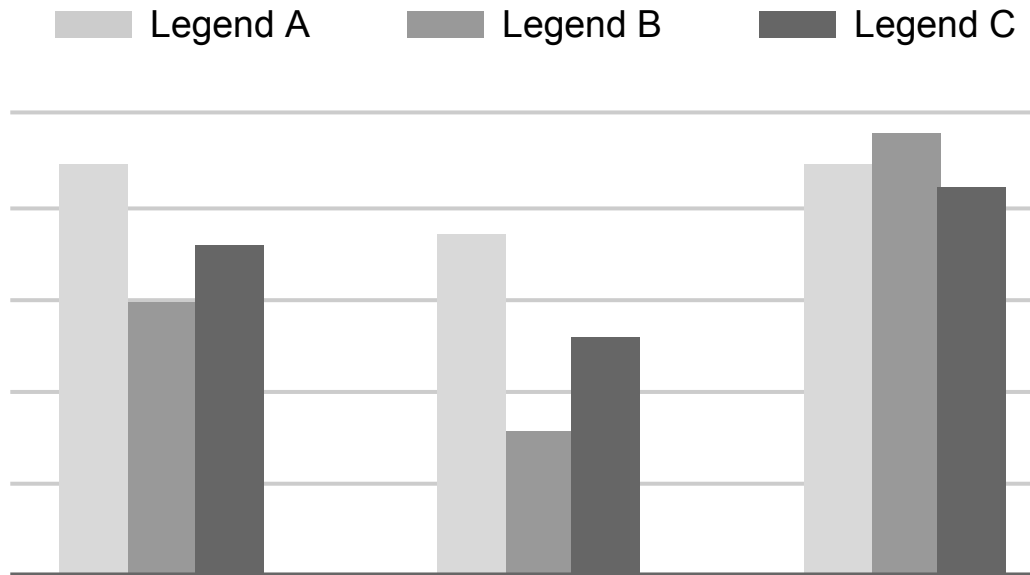
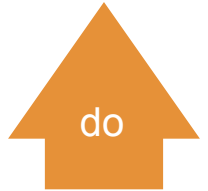
Be careful with the selection of colors:
When working with two colors avoid
complementary (opposite) colors



It's better to use different shades of the same color

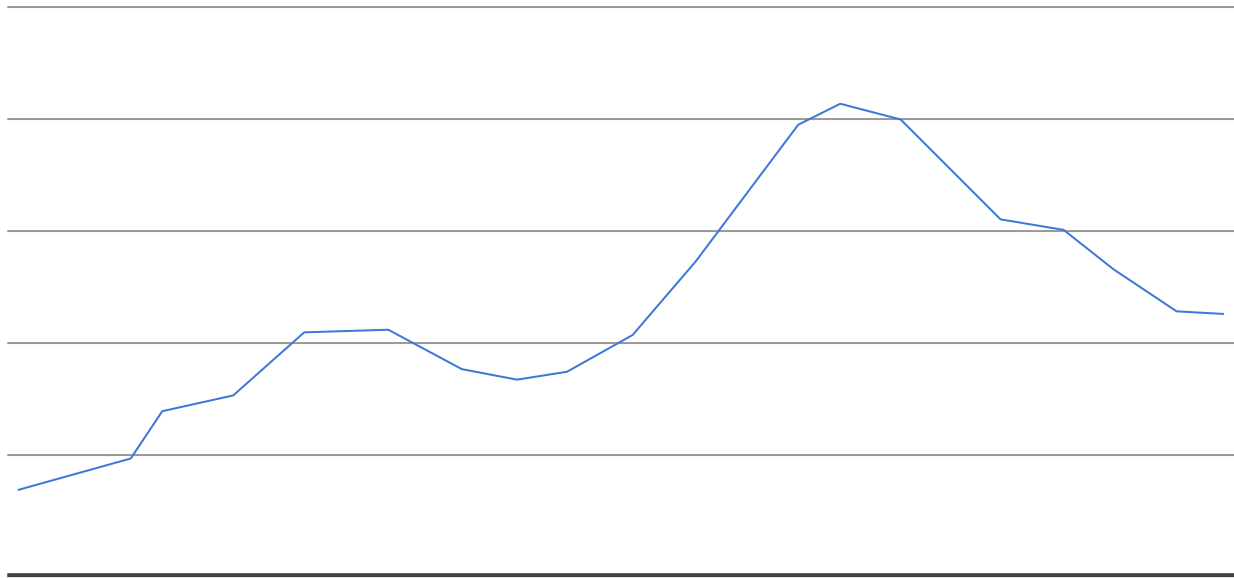


Add legends in the right sequence



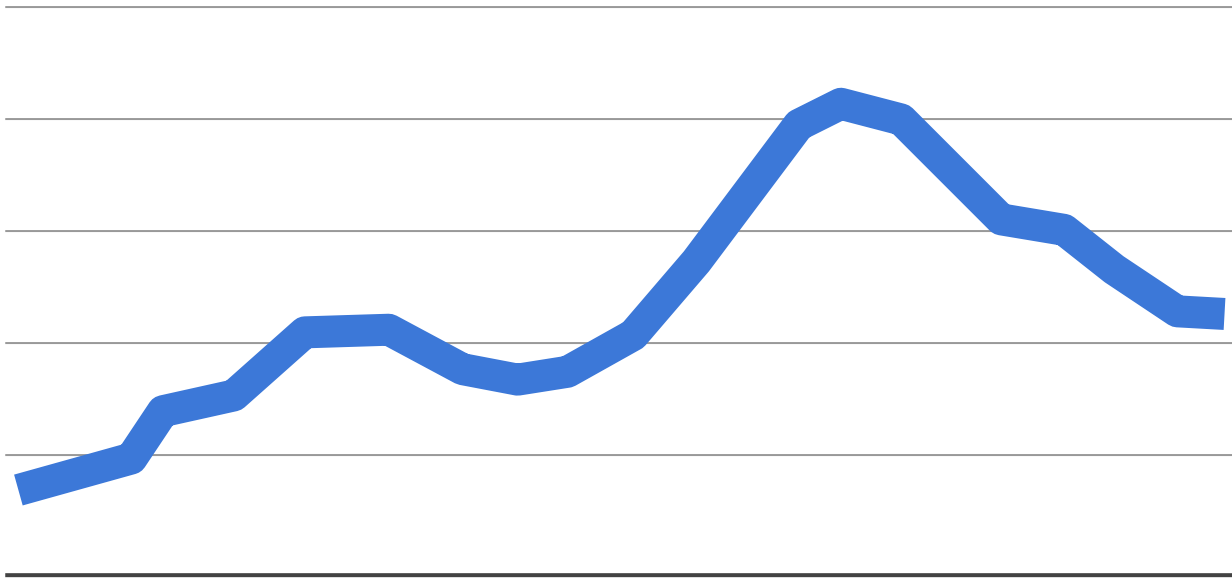


lines too thin are hard to read

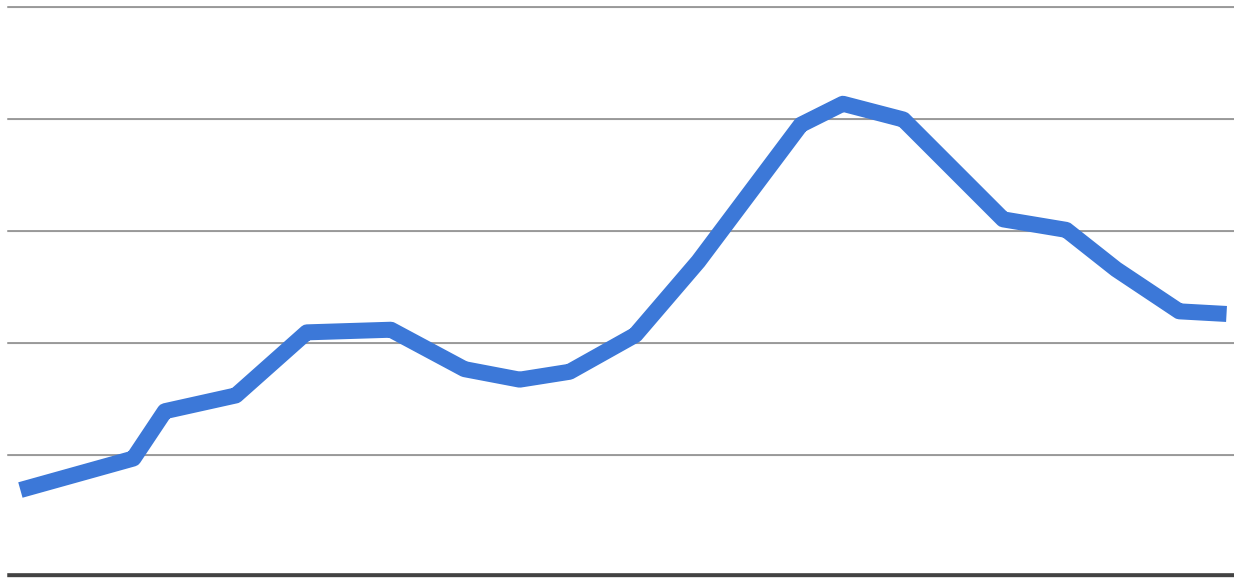
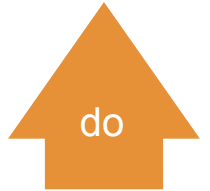




too thick lines may hide some details

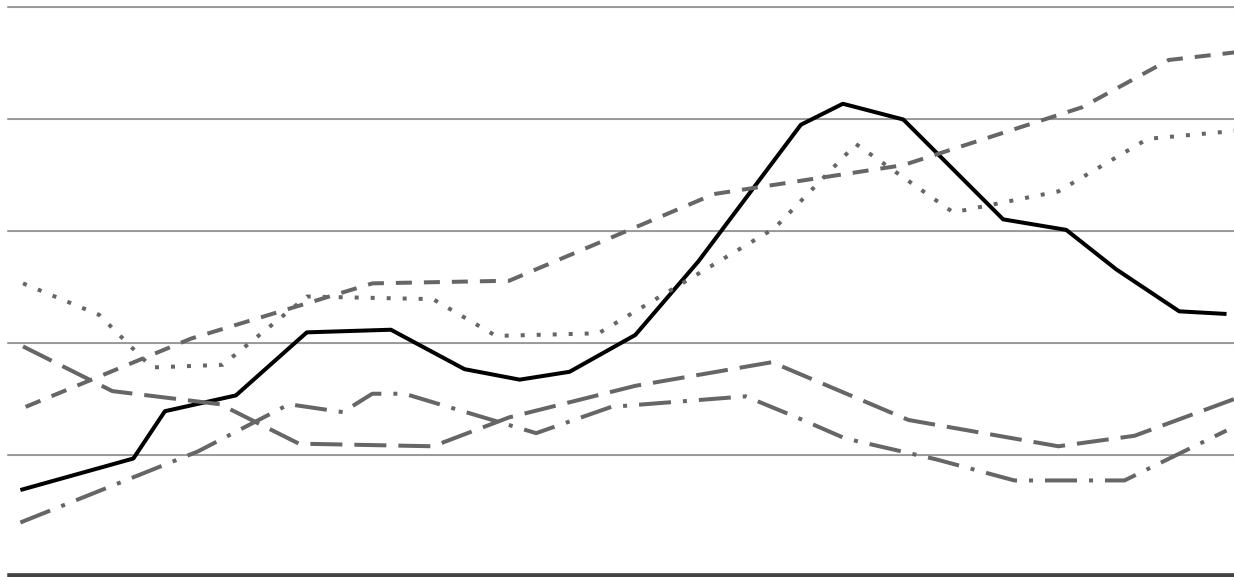


look for “an OK” thickness

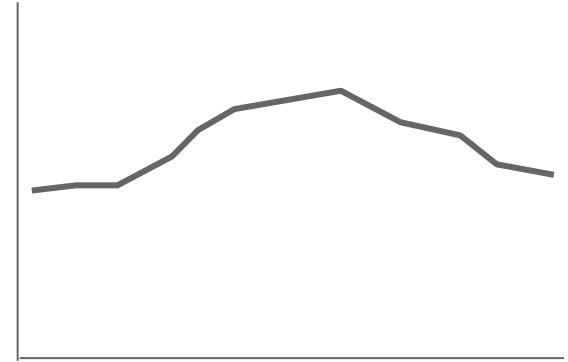
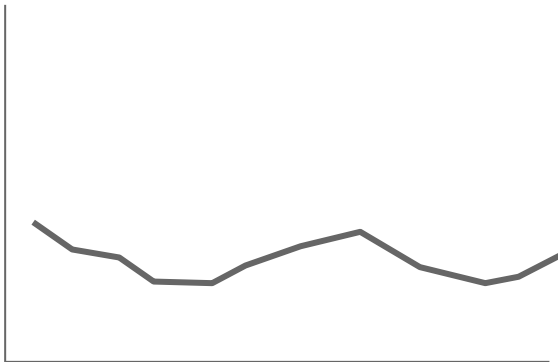
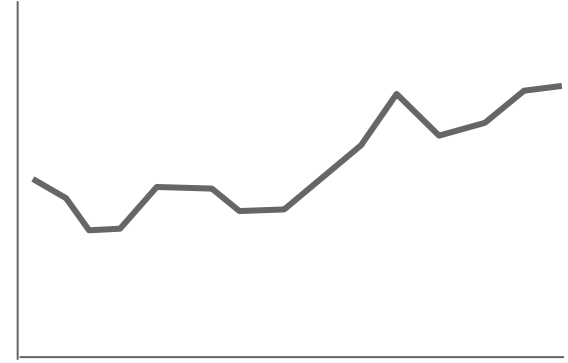
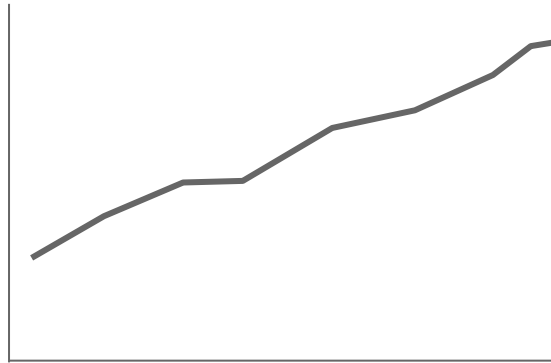
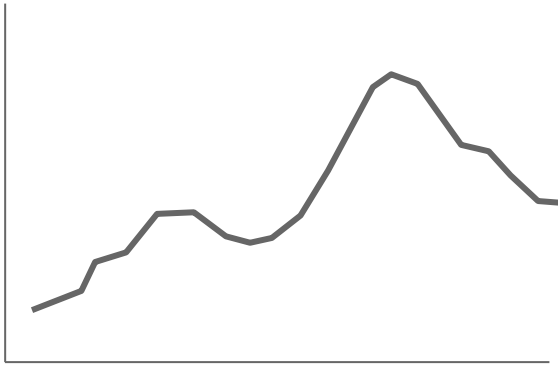
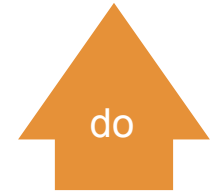




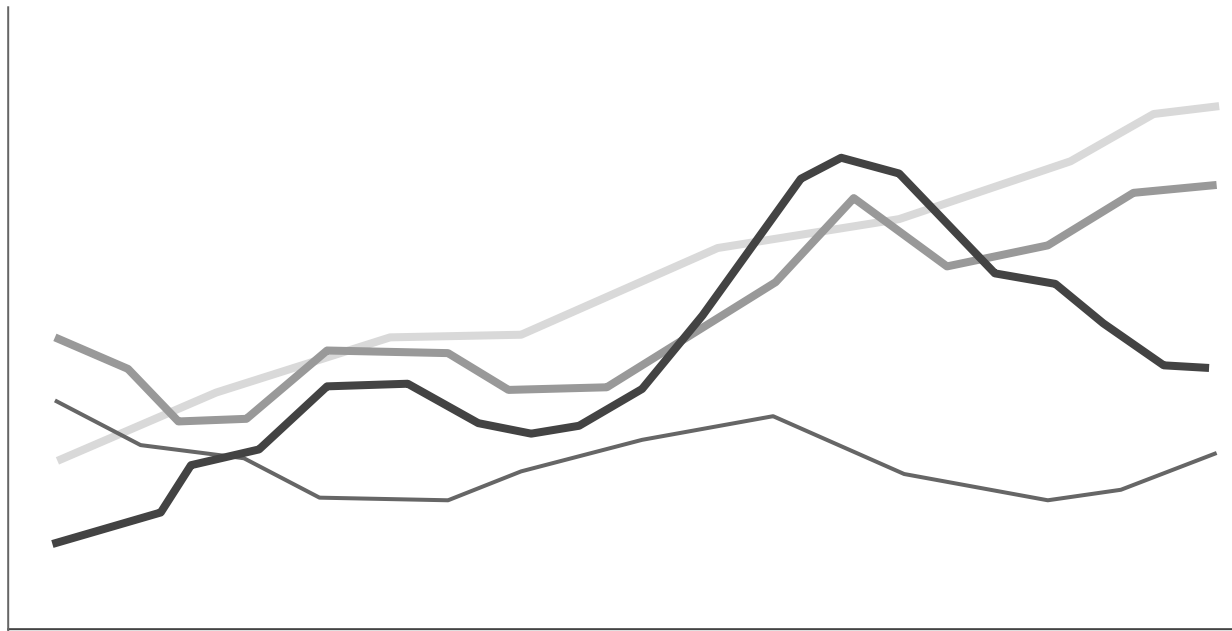
Avoid spaghetti lines:
Don't differentiate each line trying
all types of dashed lines



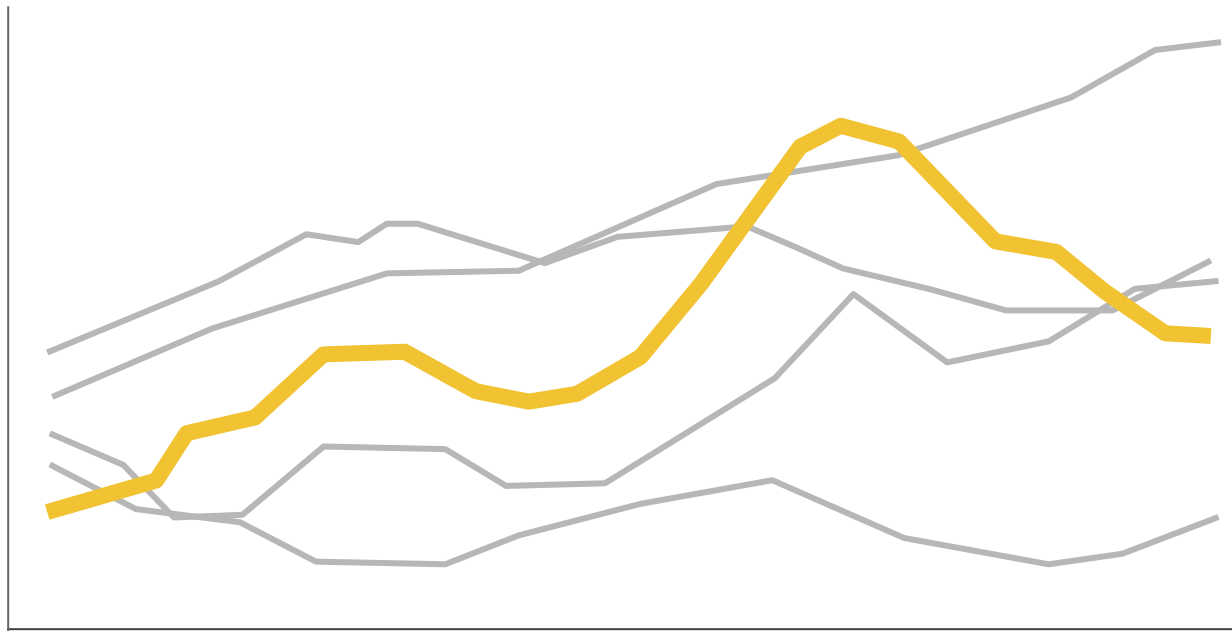
Better comparisons with panel of charts



Four or fewer lines is better
with different dark intensities

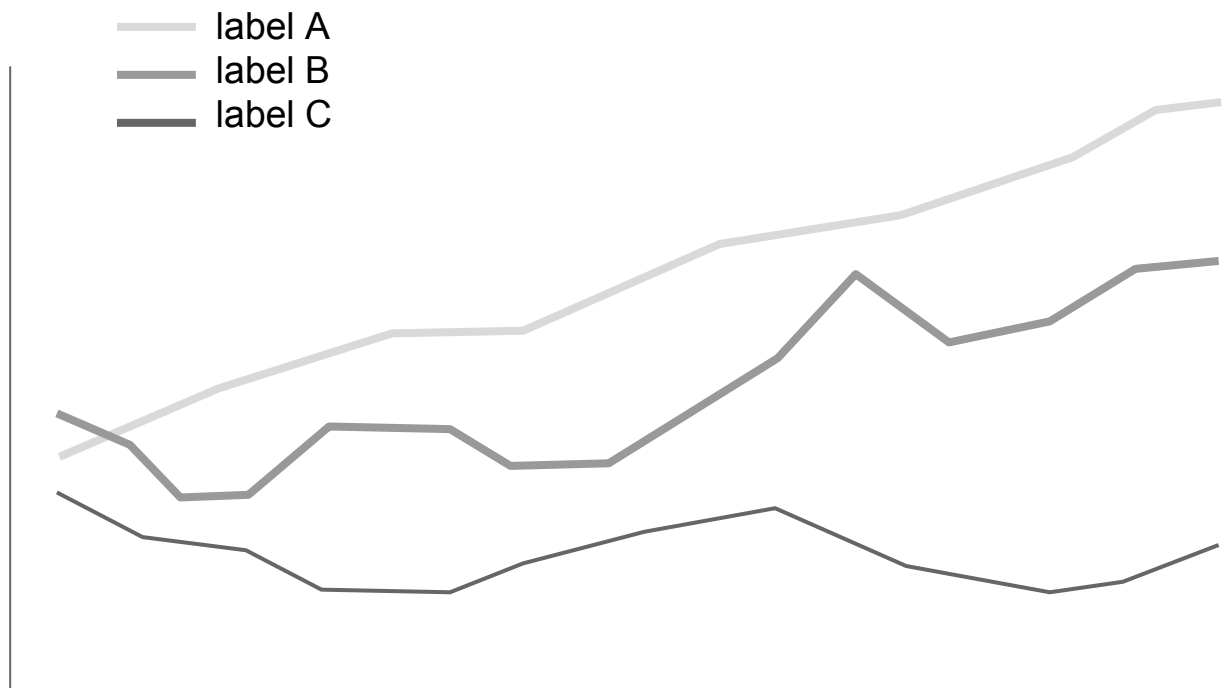


Use bright to dark colors to
emphasize the important line

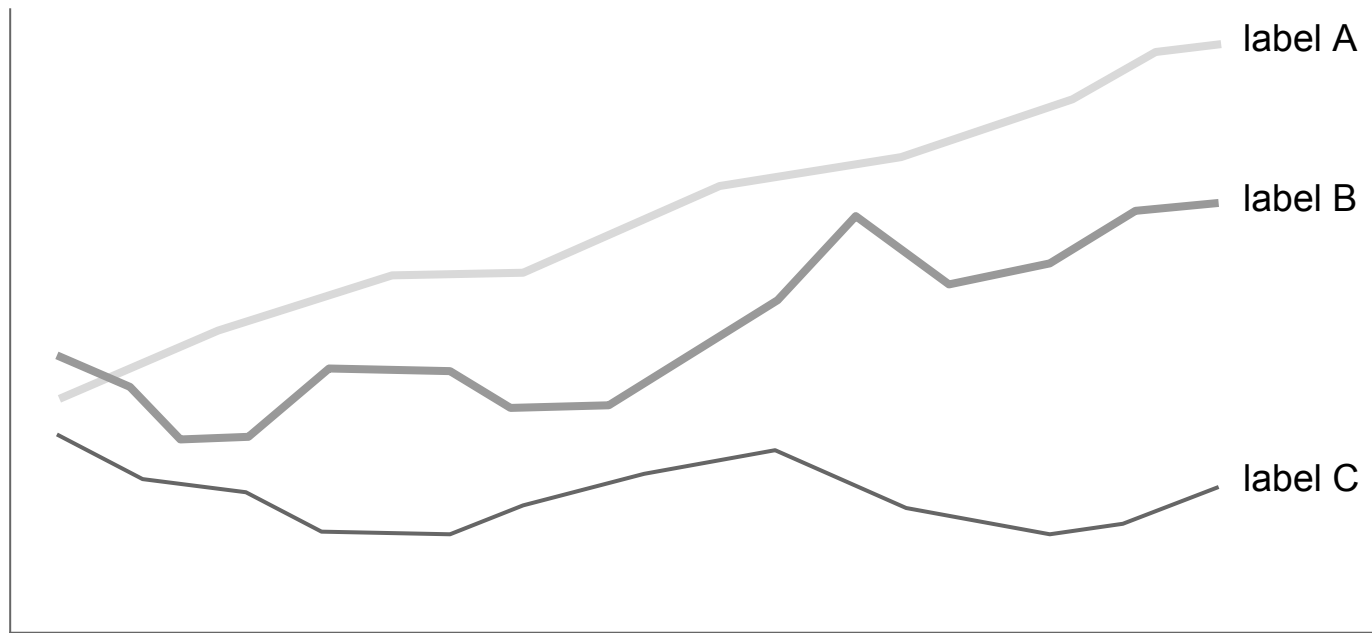




Avoid labeling at long distance

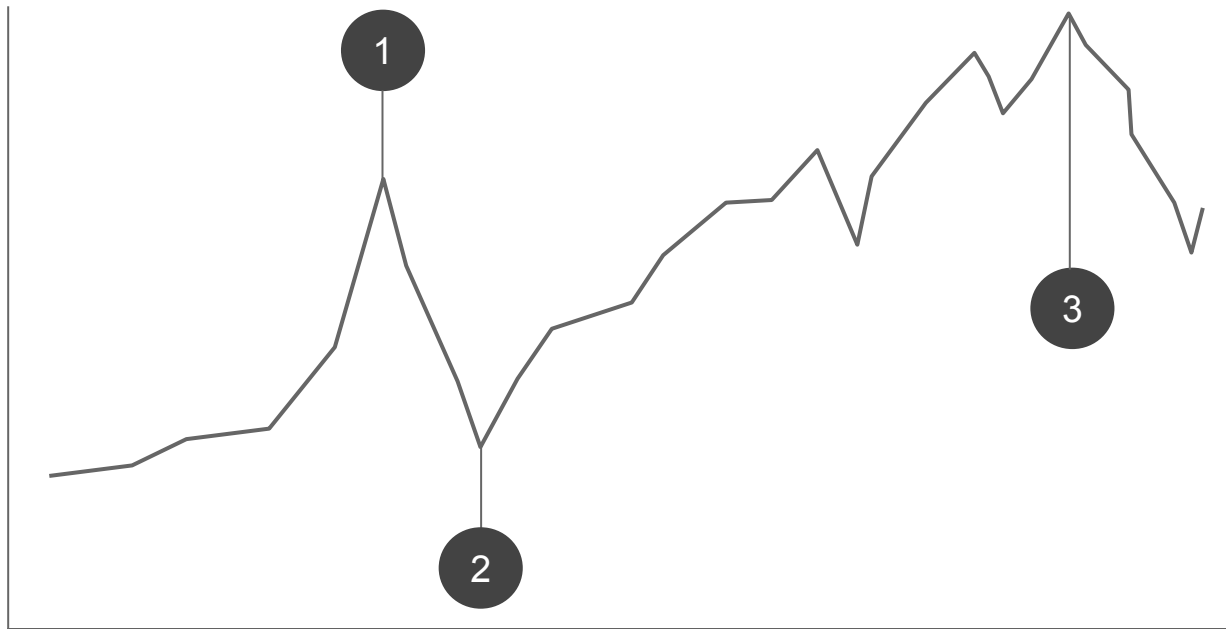


Label the lines directly



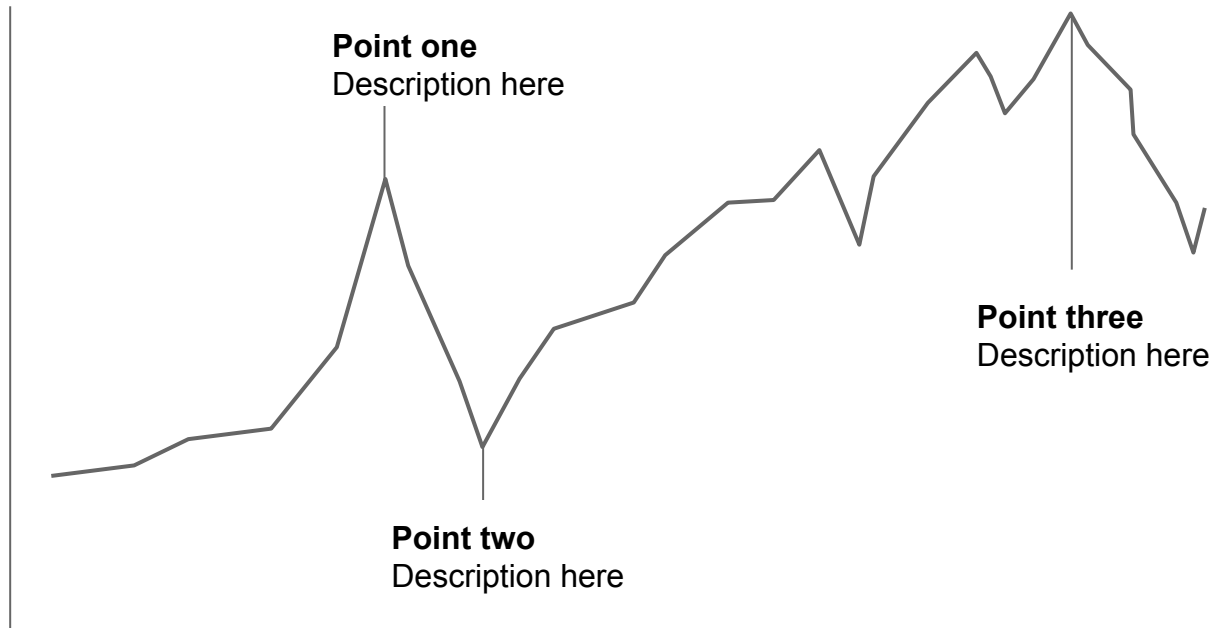
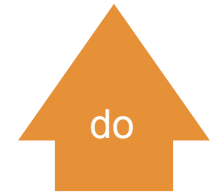


Avoid labeling at long distance



- 1 **Point one** Description text goes here
- 2 **Point two** Description text goes here
- 3 **Point three** Description text goes here

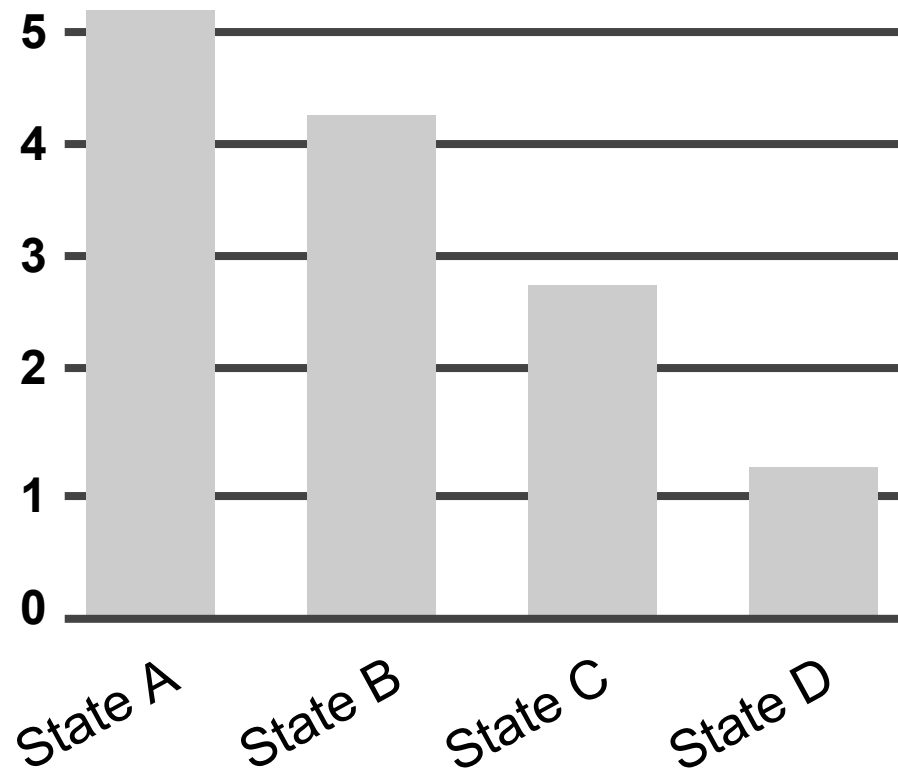
Prefer direct labeling

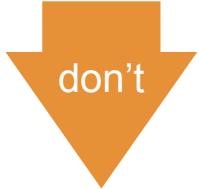




HEADLINE OF THE CHART

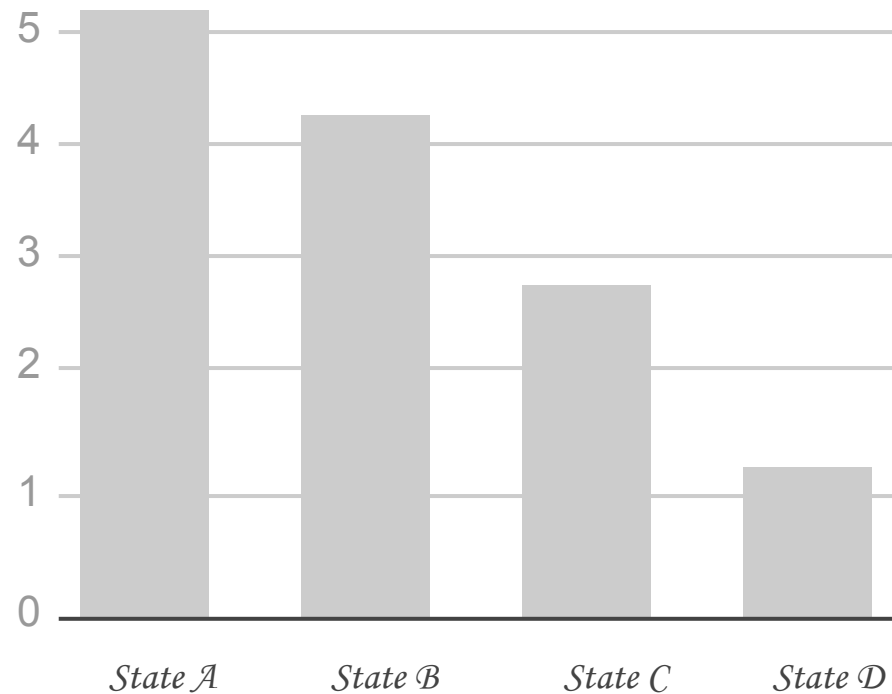
A brief description that outlines what the data shows





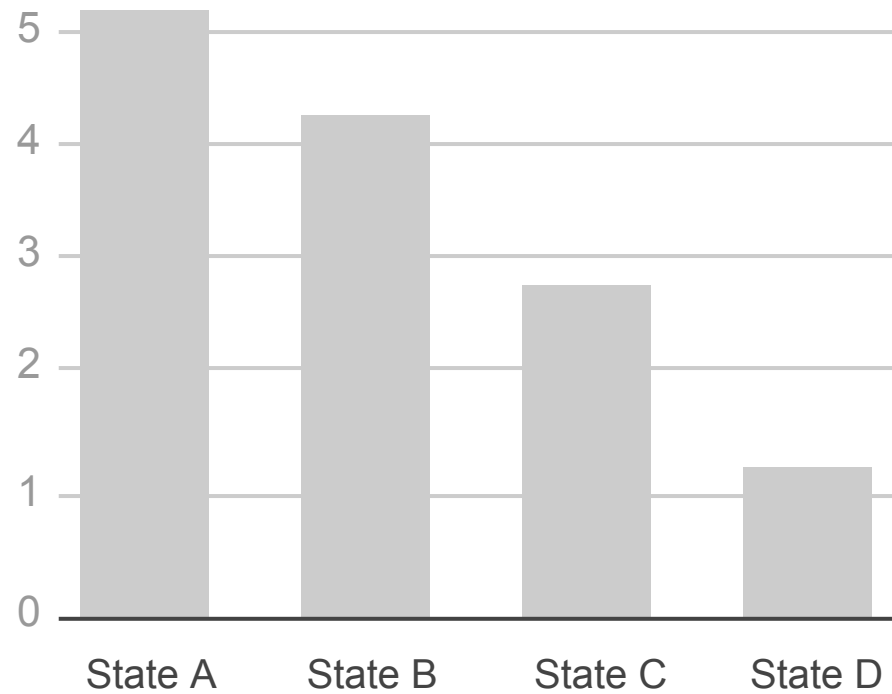
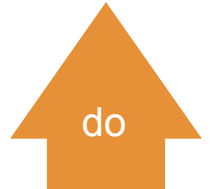
HEADLINE OF THE CHART

A brief description that outlines what the data shows



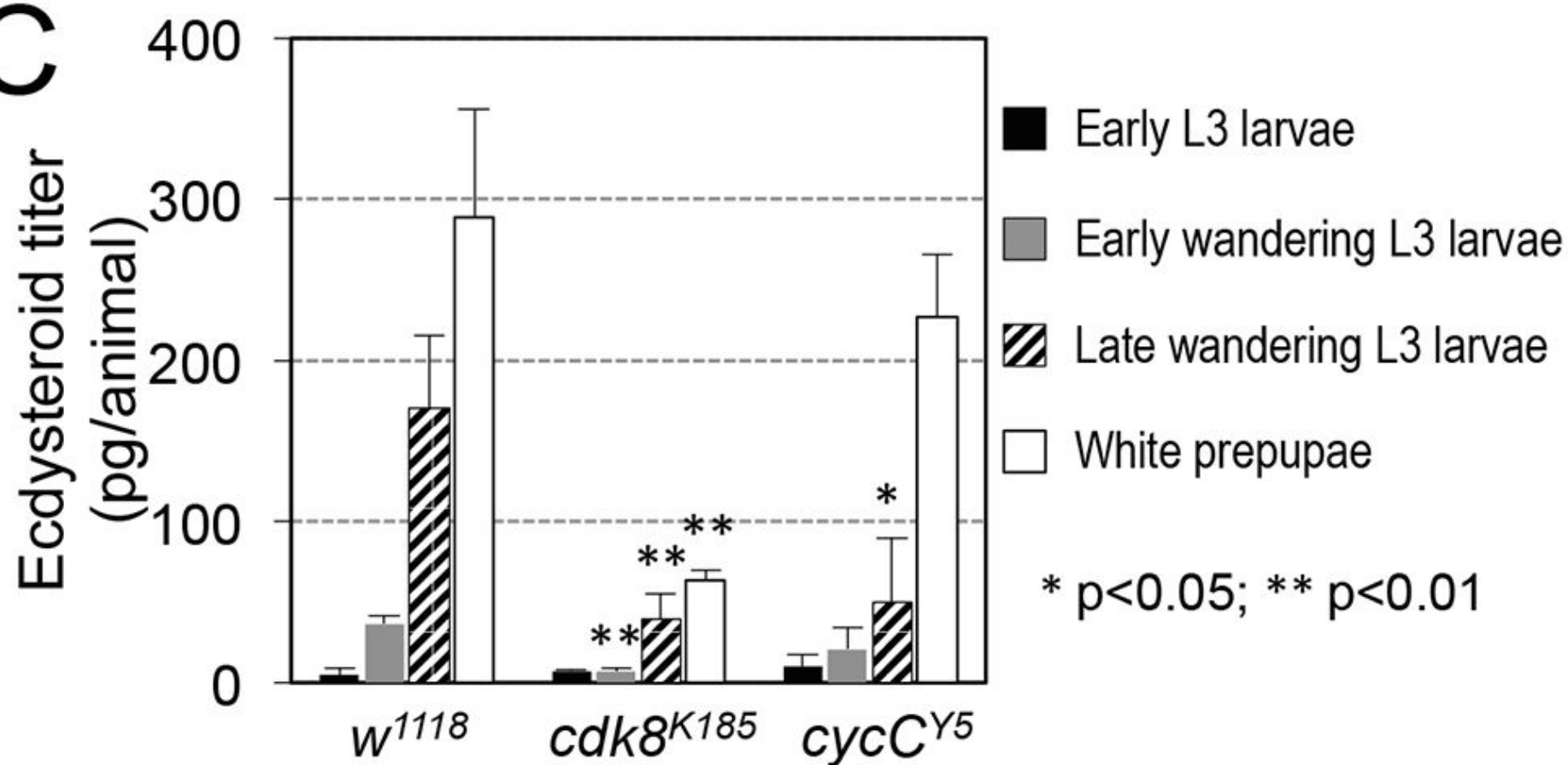
HEADLINE OF THE CHART

A brief description that outlines what the data shows



Some Examples

Scientific Journals

C

G

Time from egg laying
to pupariation (hrs)

250

200

150

100

50

0

no 20E

+ 20E

*w*¹¹¹⁸*cdk8*^{K185}*cycC*^{Y5}

N = 603

N = 813

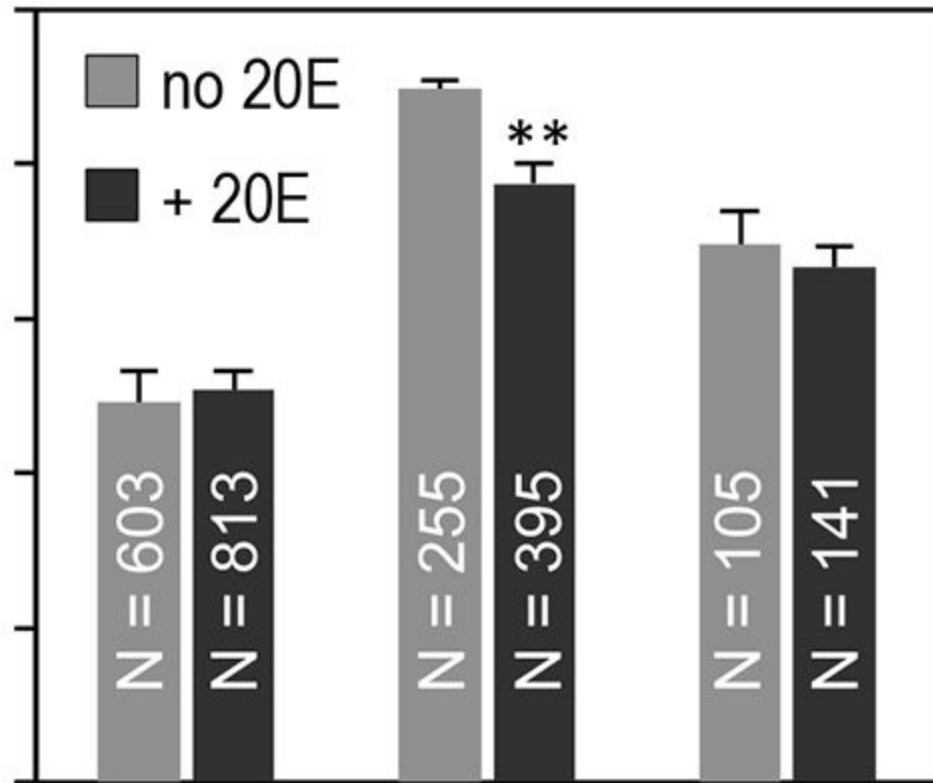
N = 255

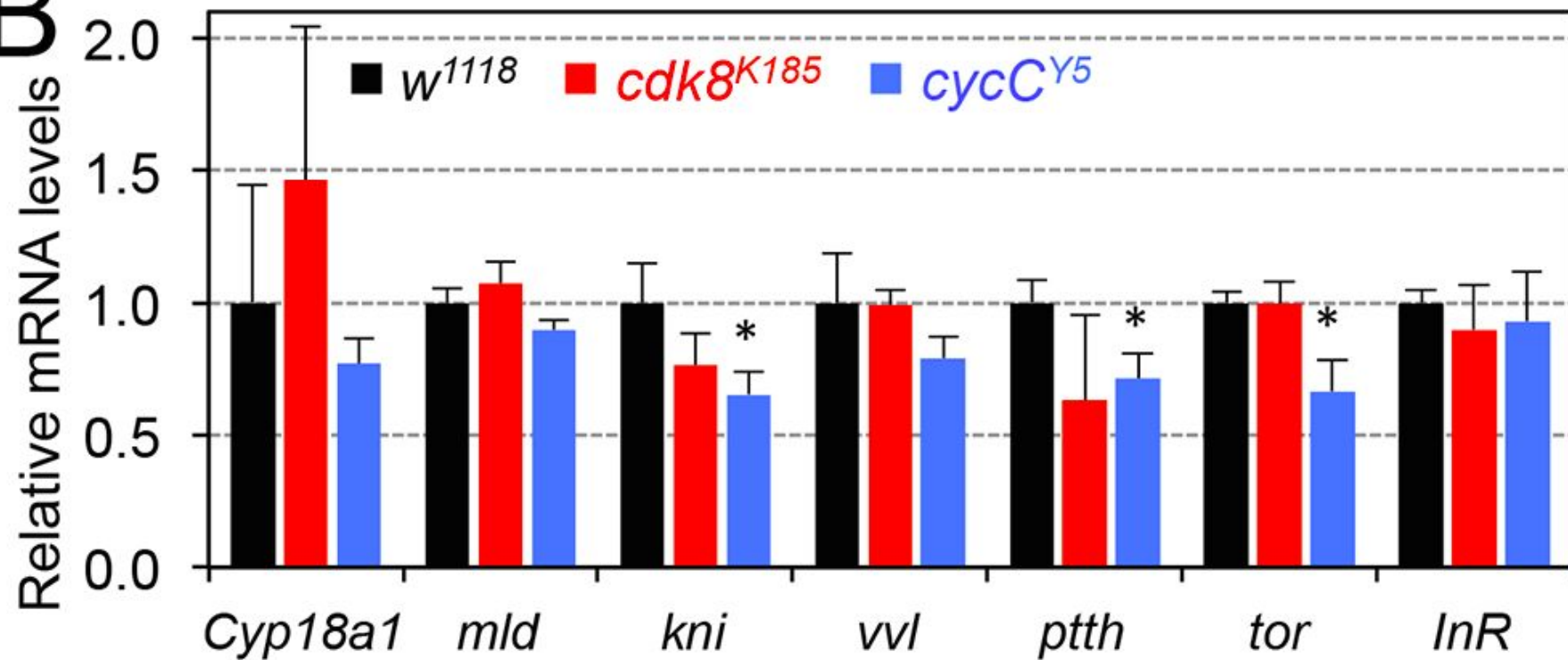
N = 395

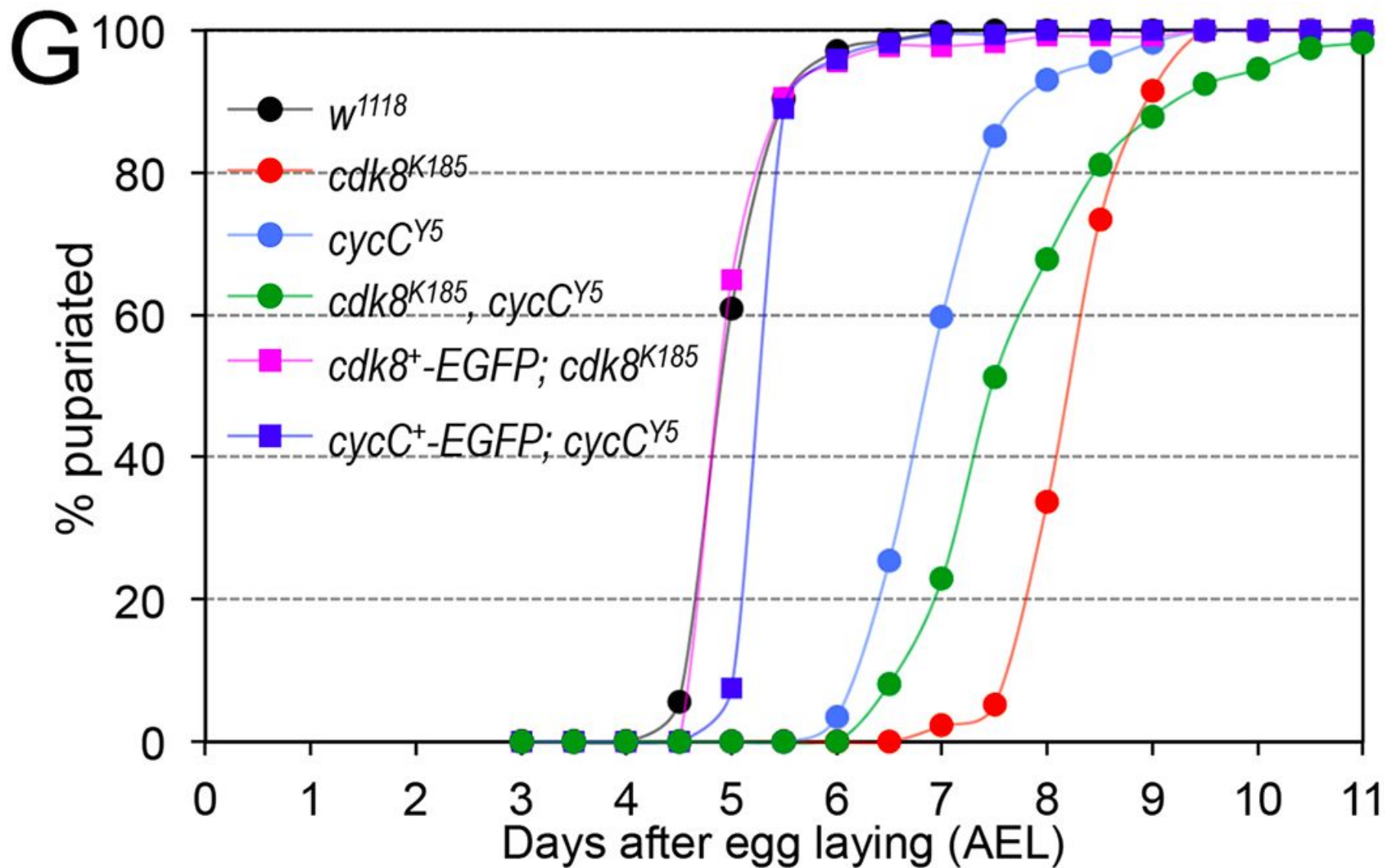
N = 105

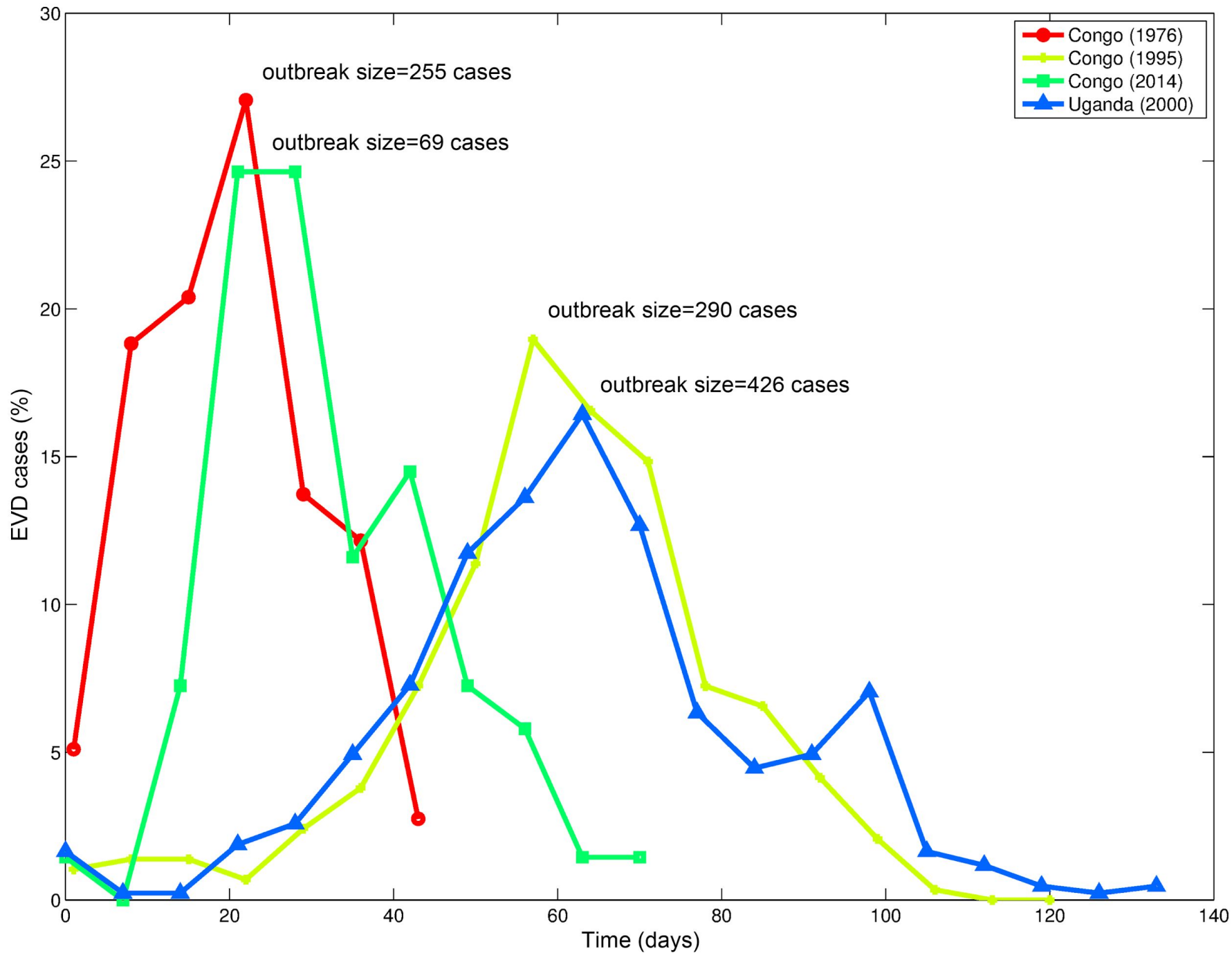
N = 141

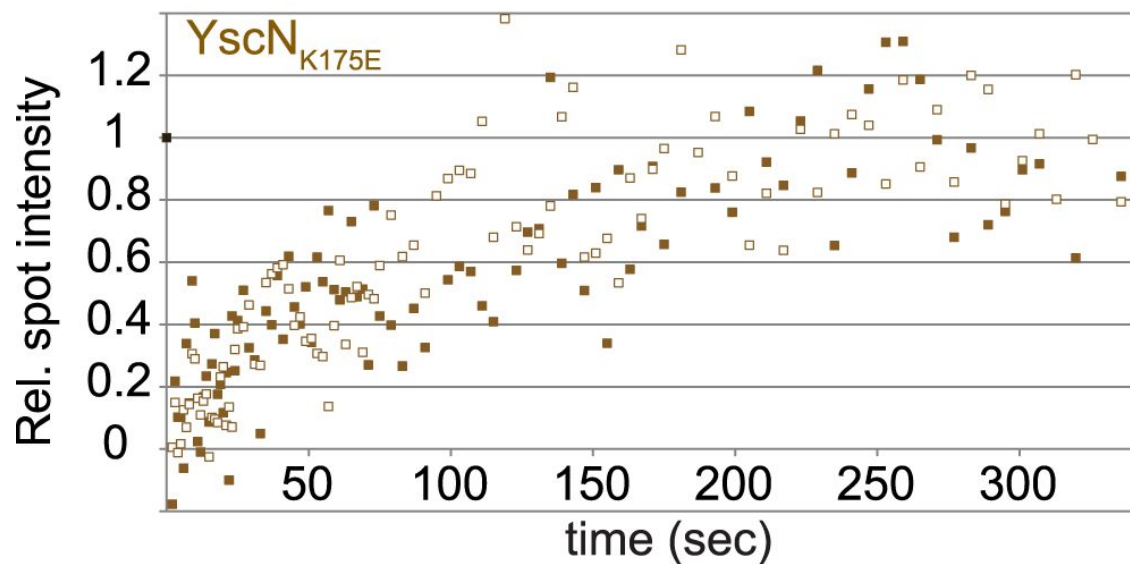
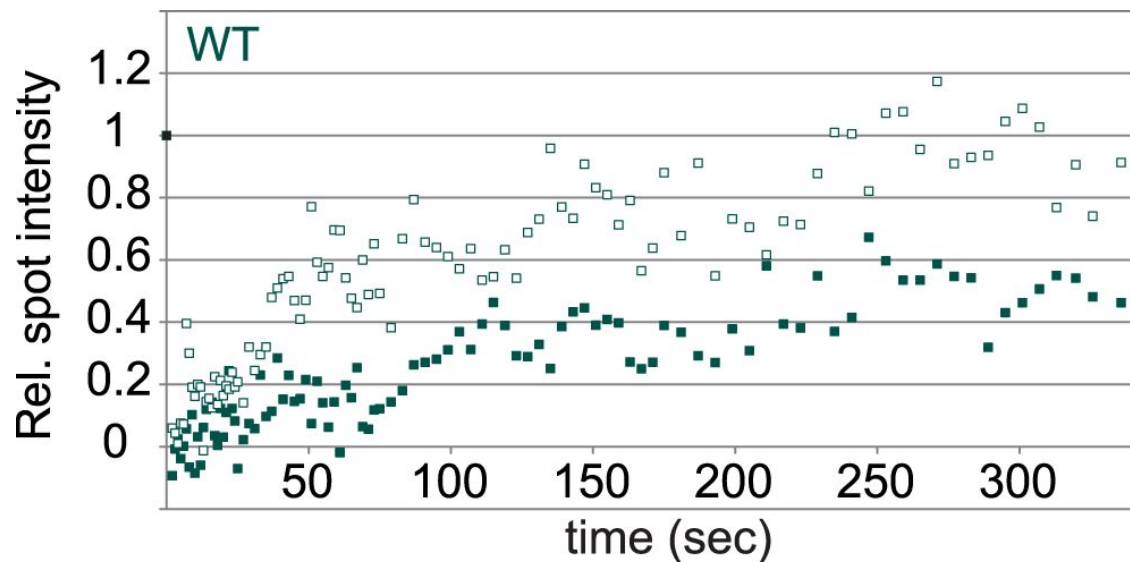
**



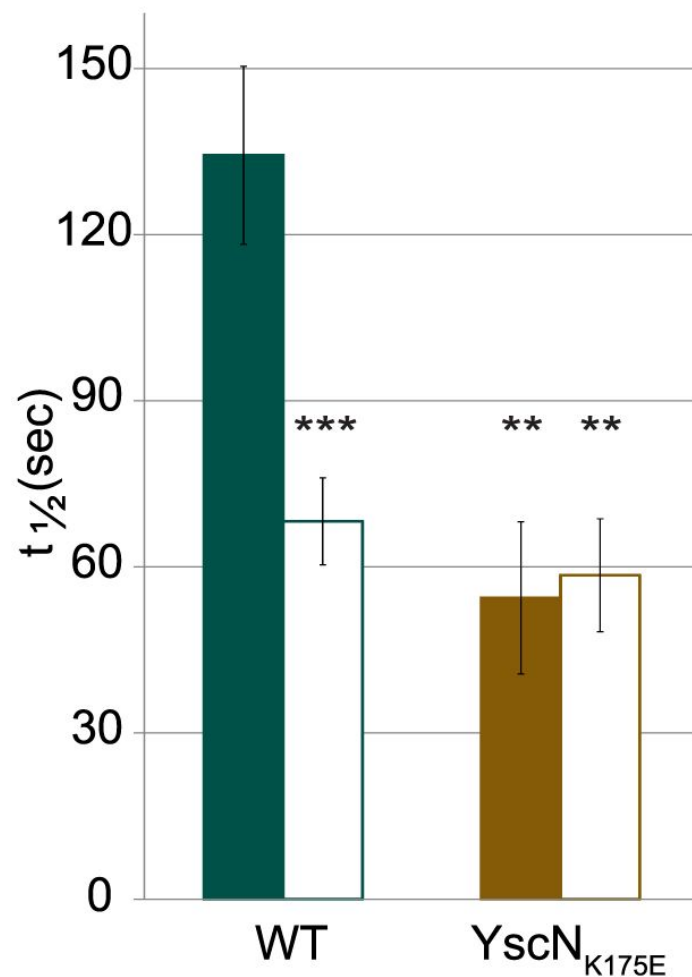
B







■ non-secreting conditions (*off*)
□ secreting conditions (*on*)



Idea

Present a couple of constructive examples: from poor to great