Data Visualization 6: Effective Charts

Stat 133 with Gaston Sanchez

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

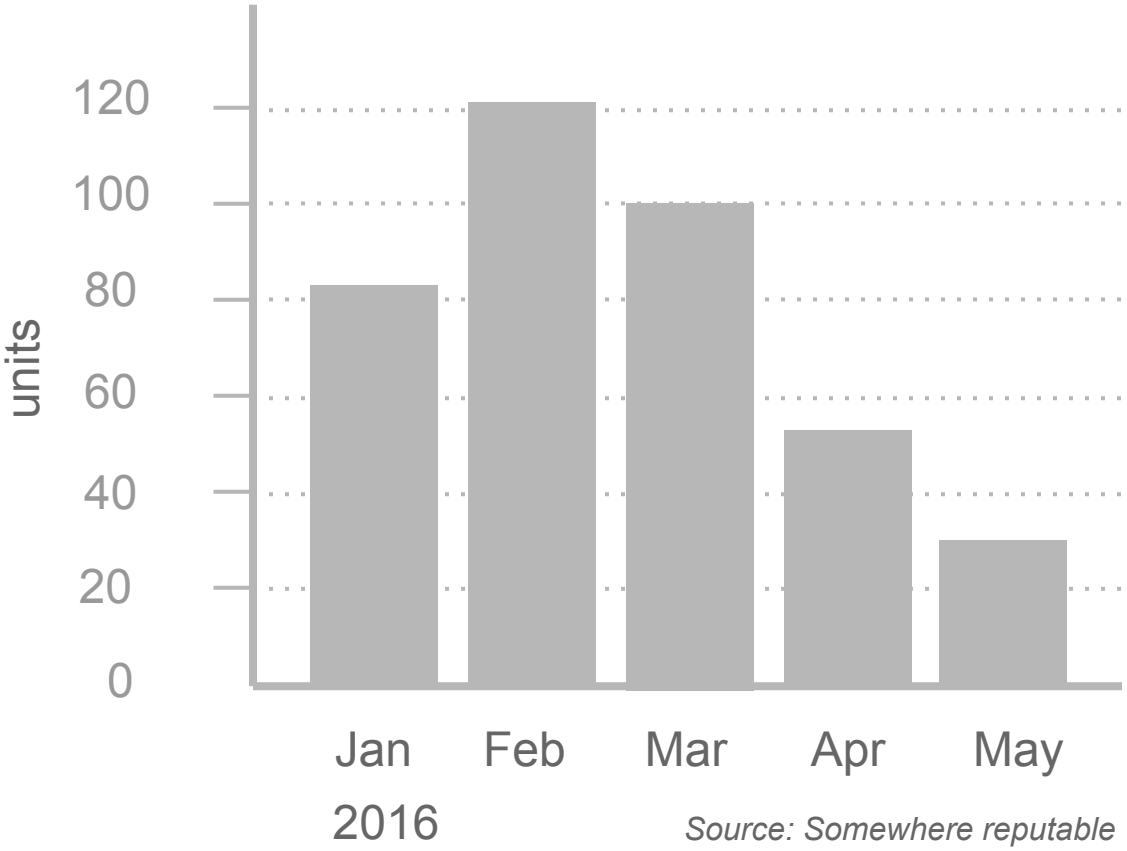
> “The ability to create effective charts and graphs has become almost as indispensable as good writing.”
>
> Donna Wong

Not yet :(

Graphical Elements and Some Vocabulary

Title of this Graph

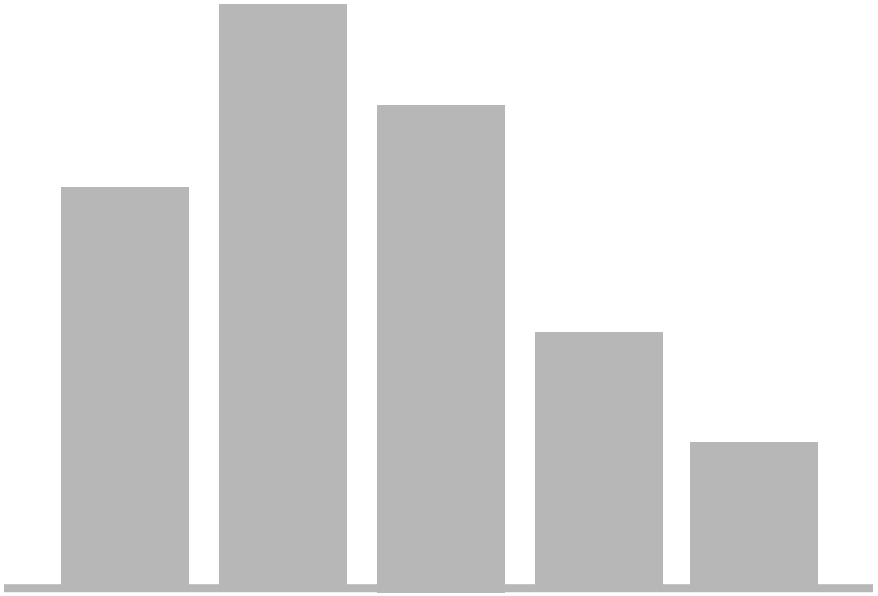
A description of the data or something worth highlighting to set the stage



Source: Somewhere reputable

Visual Cues

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

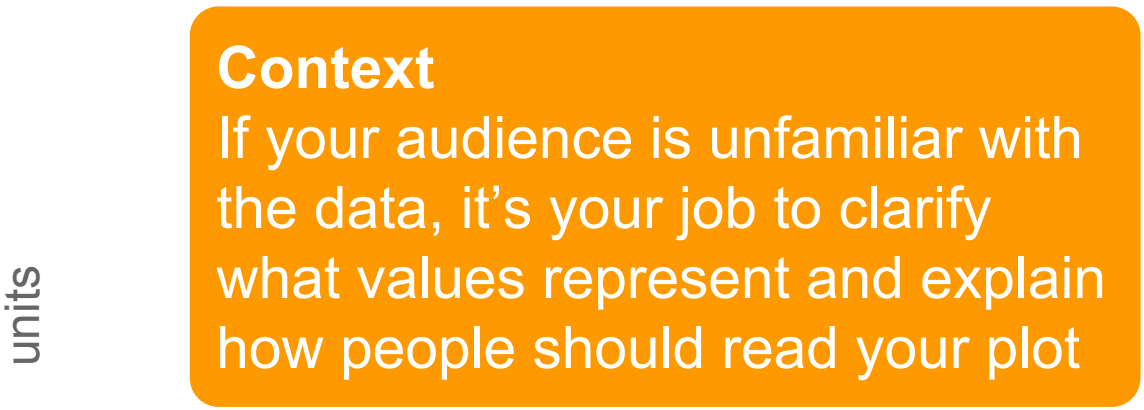


Coordinate System
Mapping data requires a system of coordinates: cartesian, polar, etc

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

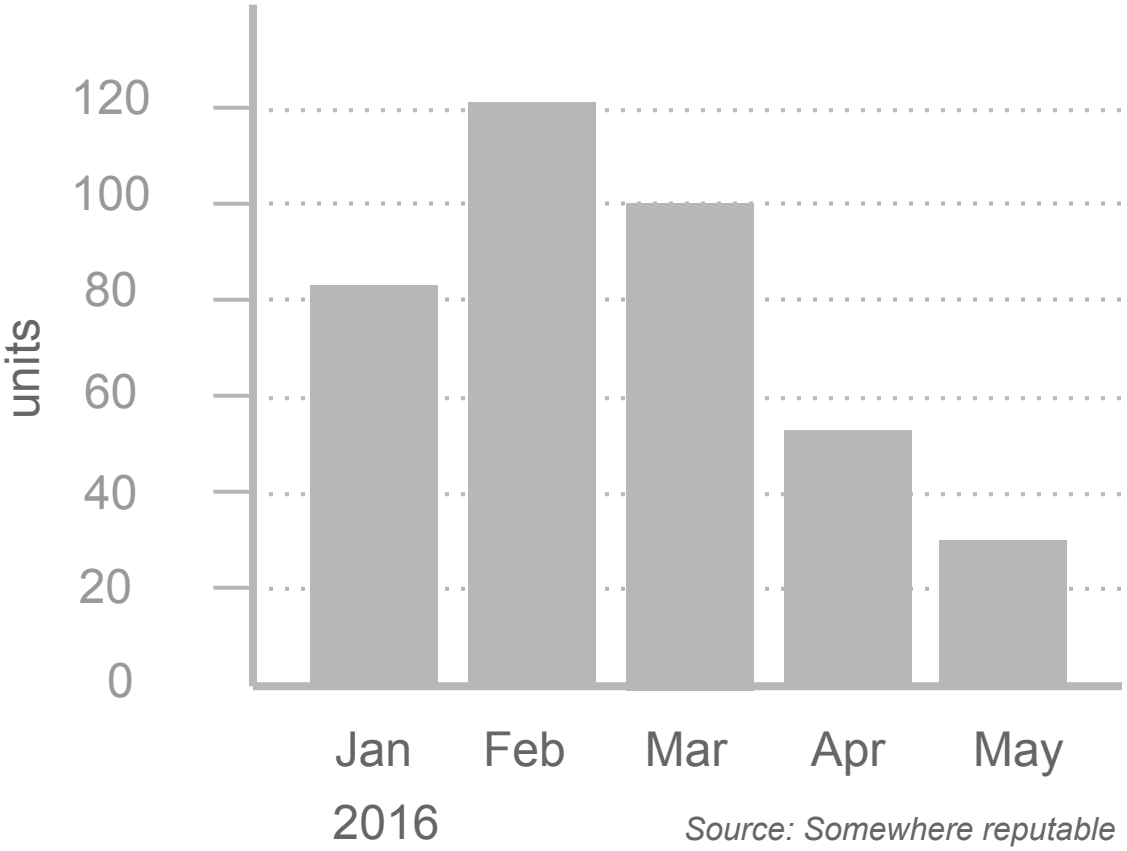
Title of this Graph

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



Title of this Graph

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

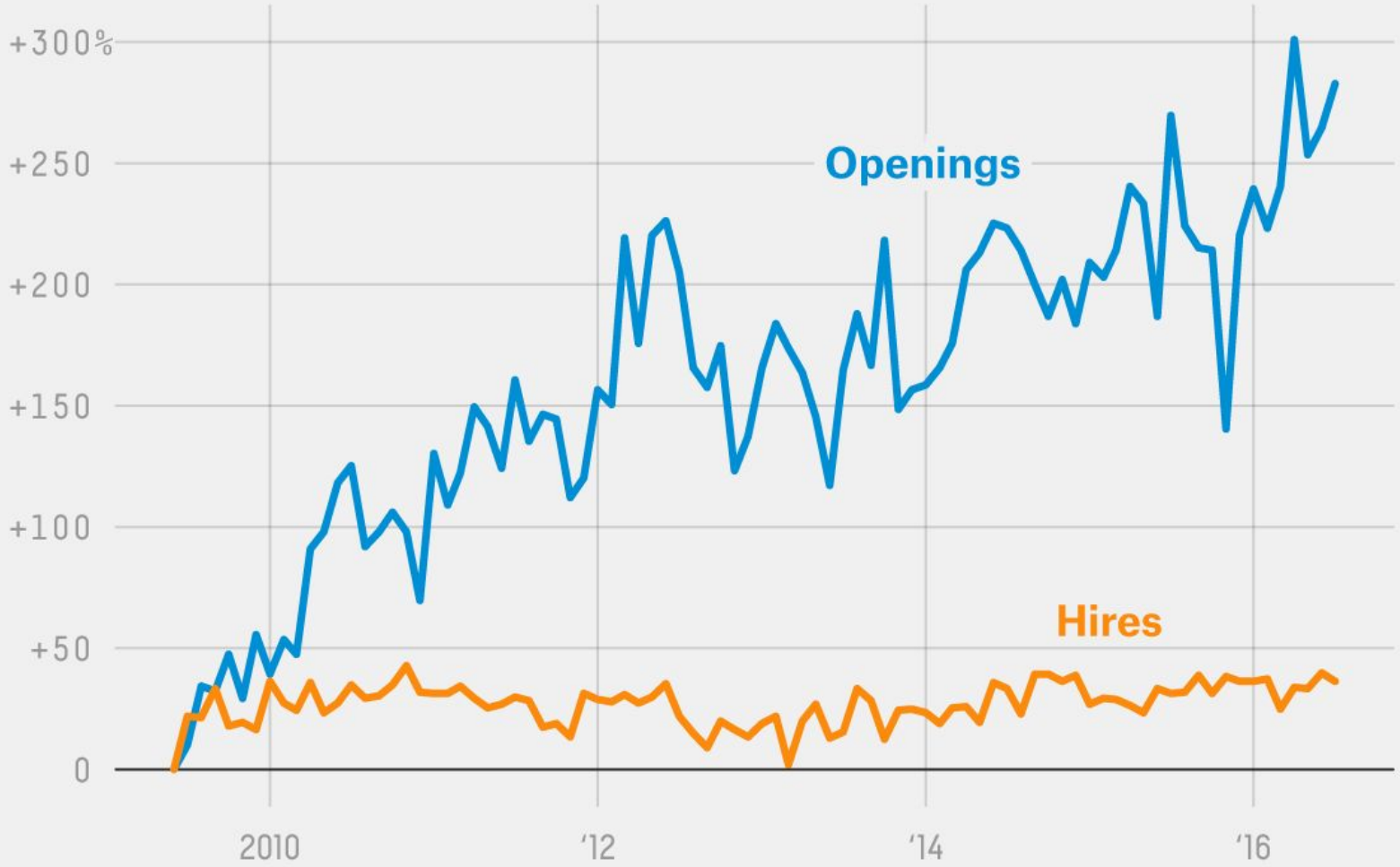


Source: Somewhere reputable

Chart Elements

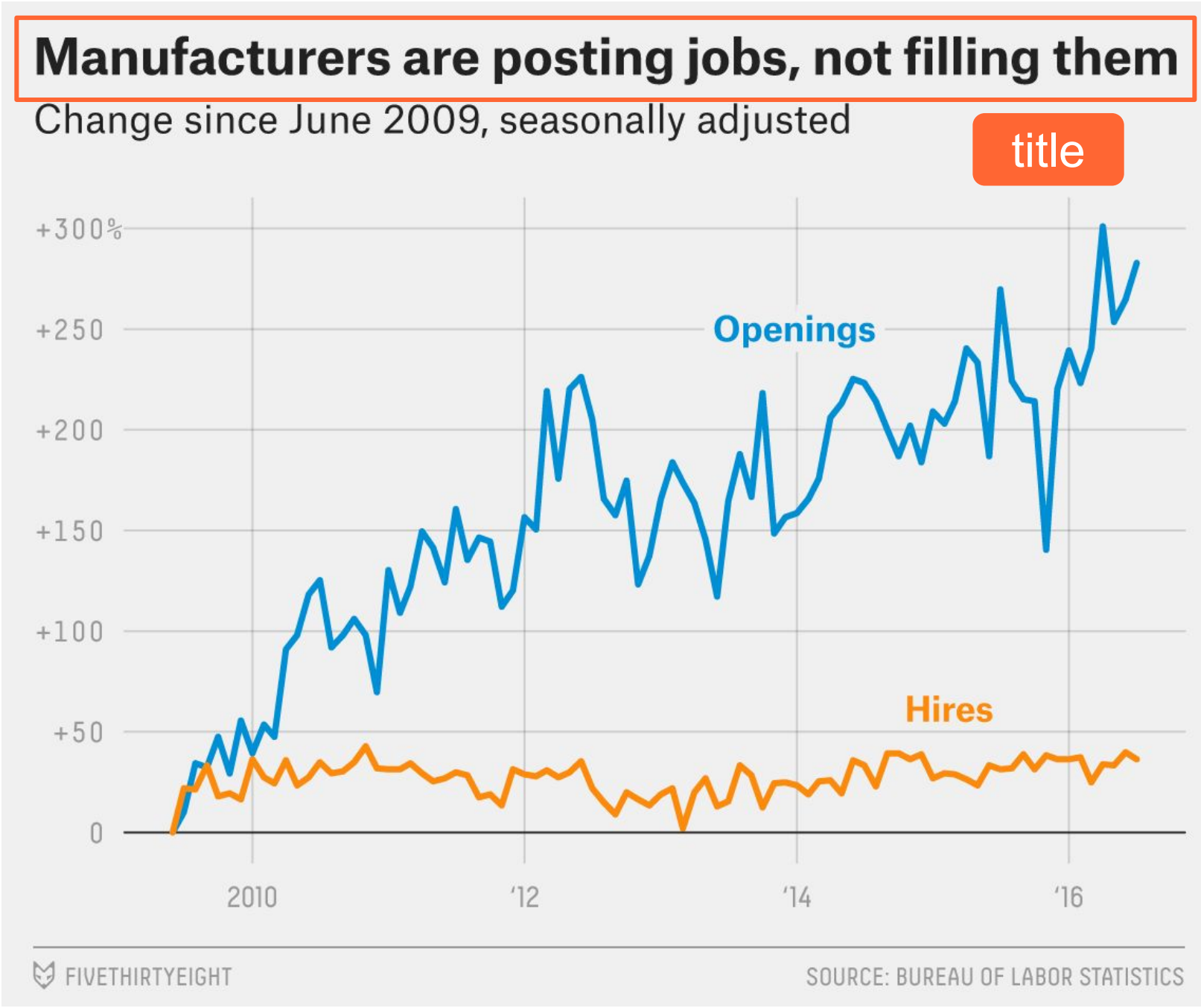
Manufacturers are posting jobs, not filling them

Change since June 2009, seasonally adjusted



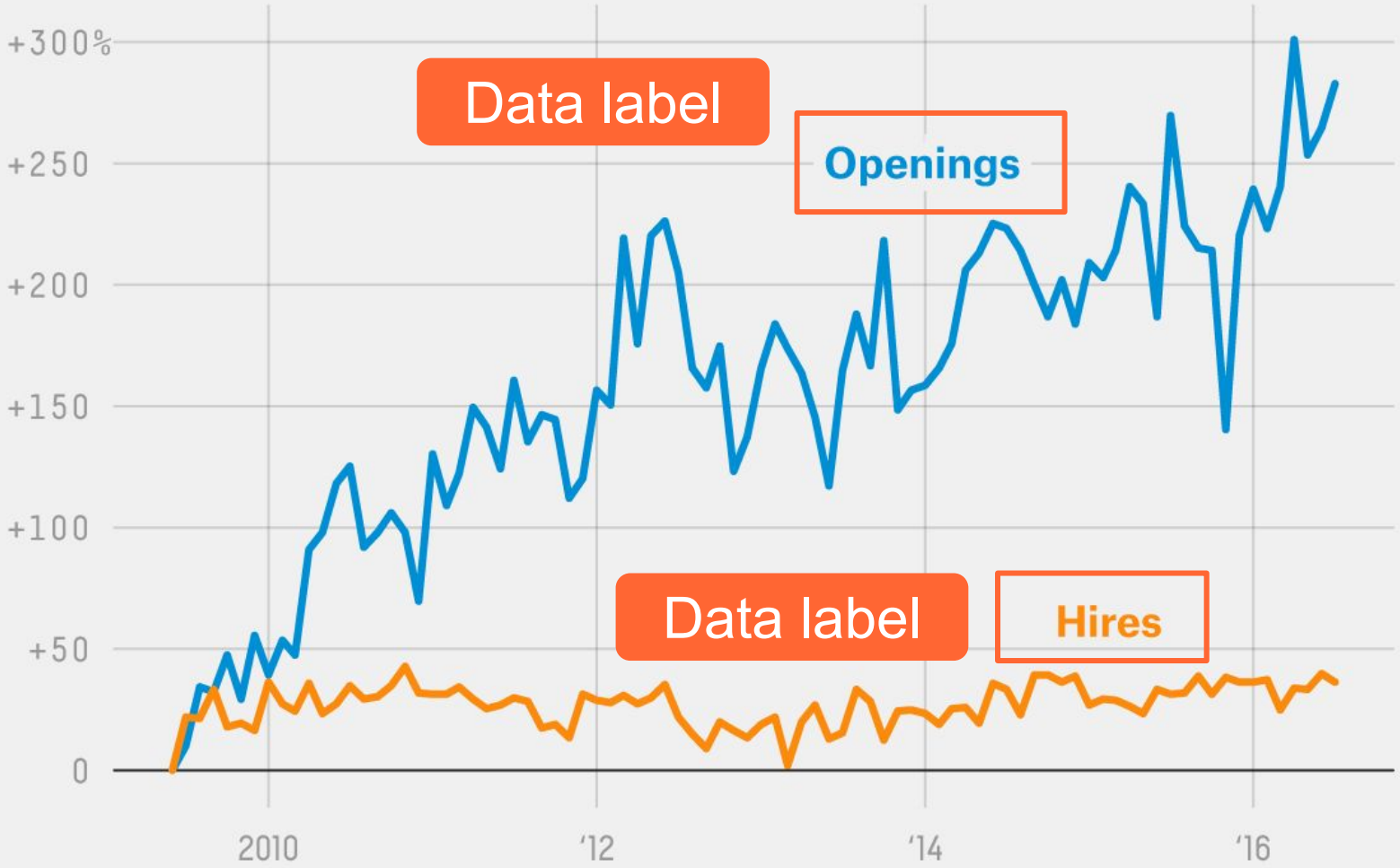
FIVETHIRTYEIGHT

SOURCE: BUREAU OF LABOR STATISTICS


Manufacturers are posting jobs, not filling them
Change since June 2009, seasonally adjusted
title
+300%
+250
+200
+150
+100
+50
0
Openings
Hires
2010
'12
'14
'16
FIVETHIRTYEIGHT
SOURCE: BUREAU OF LABOR STATISTICS

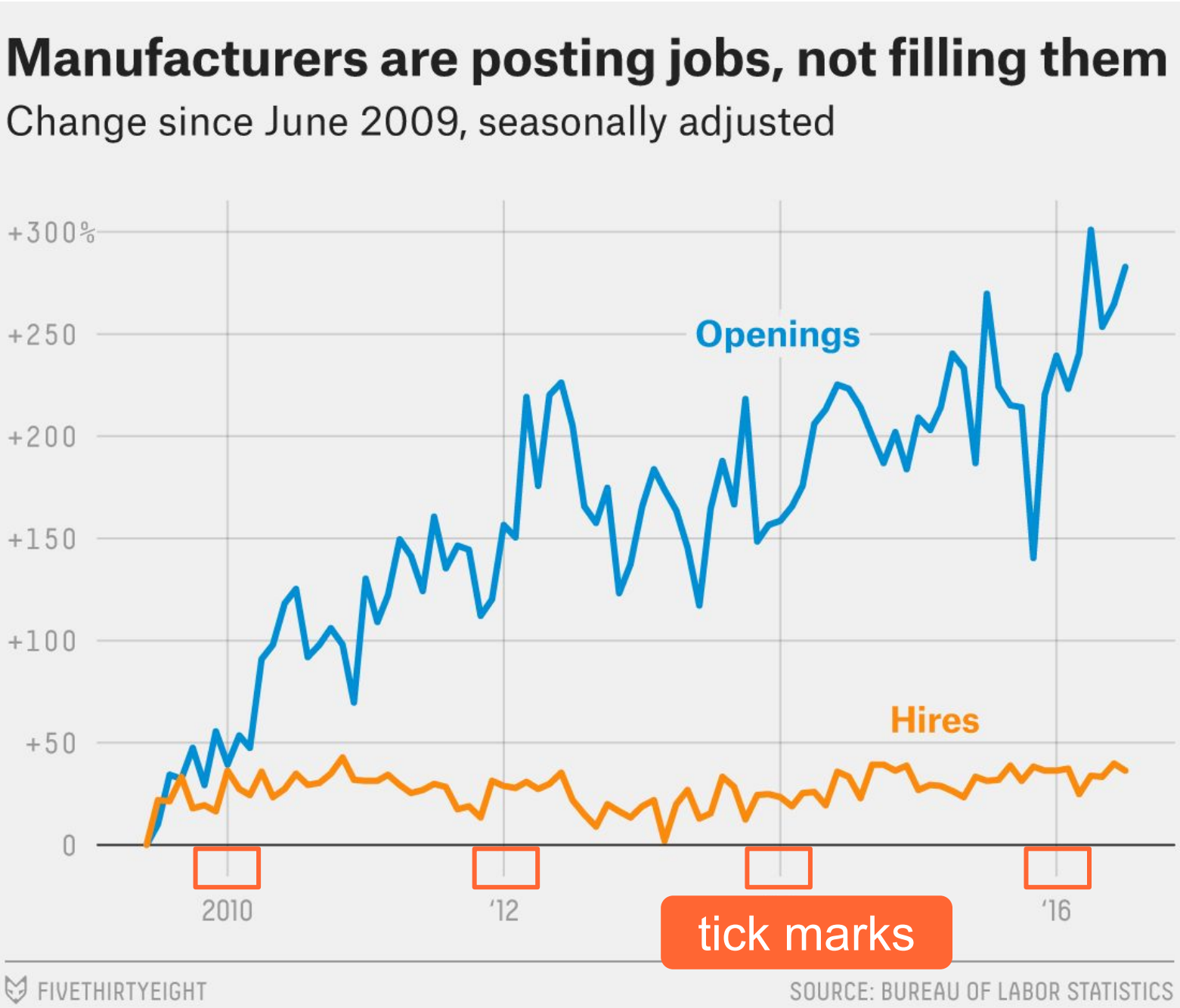
Manufacturers are posting jobs, not filling them

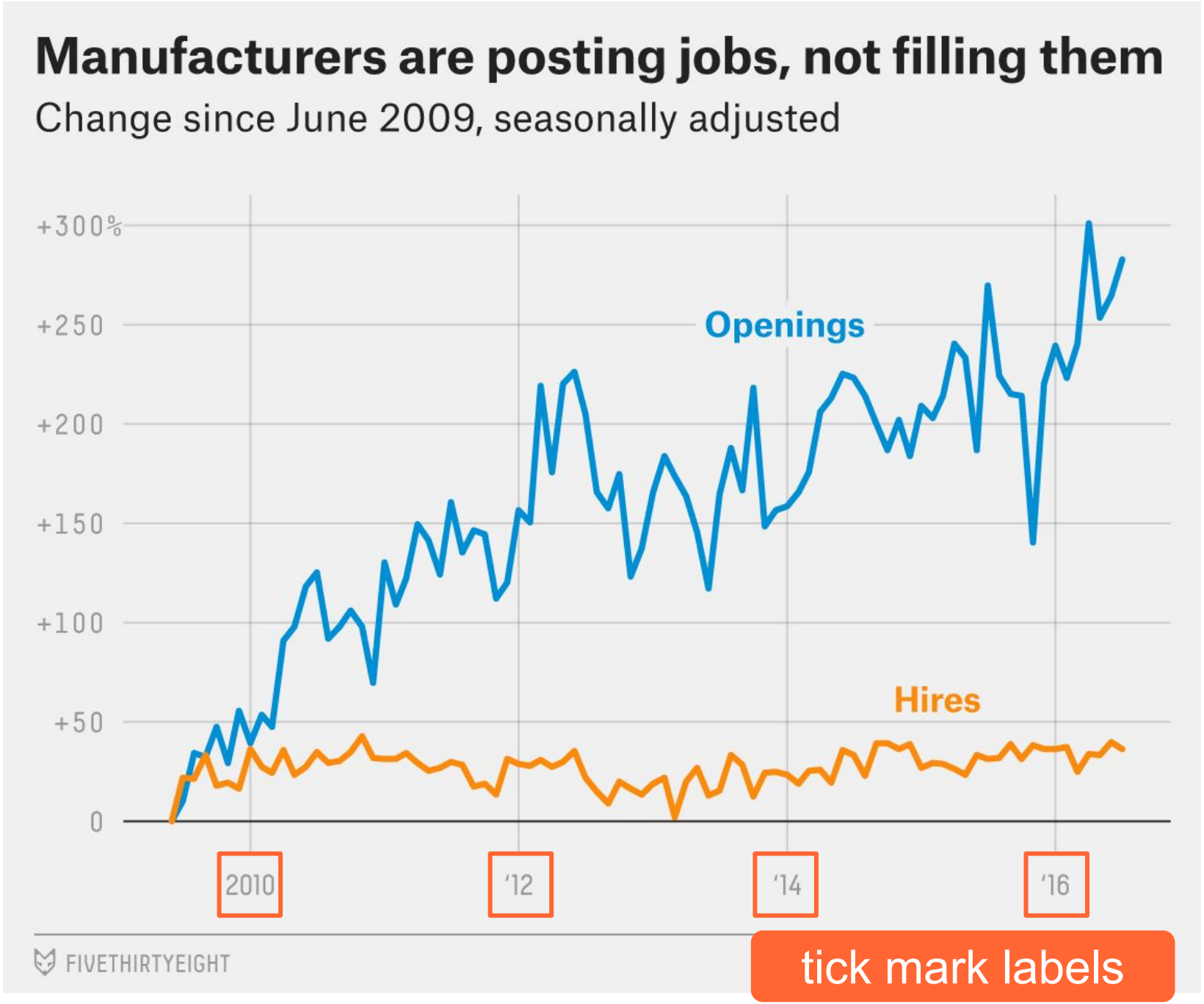
Change since June 2009, seasonally adjusted



FIVETHIRTYEIGHT

SOURCE: BUREAU OF LABOR STATISTICS


Manufacturers are posting jobs, not filling them
Change since June 2009, seasonally adjusted
+300%
+250
+200
+150
+100
+50
0
Openings
Hires
2010
'12
'16
tick marks
FIVETHIRTYEIGHT
SOURCE: BUREAU OF LABOR STATISTICS



tick mark labels

Manufacturers are posting jobs, not filling them

Change since June 2009, seasonally adjusted

Grid lines

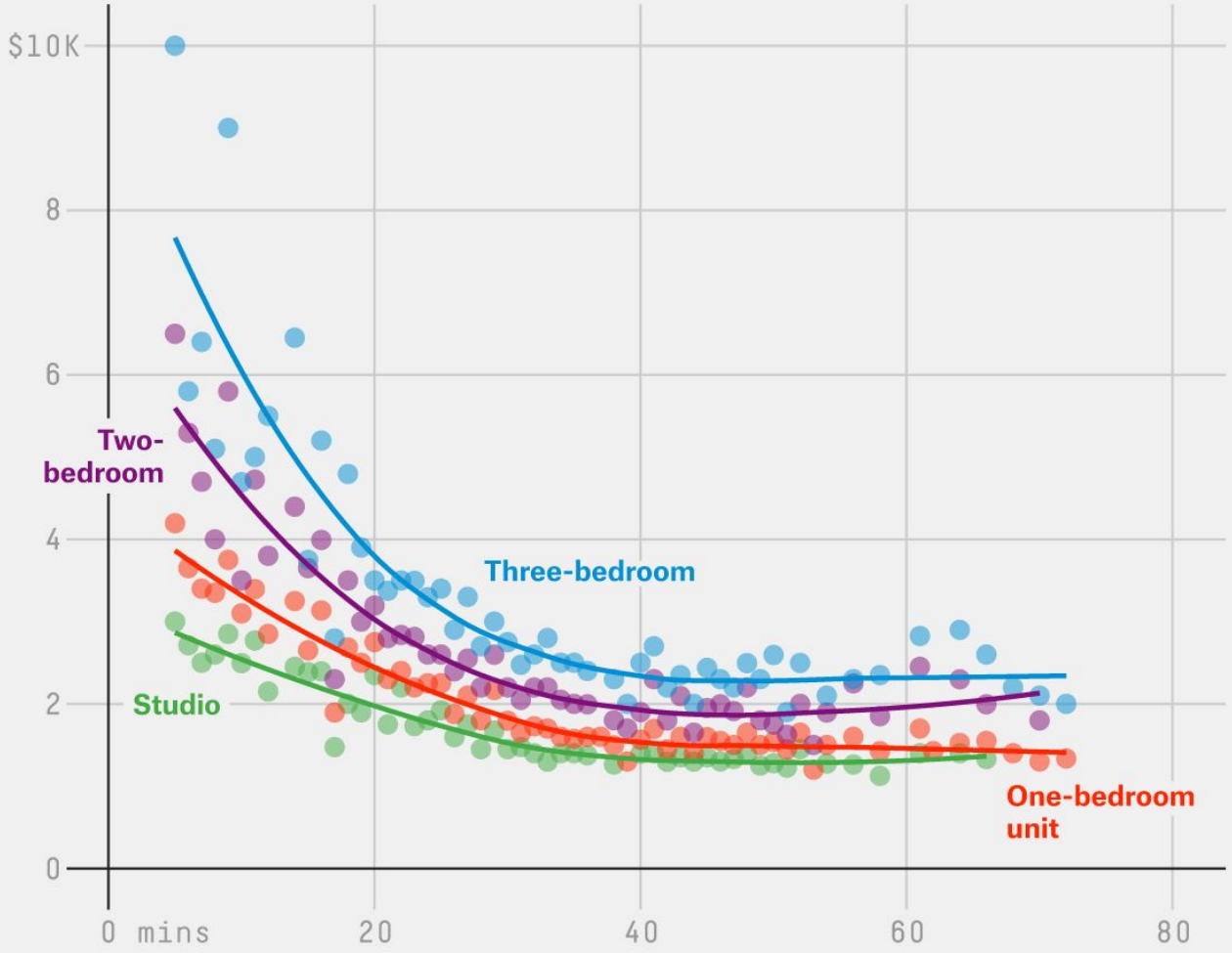
FIVETHIRTYEIGHT

SOURCE: BUREAU OF LABOR STATISTICS

Another example

New Yorkers pay up for a shorter commute

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



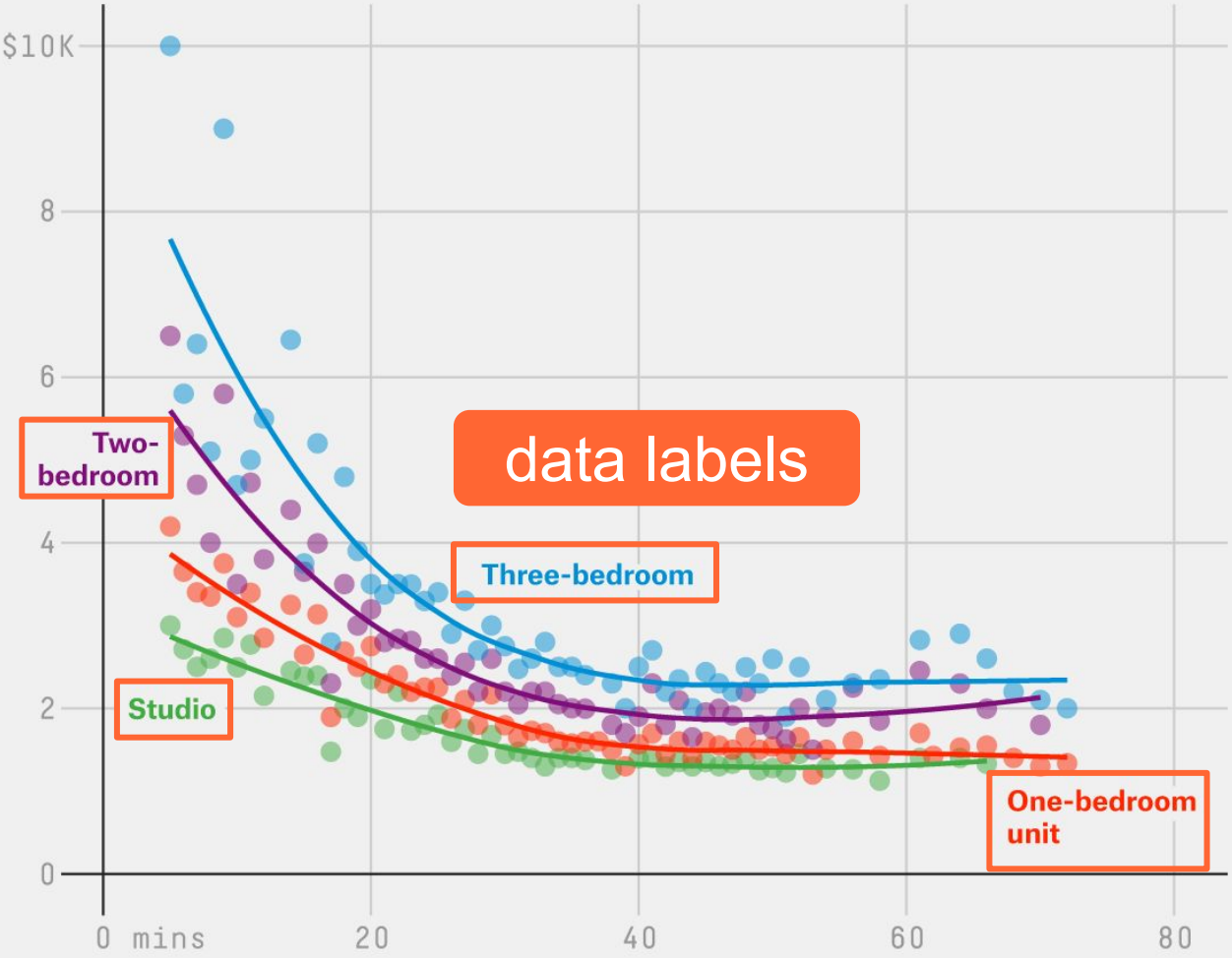
Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

FIVETHIRTYEIGHT

SOURCE: STREETEASY

New Yorkers pay up for a shorter commute

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



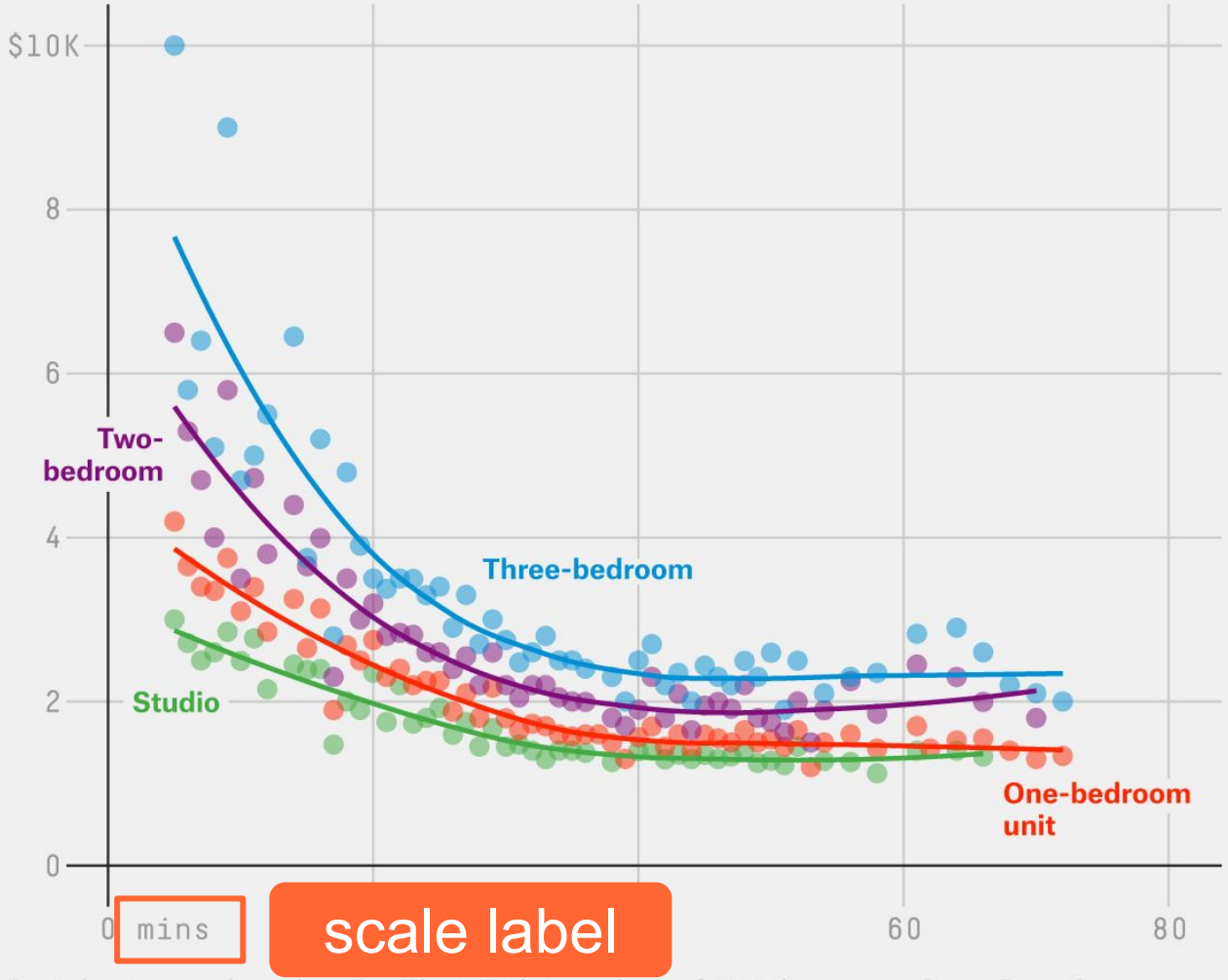
Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

FIVETHIRTYEIGHT

SOURCE: STREETEASY

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.

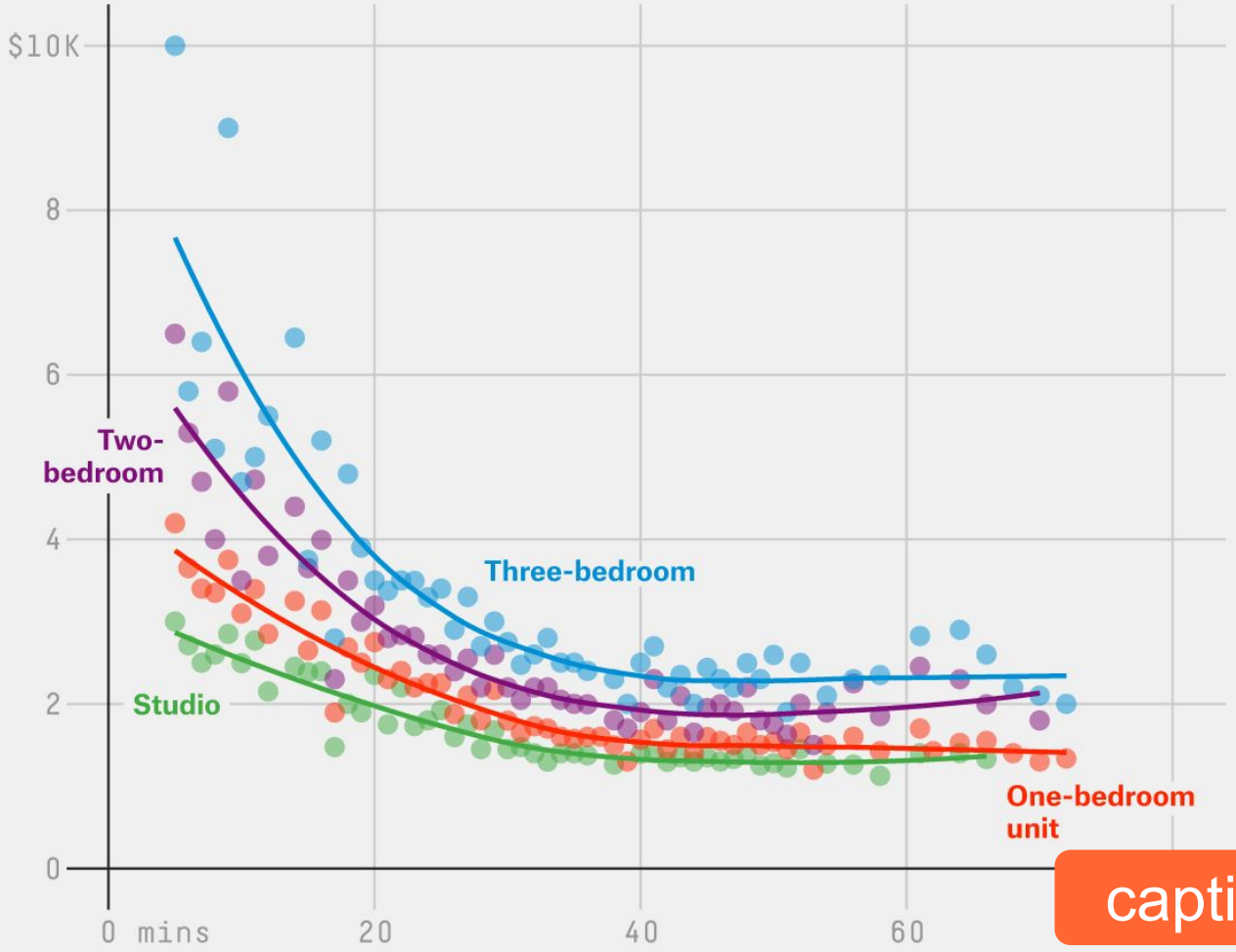
FIVETHIRTYEIGHT

SOURCE: STREETEASY

New Yorkers pay up for a shorter commute
Median monthly NYC rent in 2015 vs. commute time by subway
$10K
scale label
8
6
4
2
0
Two-bedroom
Three-bedroom
Studio
One-bedroom unit
0 mins
20
40
60
80
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.
FIVETHIRTYEIGHT
SOURCE: STREETEASY

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.

FIVETHIRTYEIGHT

SOURCE: STREETEASY

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

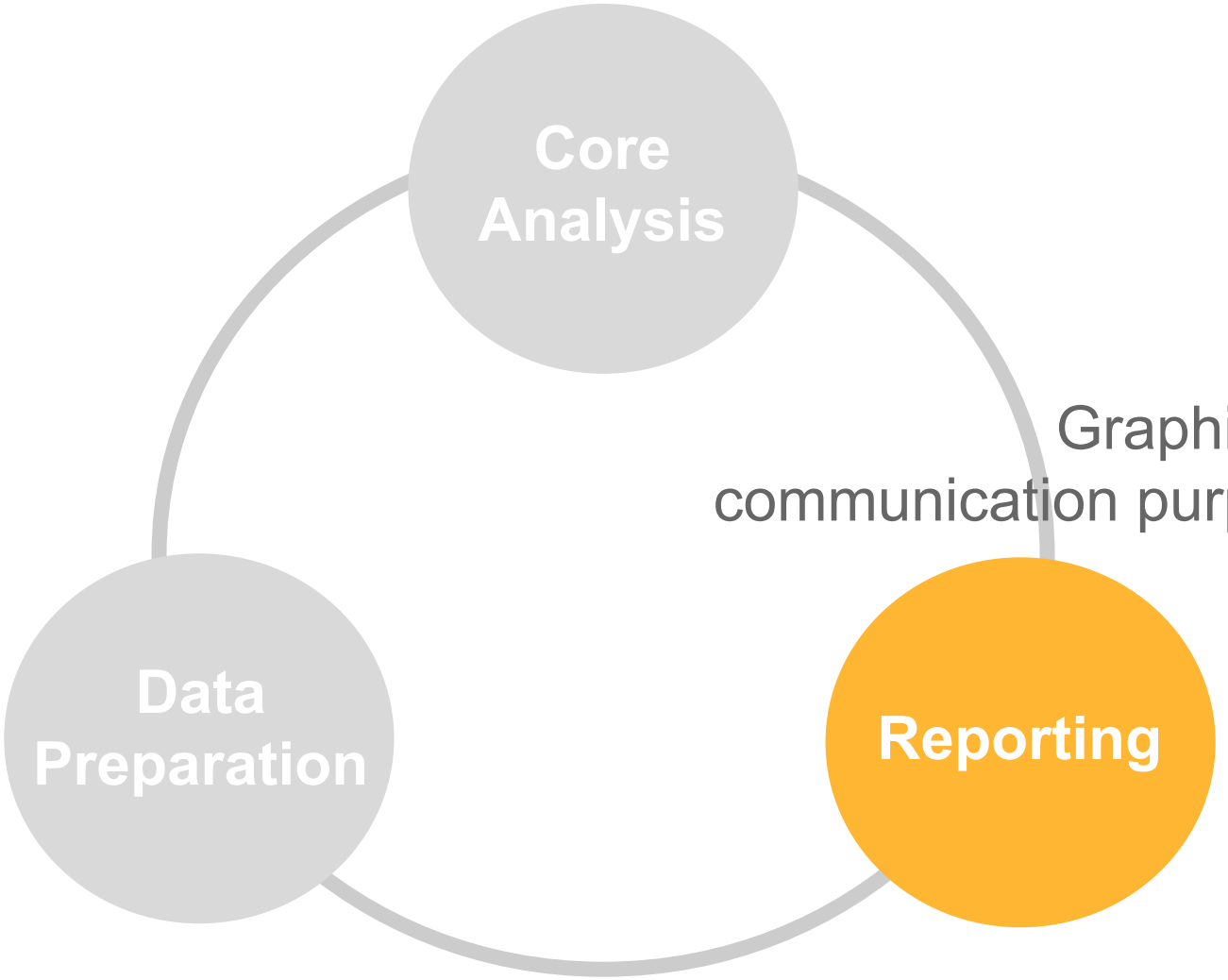
Charts for ...

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

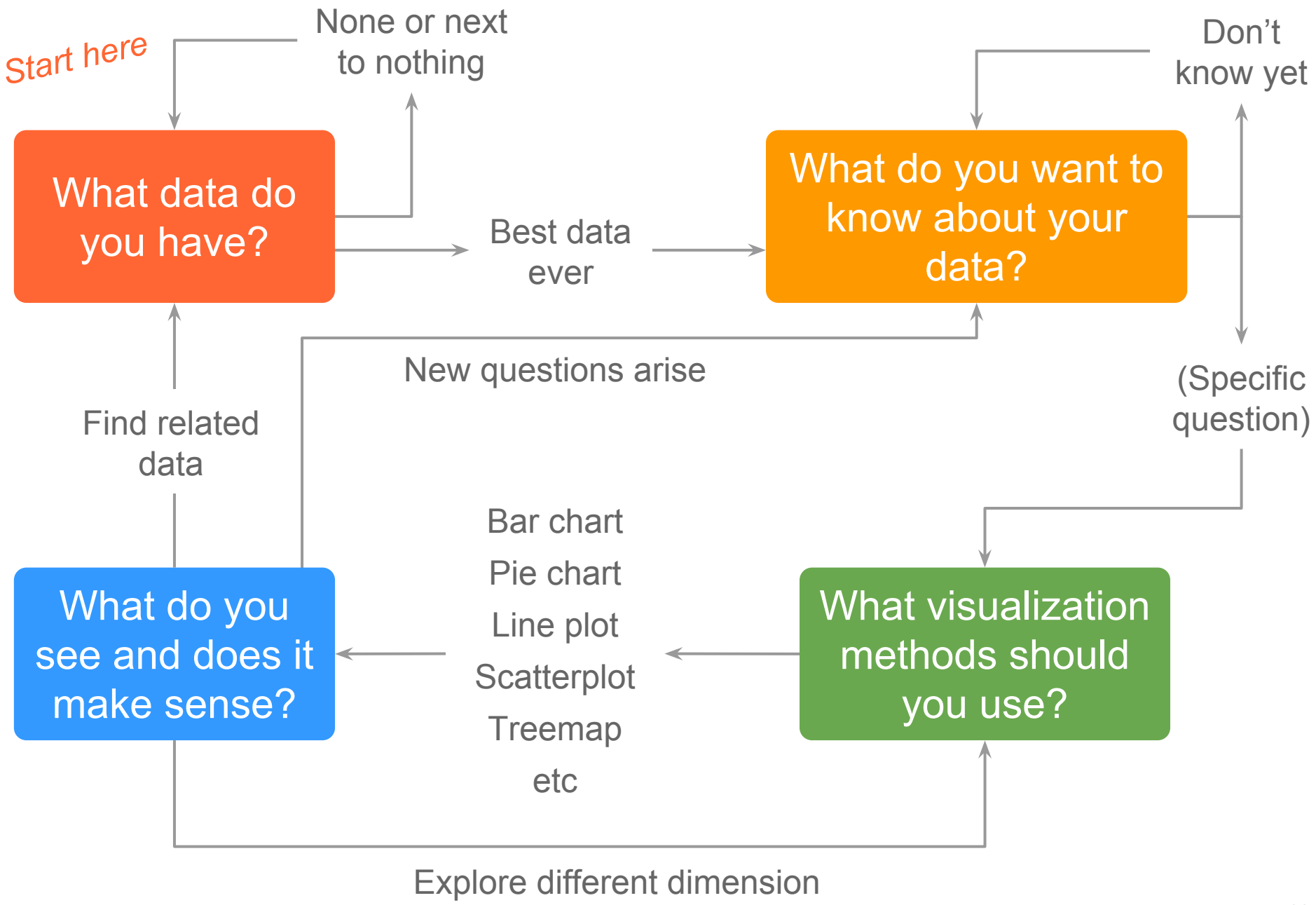
(The Elements of Graphig Data)

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?



The iterative data exploration process by Nathan Yau

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.

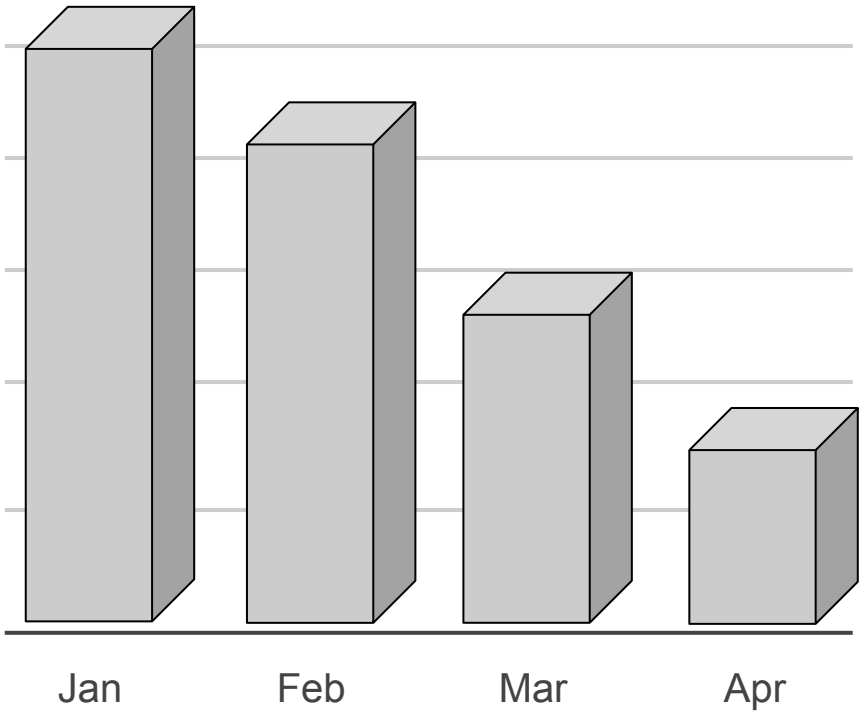
Once you have a good draft, then switch to the computer.

Polish your graph until you get a nice display.

Some Guidelines

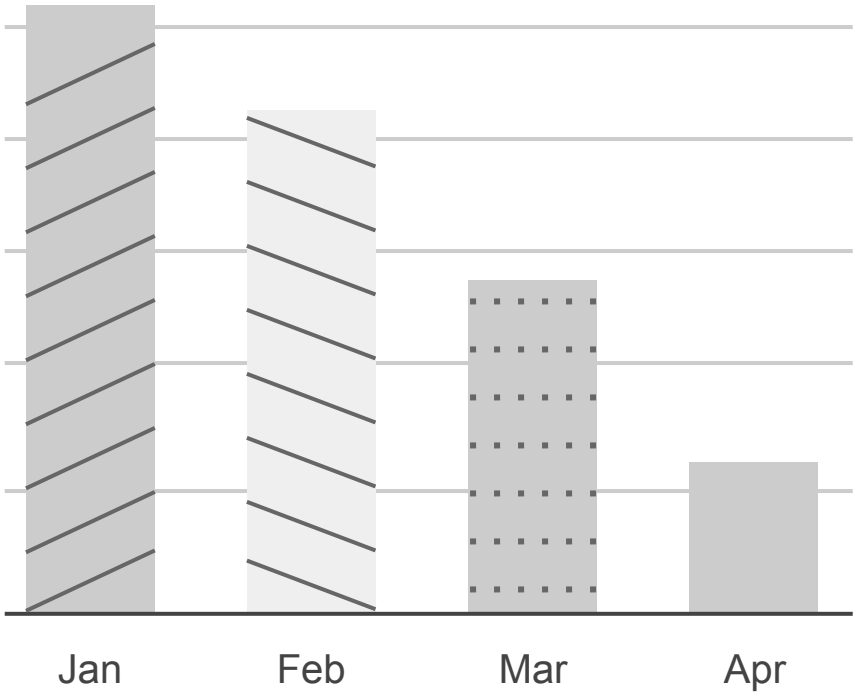
don’t

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



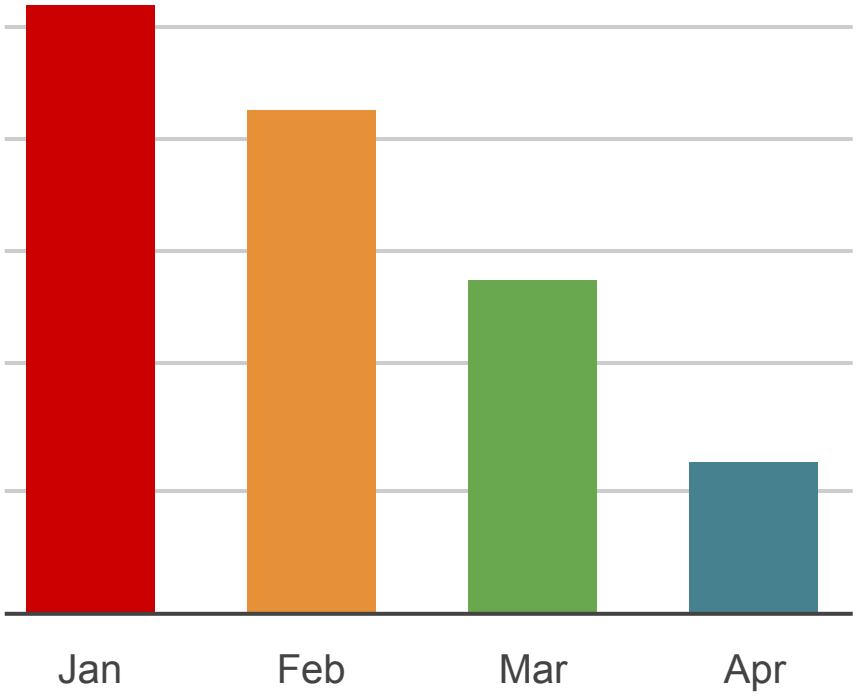
don’t

Don’t use multiple distracting shades



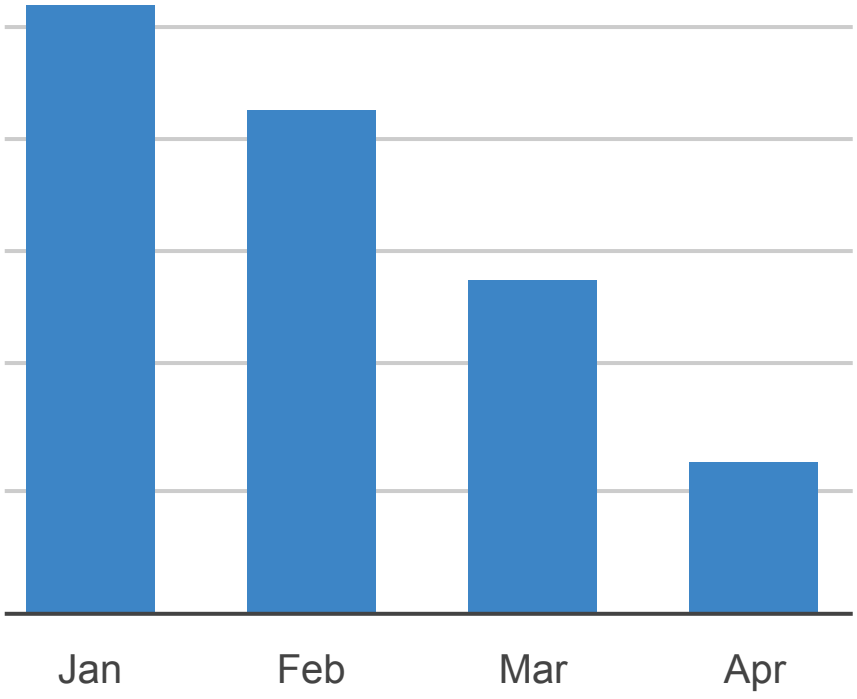
don’t

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



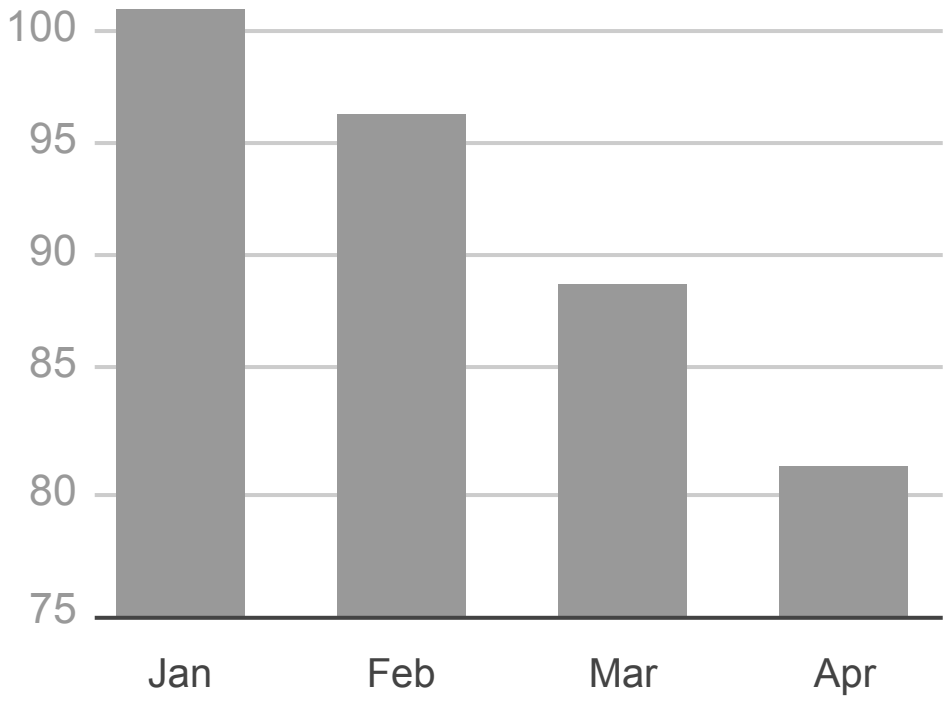
Use the same color to represent the same variable

do



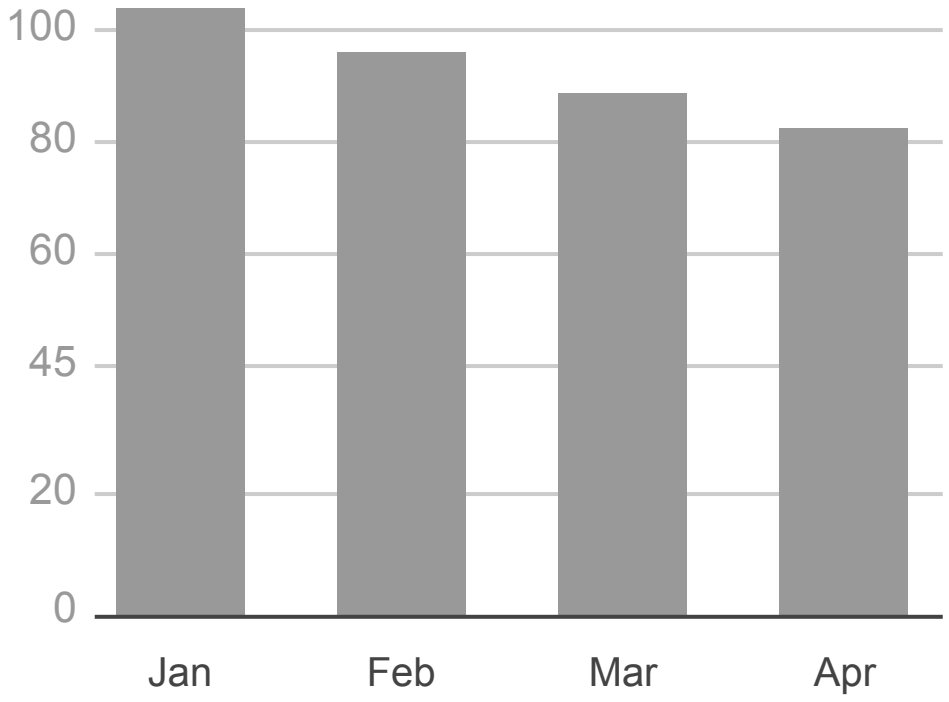
don’t

Don’t truncate baseline of bar charts



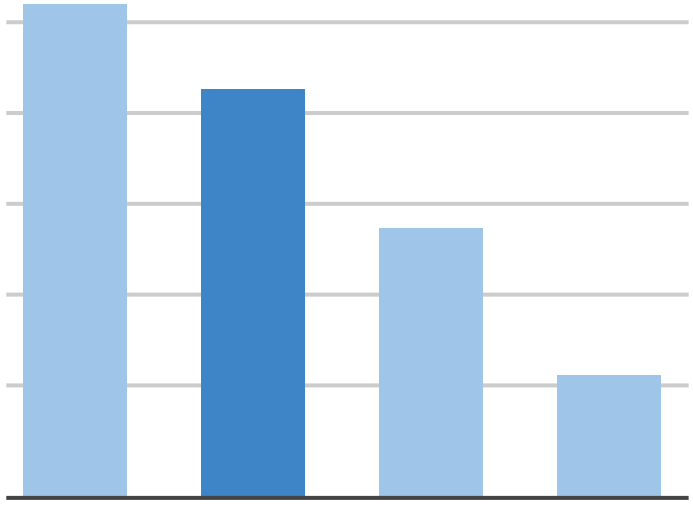
Always start at the zero baseline
No exceptions!

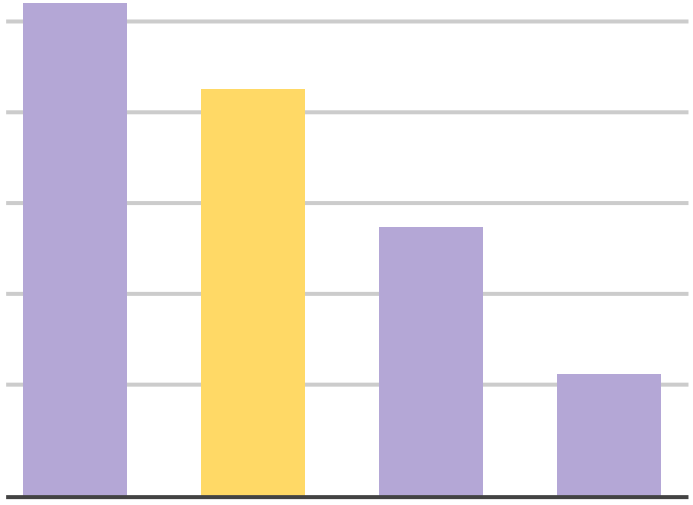
do



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

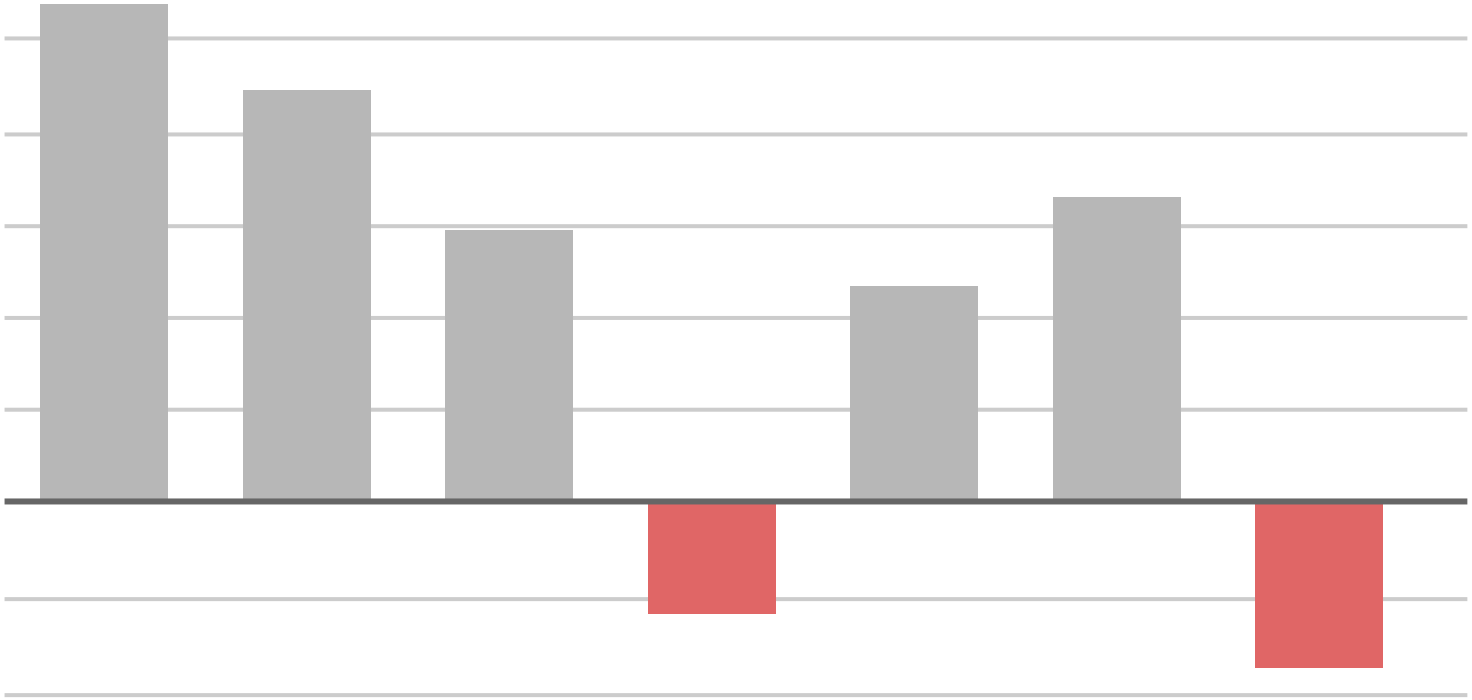
do





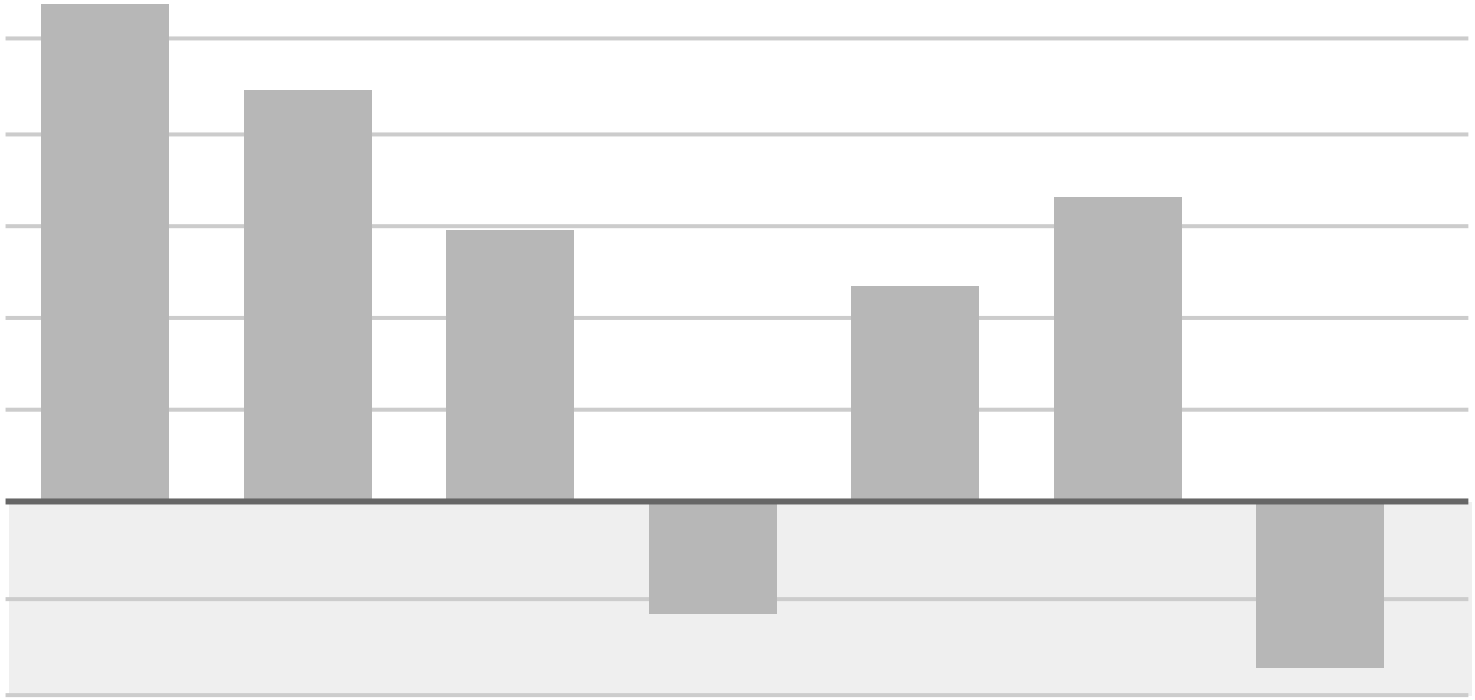
You can use contrasting colors to depict negative values

do



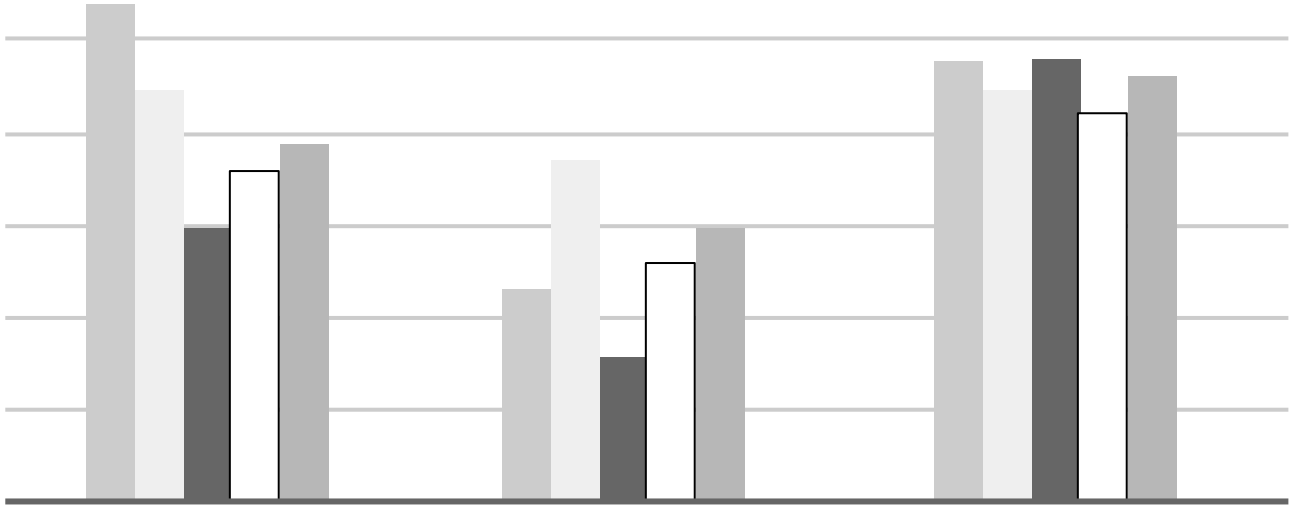
You can use a gray background to identify the negative zone

do

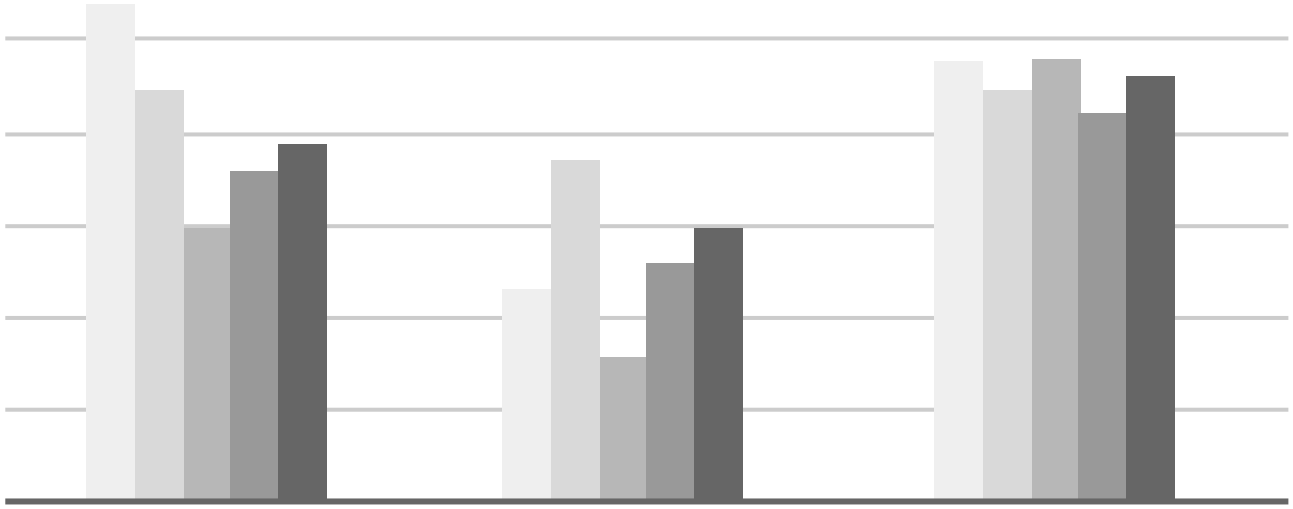


don’t

No zebra pattern



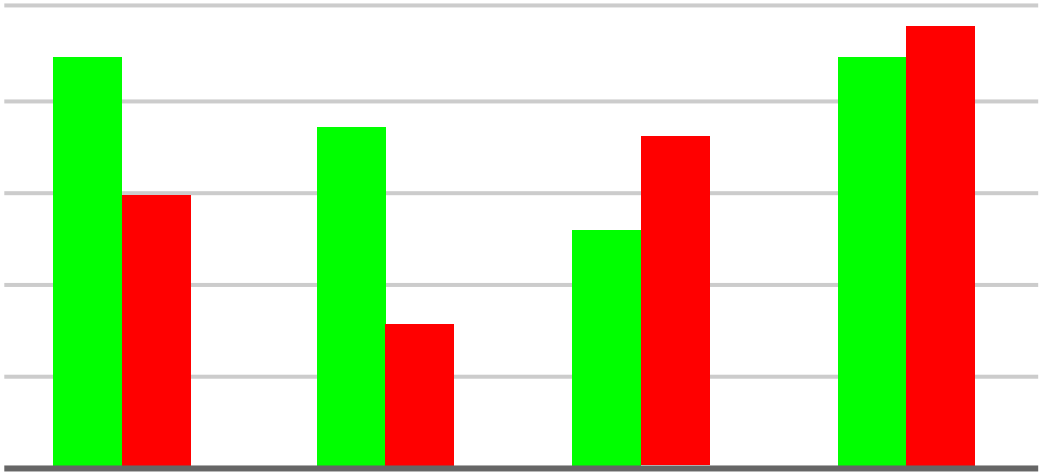
From lightest to darkest



do

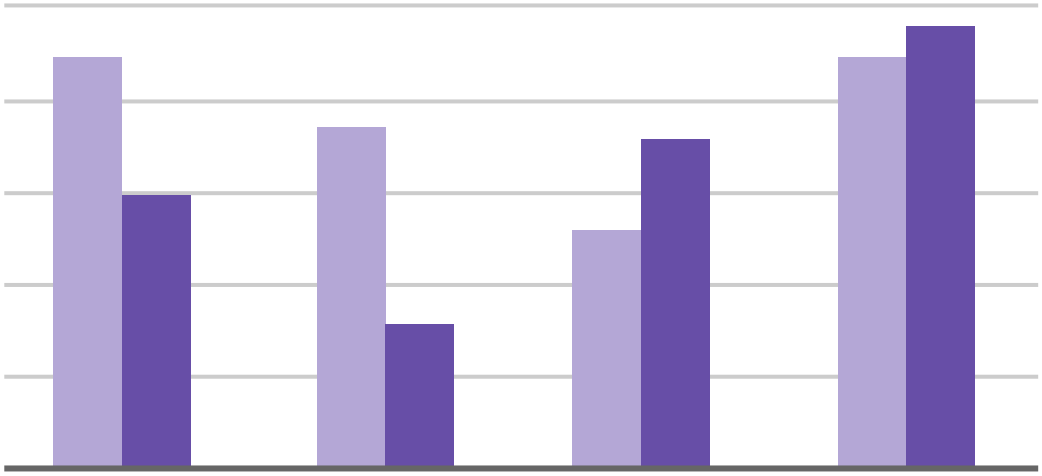
don’t

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



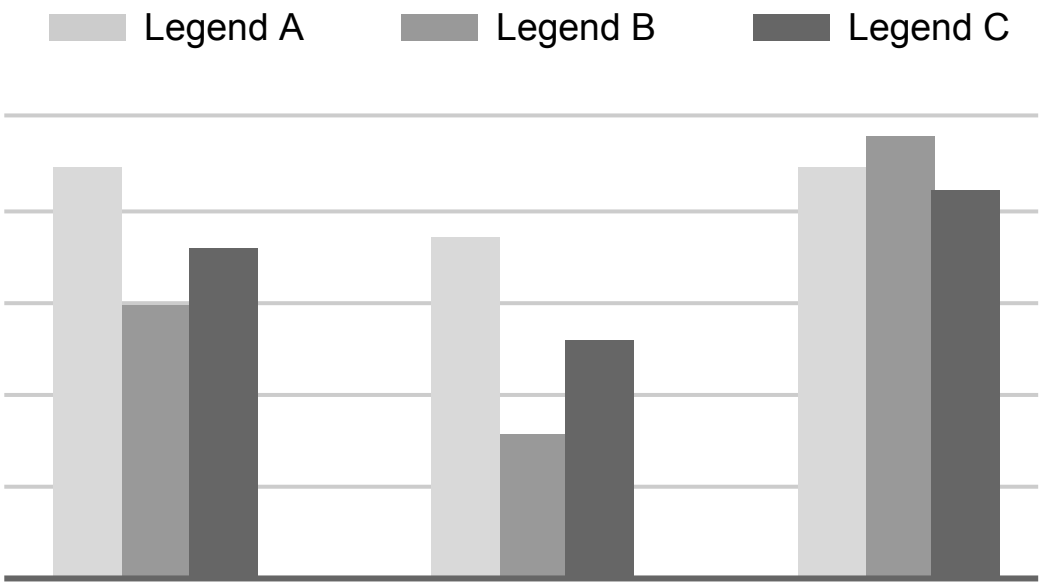
It’s better to use different shades of the same color

do



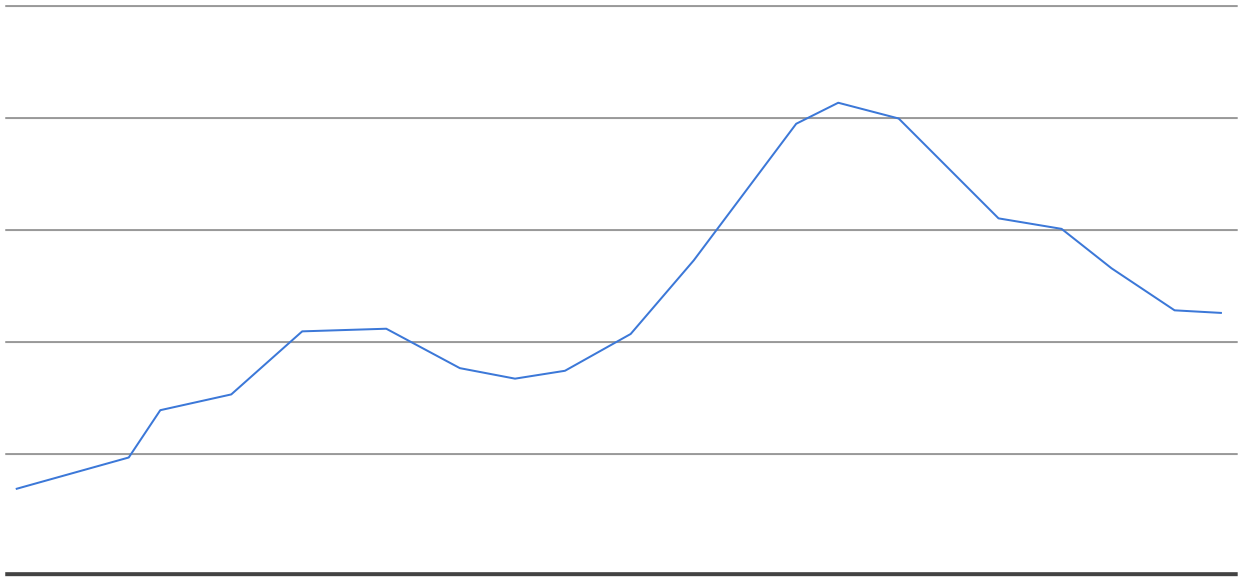
Add legends in the right sequence

do



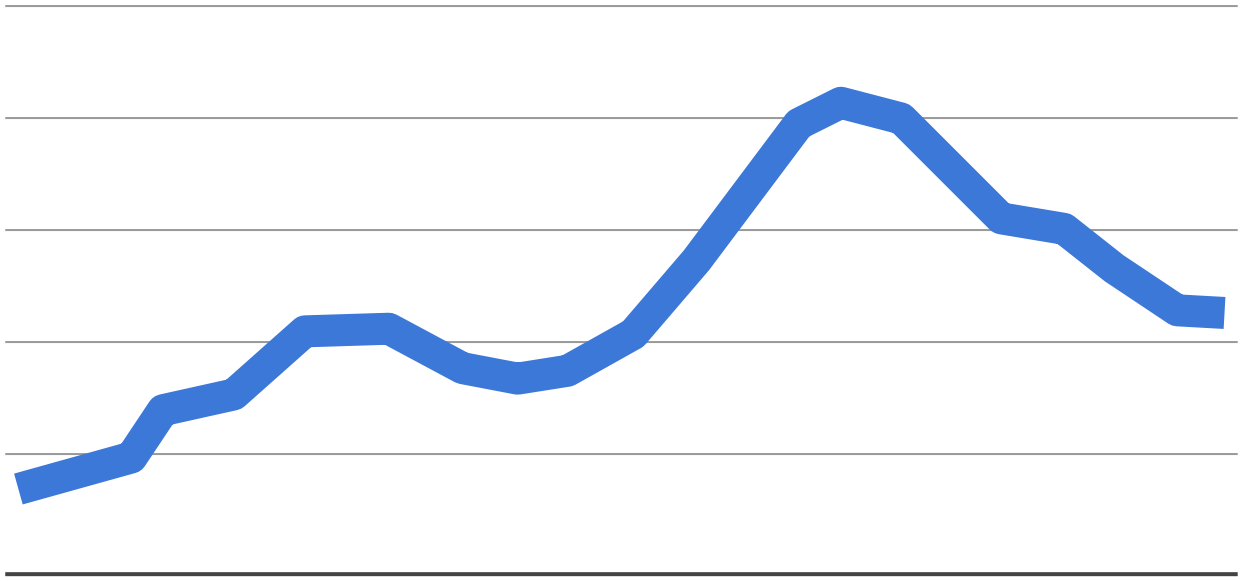
don’t

lines too thin are hard to read

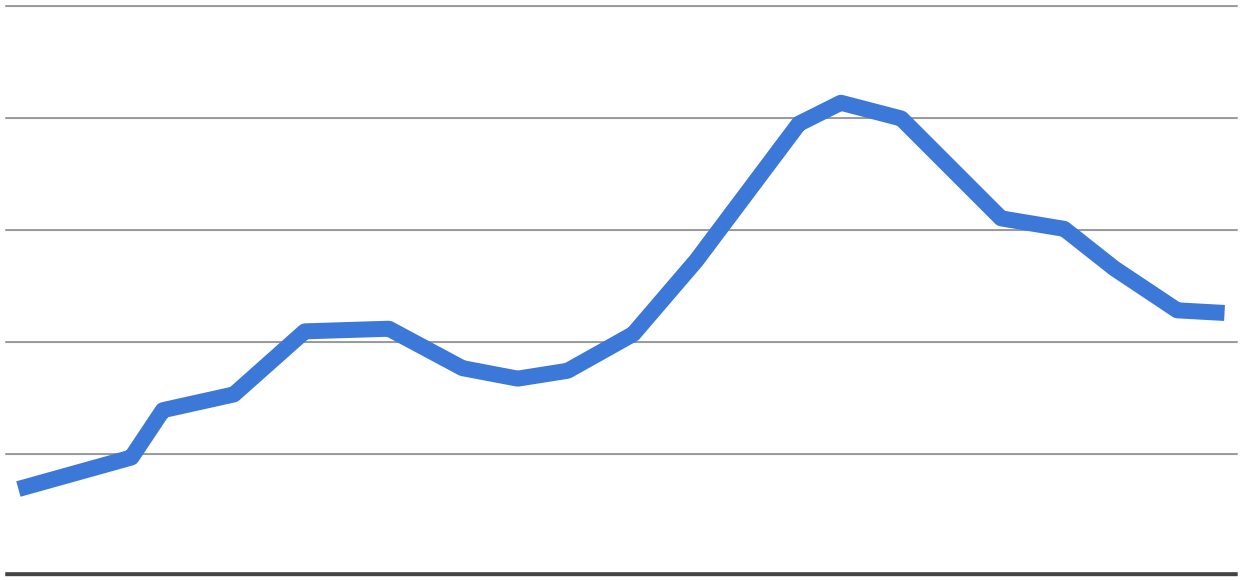


don’t

too thick lines may hide some details



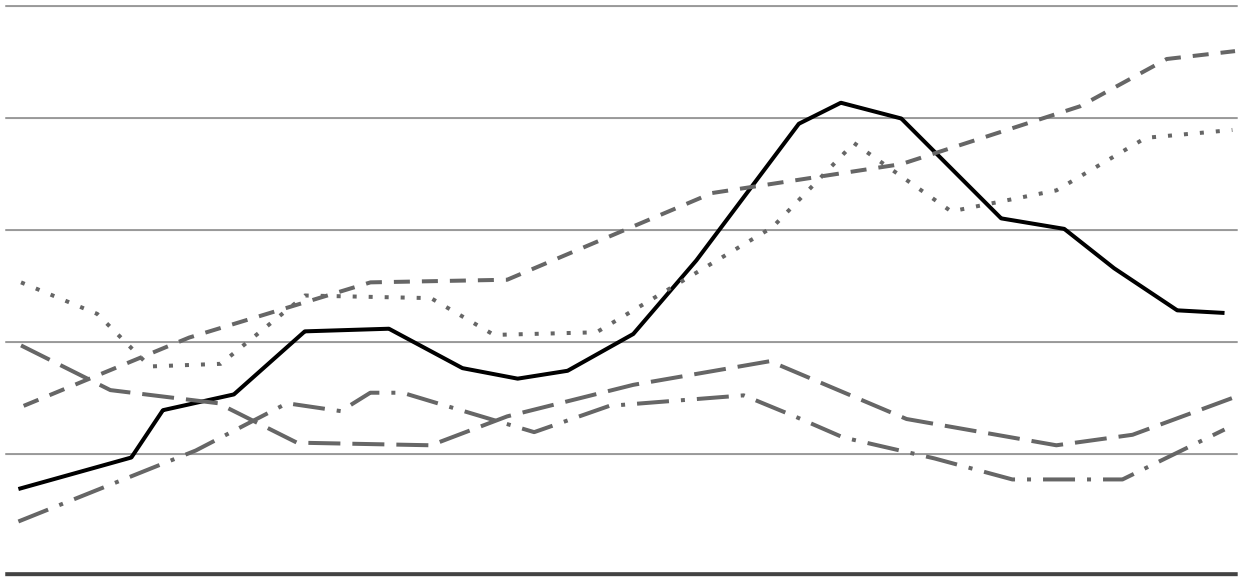
look for “an OK” thickness



do

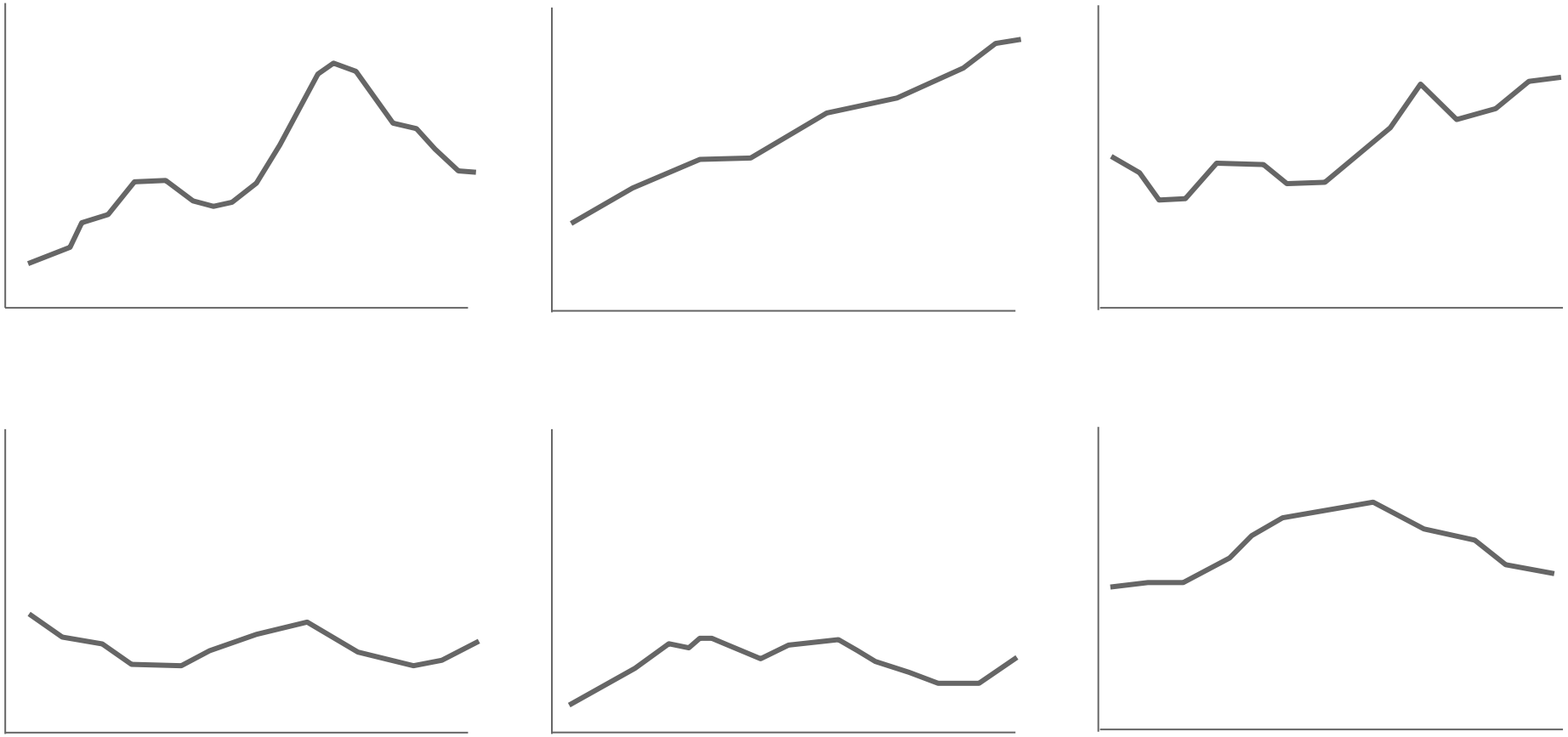
don’t

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



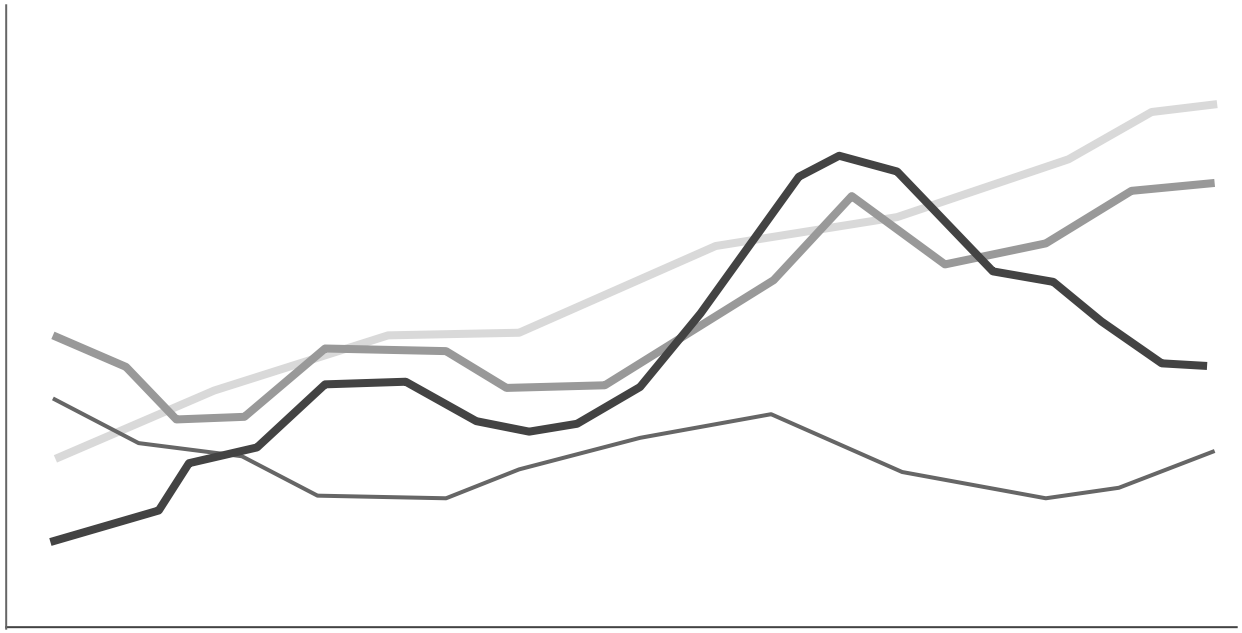
Better comparisons with panel of charts

do



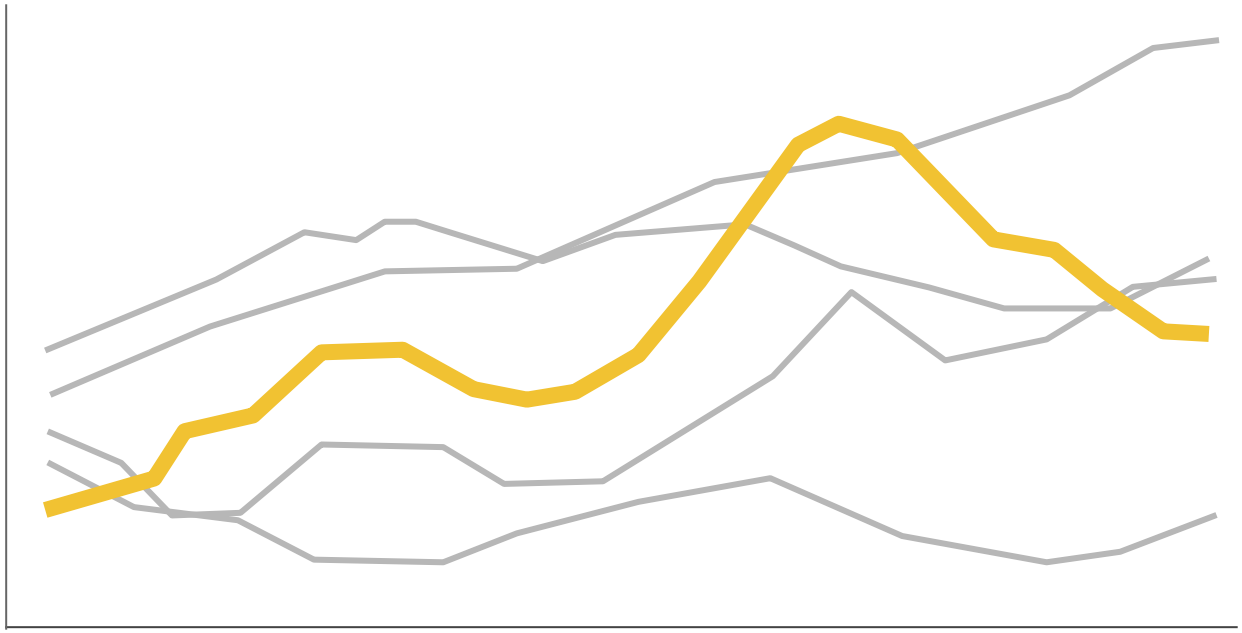
do

Four or fewer lines is better with different dark intensities



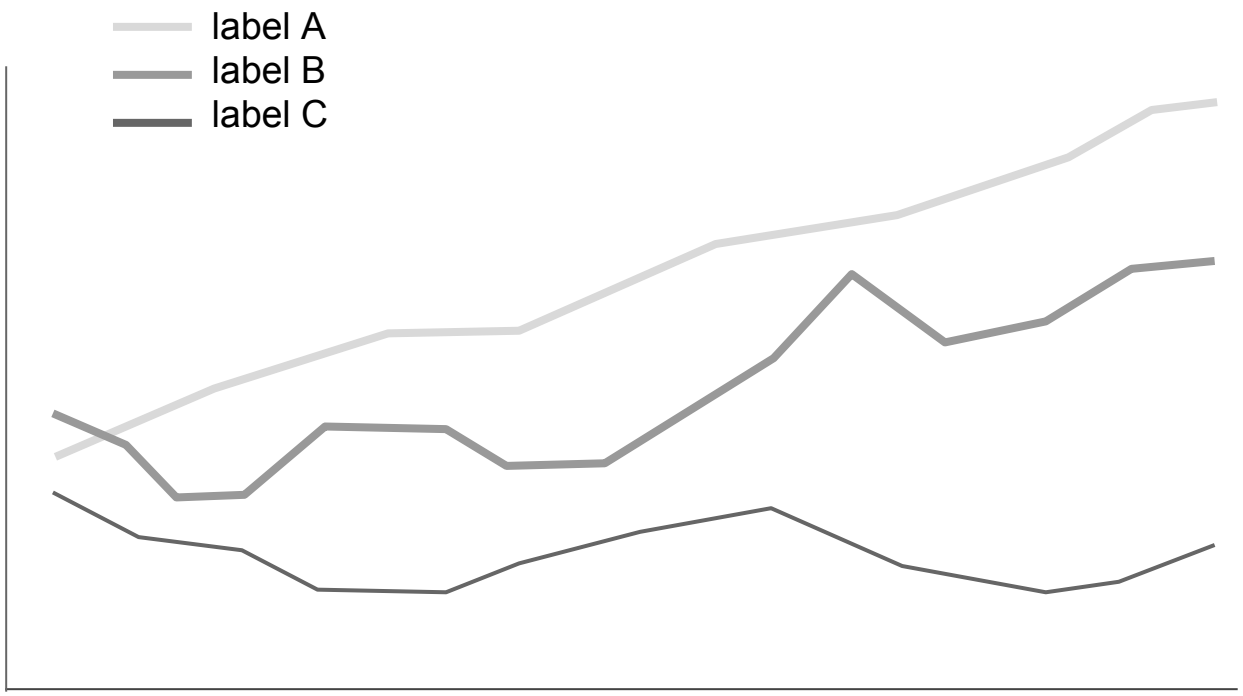
Use bright to dark colors to emphasize the important line

do



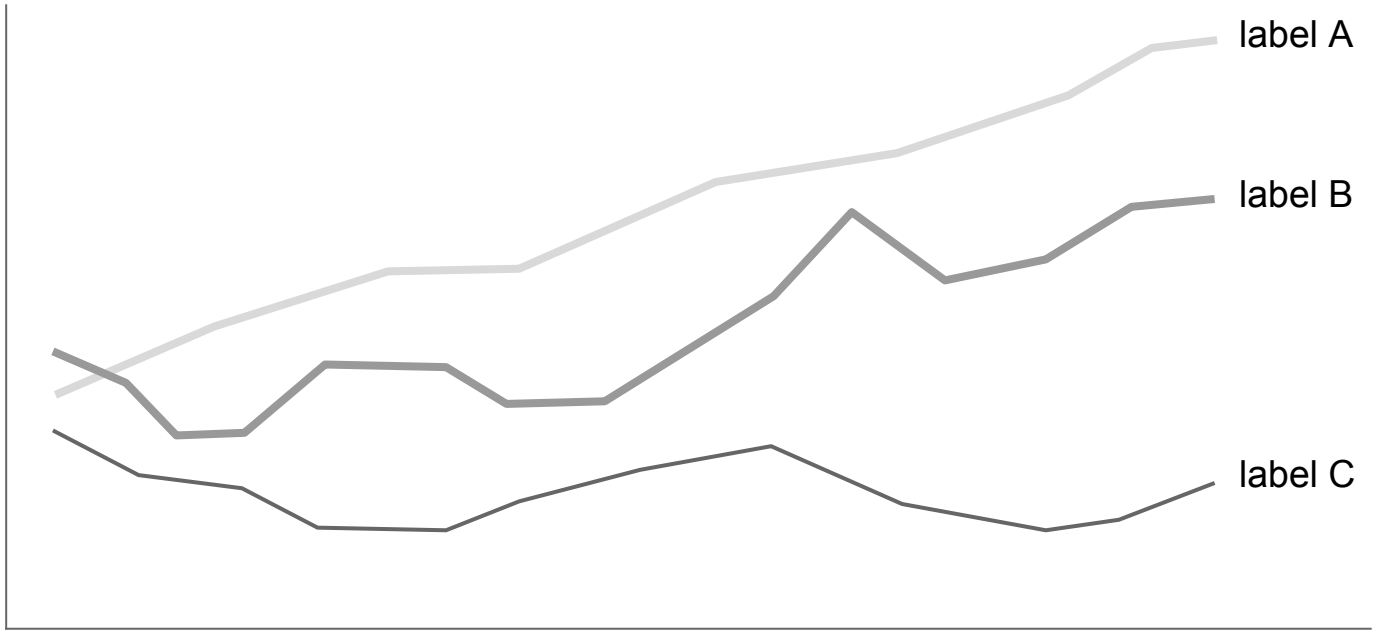
don’t

Avoid labeling at long distance



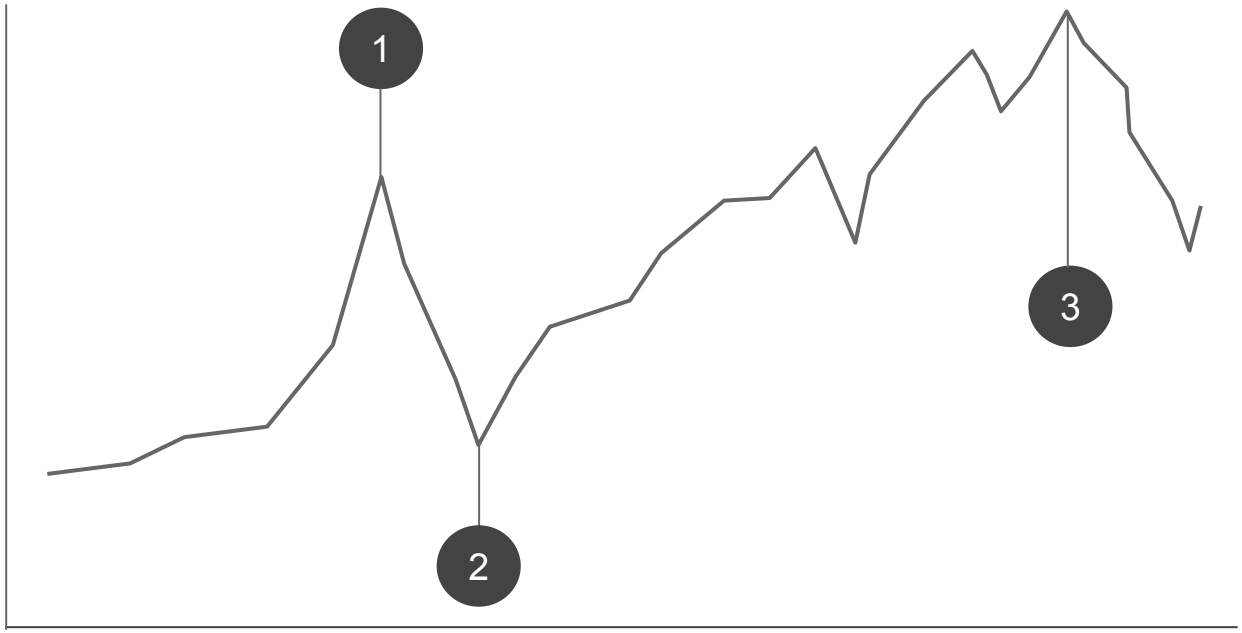
Label the lines directly

do



don’t

Avoid labeling at long distance



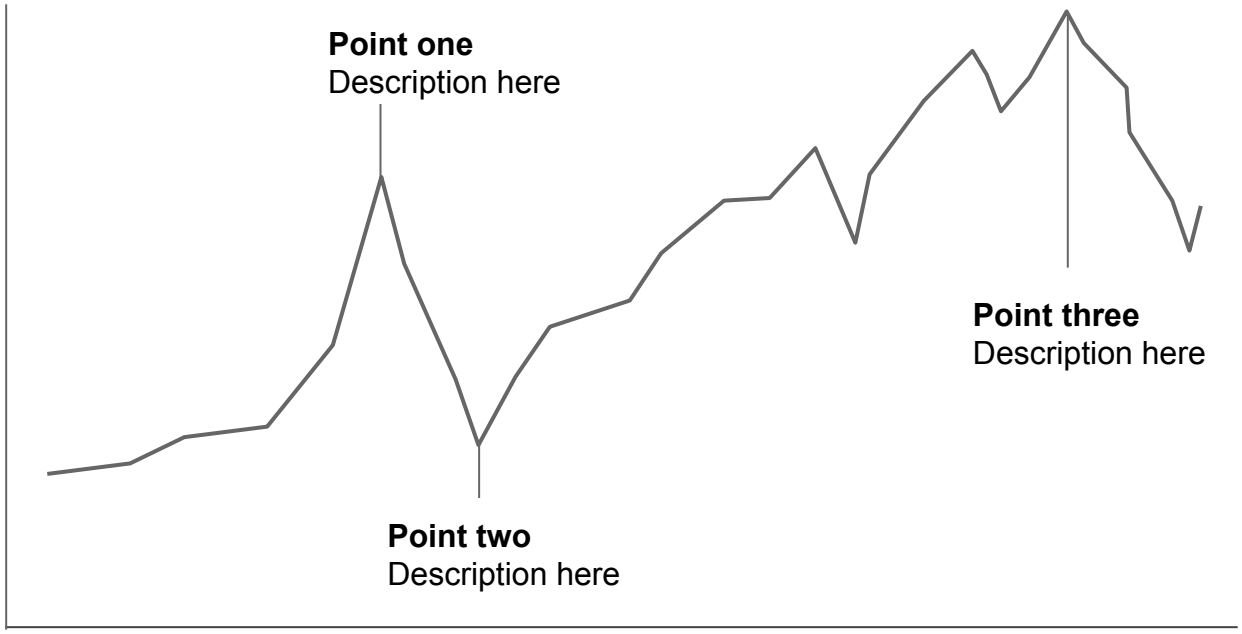
1 **Point one** Description text goes here

2 **Point two** Description text goes here

3 **Point three** Description text goes here

Prefer direct labeling

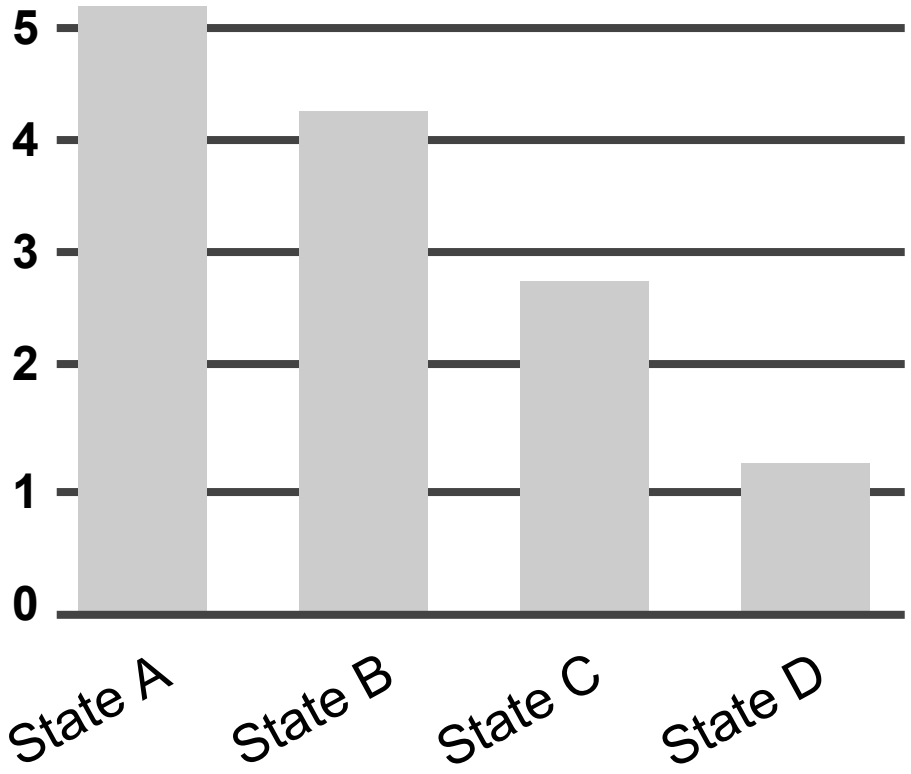
do



don’t

HEADLINE OF THE CHART

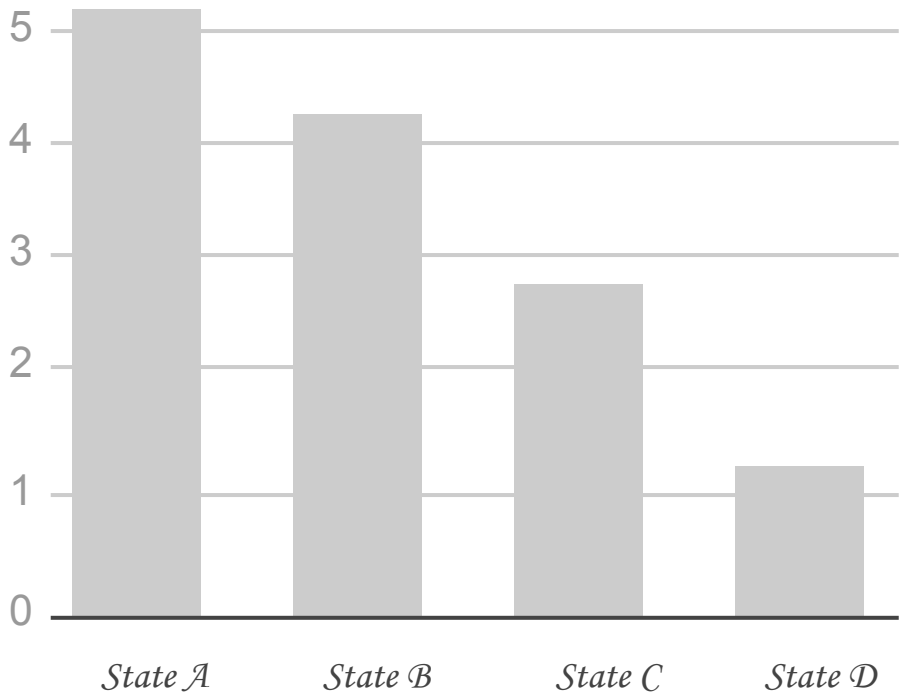
A brief description that outlines what the data shows



don't

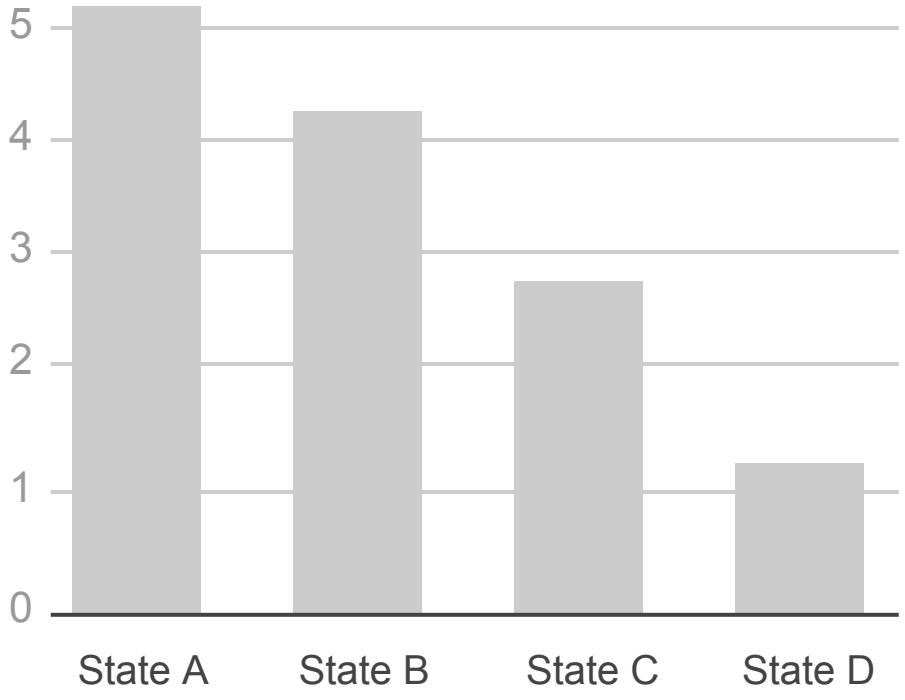
HEADLINE OF THE CHART

A brief description that outlines what the data shows



HEADLINE OF THE CHART

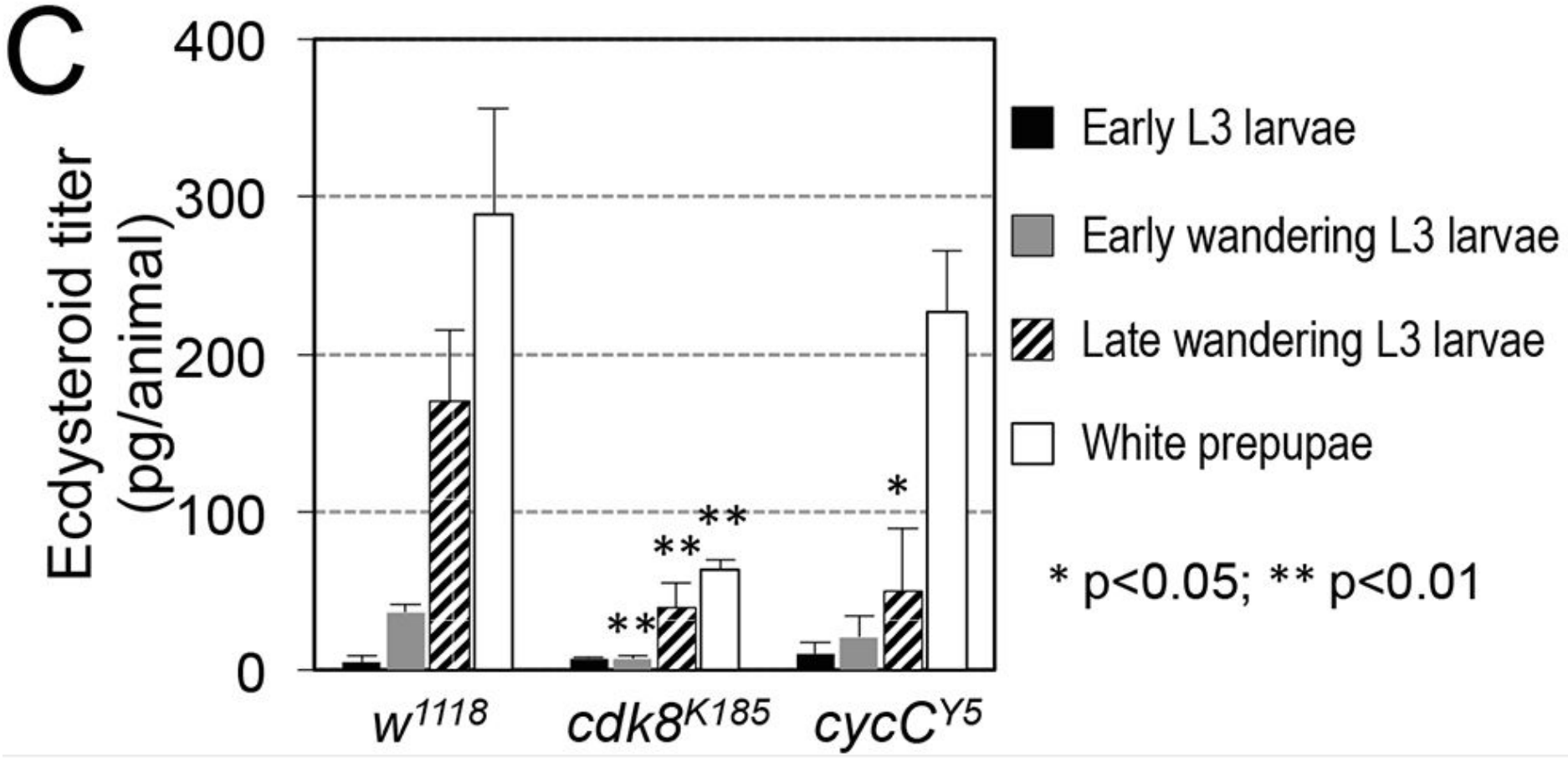
A brief description that outlines what the data shows

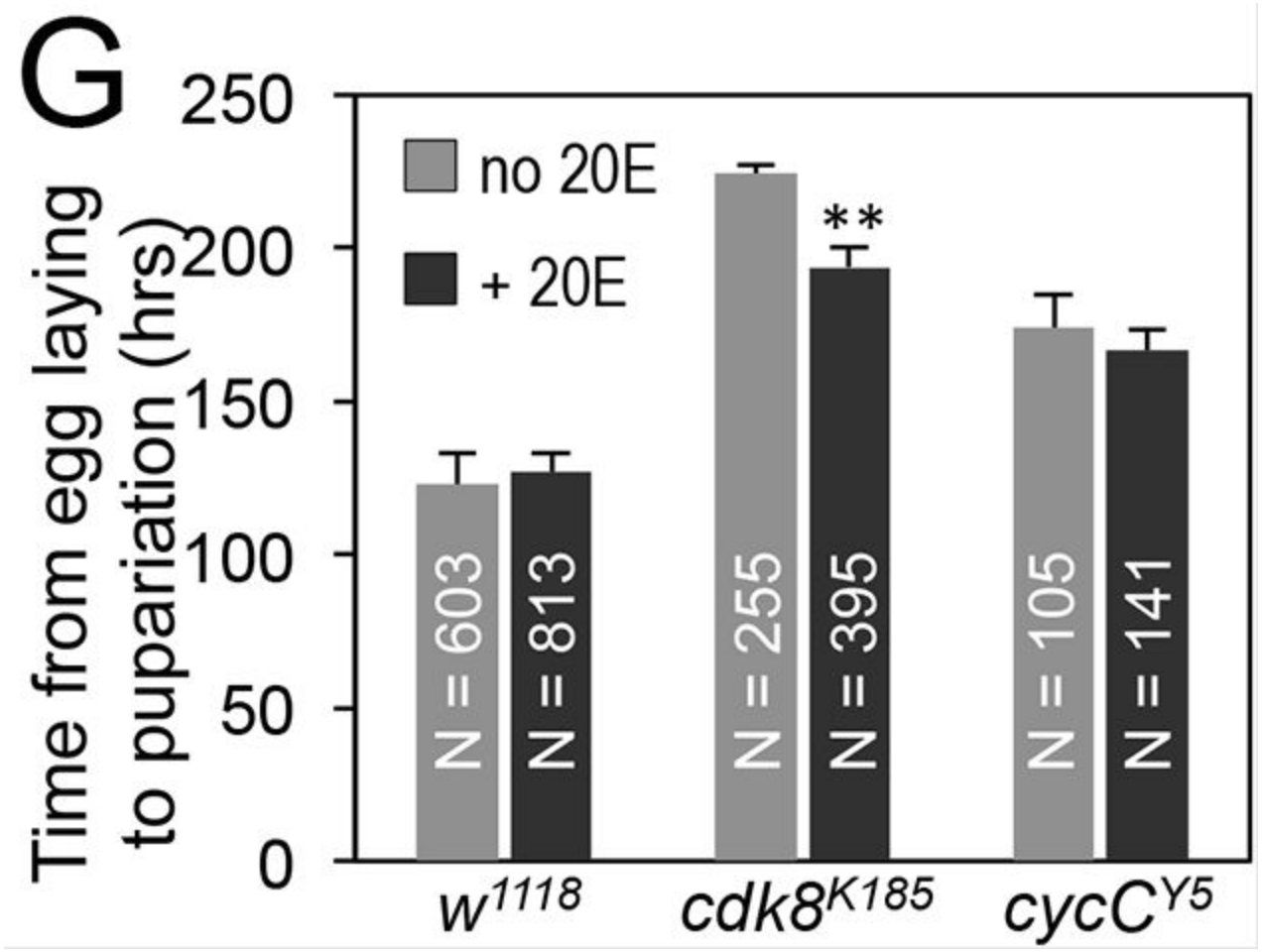


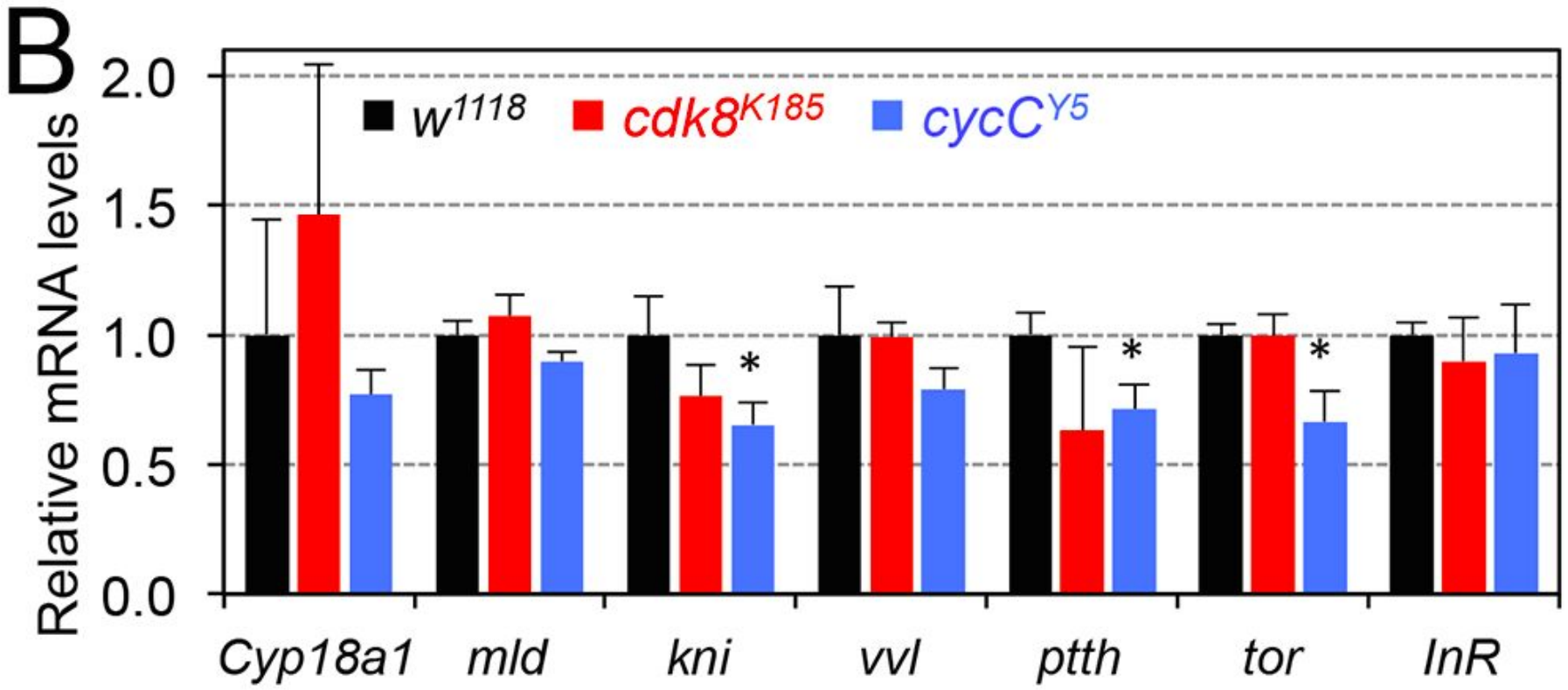
do

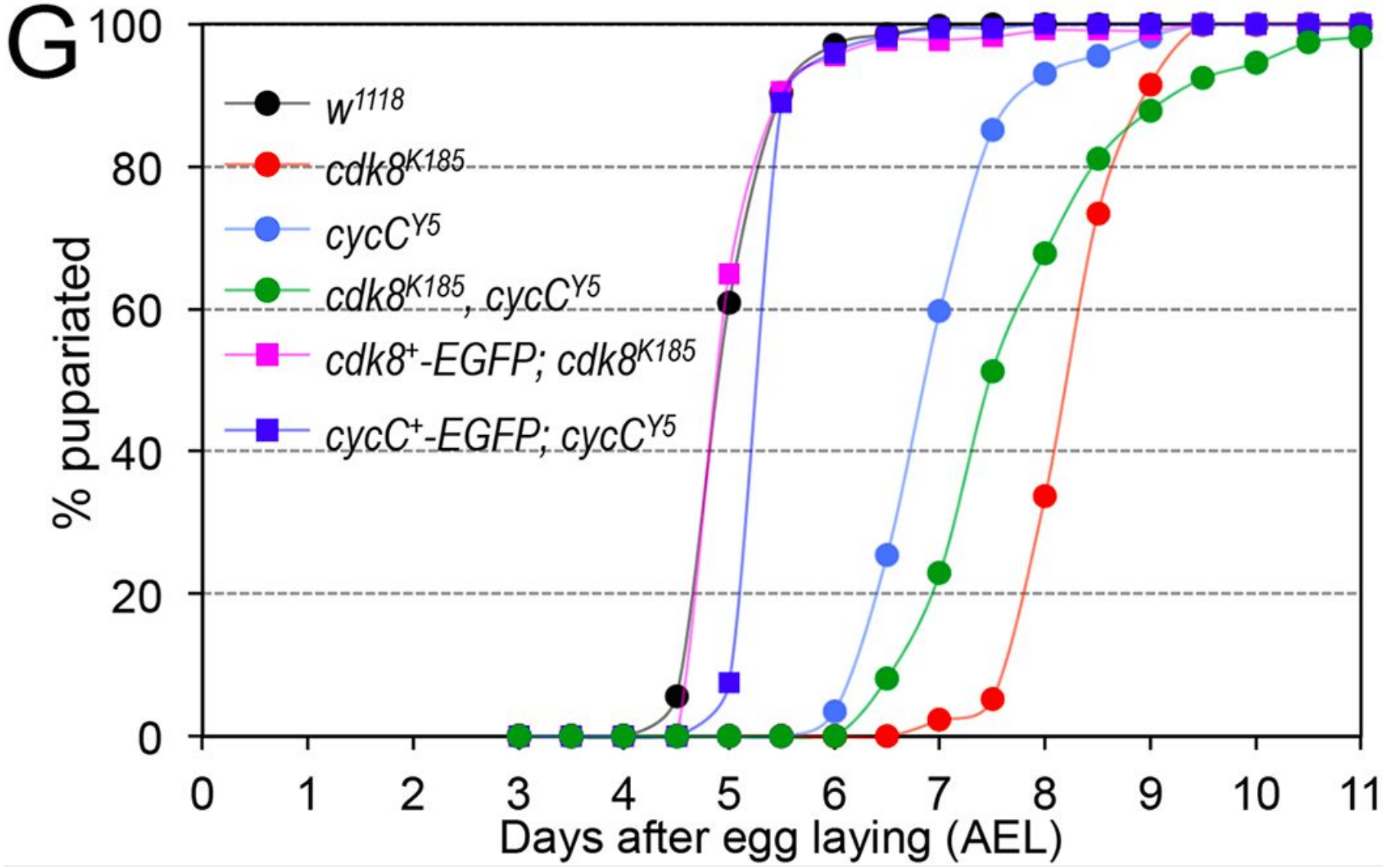
Some Examples

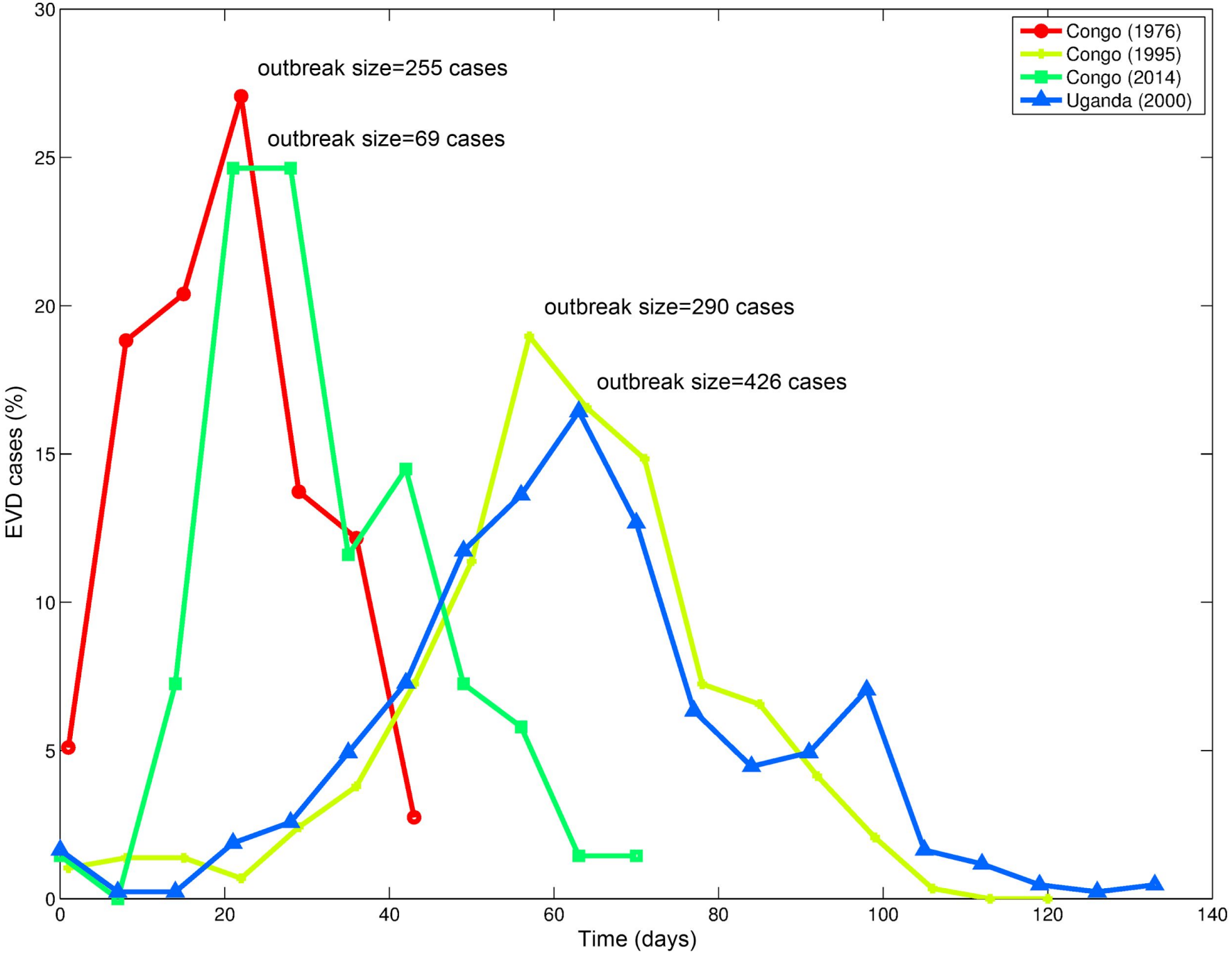
Scientific Journals

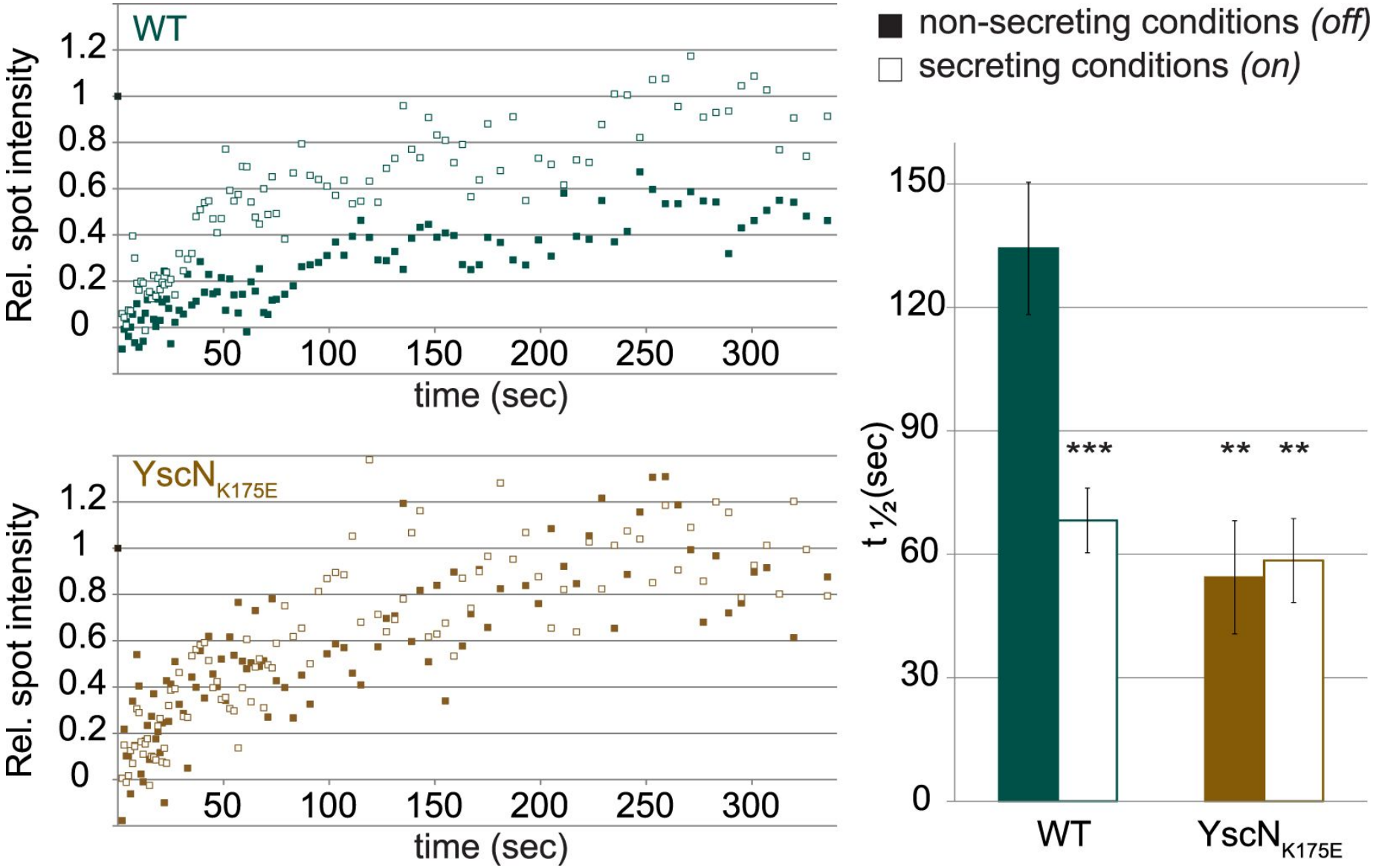

C
Ecdysteroid titer (pg/animal)
400
300
200
100
0
w1118
cdk8K185
cycCY5
Early L3 larvae
Early wandering L3 larvae
Late wandering L3 larvae
White prepupae
* p<0.05; ** p<0.01


G
Time from egg laying to pupariation (hrs)
250
200
150
100
50
0
no 20E
+ 20E
**
N = 603
N = 813
N = 255
N = 395
N = 105
N = 141
w1118
cdk8K185
cycCY5


B
Relative mRNA levels
2.0
1.5
1.0
0.5
0.0
w1118
cdk8K185
cycCY5
*
*
*
Cyp18a1
mld
kni
vvl
ptth
tor
InR


G
100
80
60
40
20
0
% pupariated
w1118
cdk8K185
cycCY5
cdk8K185, cycCY5
cdk8+-EGFP; cdk8K185
cycC+-EGFP; cycCY5
0
1
2
3
4
5
6
7
8
9
10
11
Days after egg laying (AEL)


Congo (1976)
Congo (1995)
Congo (2014)
Uganda (2000)
outbreak size=255 cases
outbreak size=69 cases
outbreak size=290 cases
outbreak size=426 cases
EVD cases (%)
Time (days)
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140


WT
Rel. spot intensity
1.2
1
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300
time (sec)
YscN K175E
Rel. spot intensity
time (sec)
non-secreting conditions (off)
secreting conditions (on)
t ½(sec)
150
120
90
60
30
0
***
**
**
WT
YscN K175E