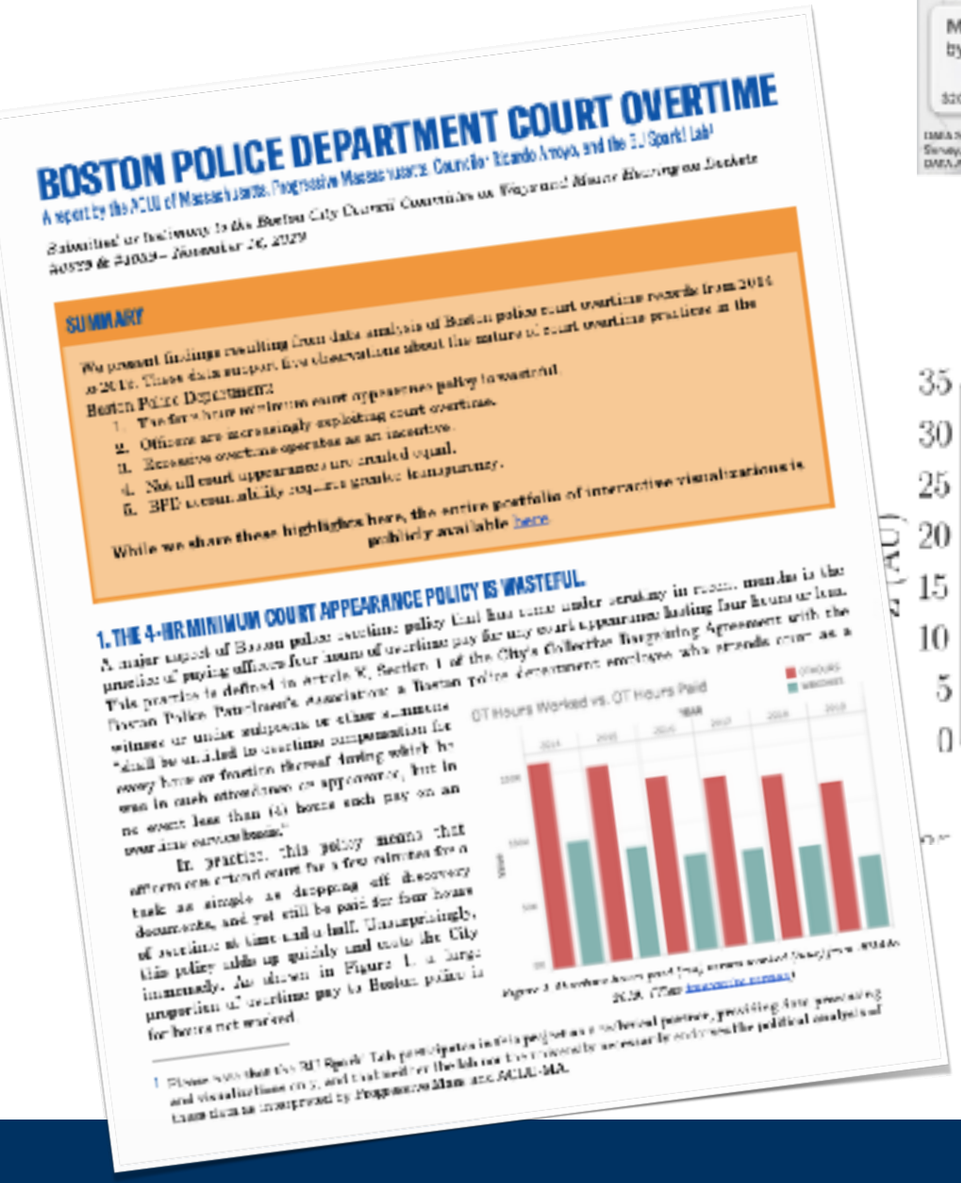
Effective Data Visualization

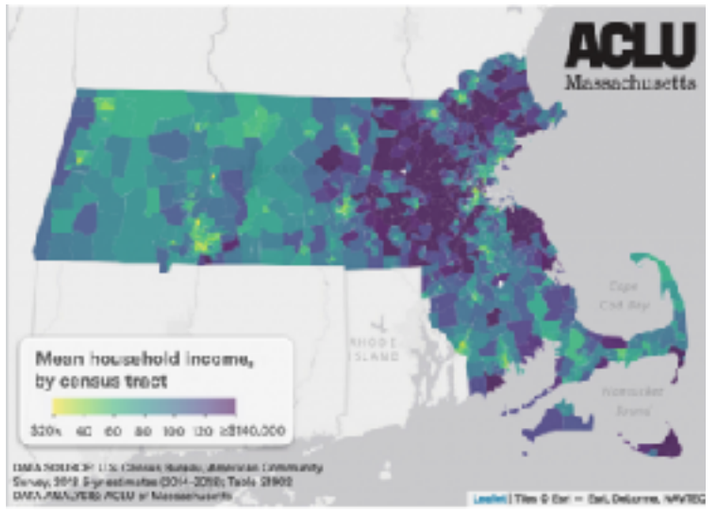
INFO 290 - Public Policy Data Lab

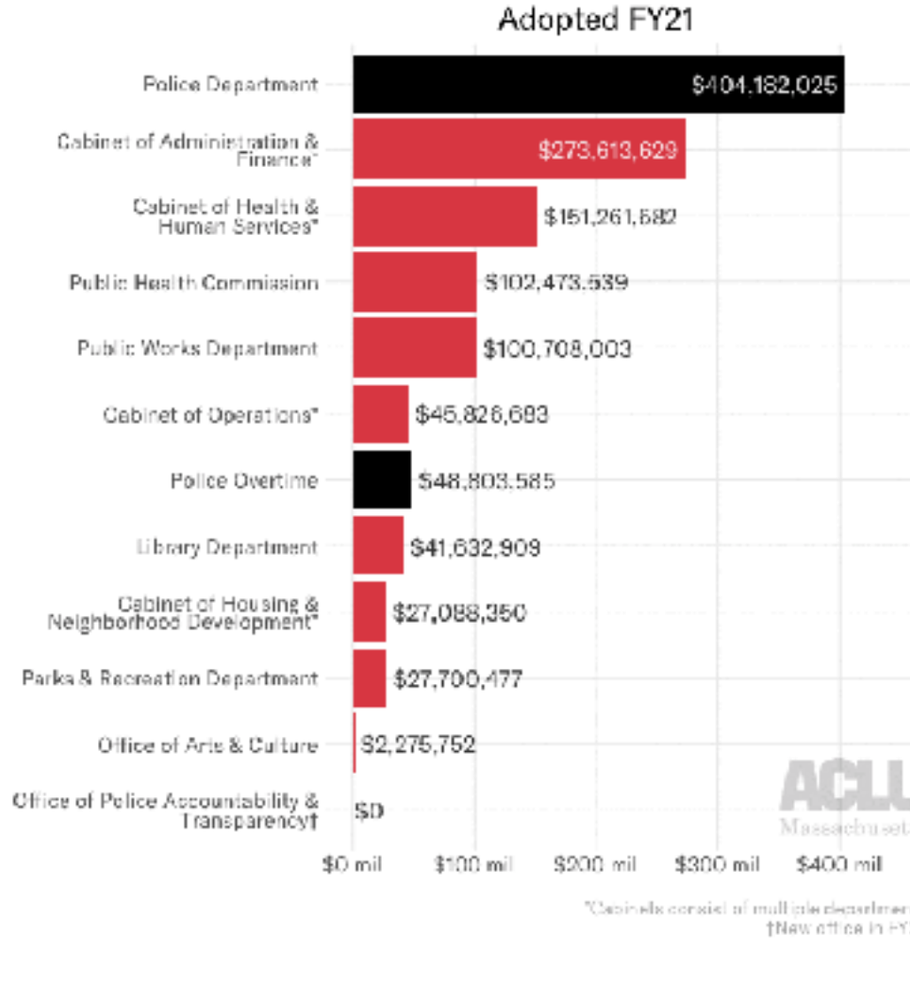
April 10, 2025

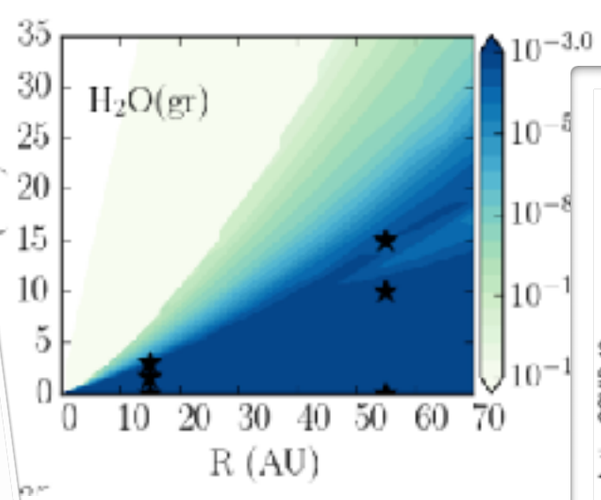
Berkeley SCHOOL OF INFORMATION

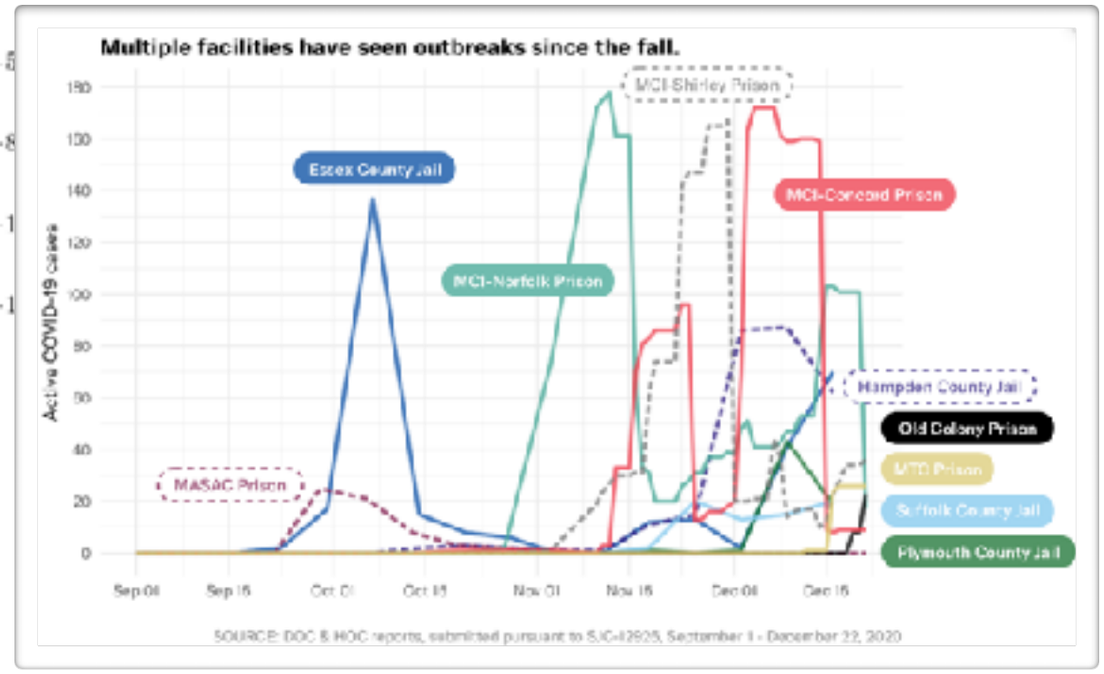
Data viz experience











Today we will explore...

- why data visualization is important
- how to identity the different components and types of data visualizations
- how to interpret visualizations critically
- how to intentionally tell a story by visualizing data

To learn more, consider taking INFO 247: Information Visualization & Presentation

All resources shared today are available at:

https://github.com/laurenmarietta/effective-data-viz-2025

This presentation is adapted from a workshop designed by Kira Tebbe in February 2020

Workshop outline

- What is data visualization 🗣️
- Anatomy of a data visualization
- Types of data visualization
- Roast! This! Viz! 🗣️
- ✨ 10 Best Practices ✨
- Improve! This! Viz!
- Exercises 🗣️
- Interactivity, inspiration, and learning more

What does the term "data visualization" mean to you? Why is it important?

Why is data visualization important?

data = information

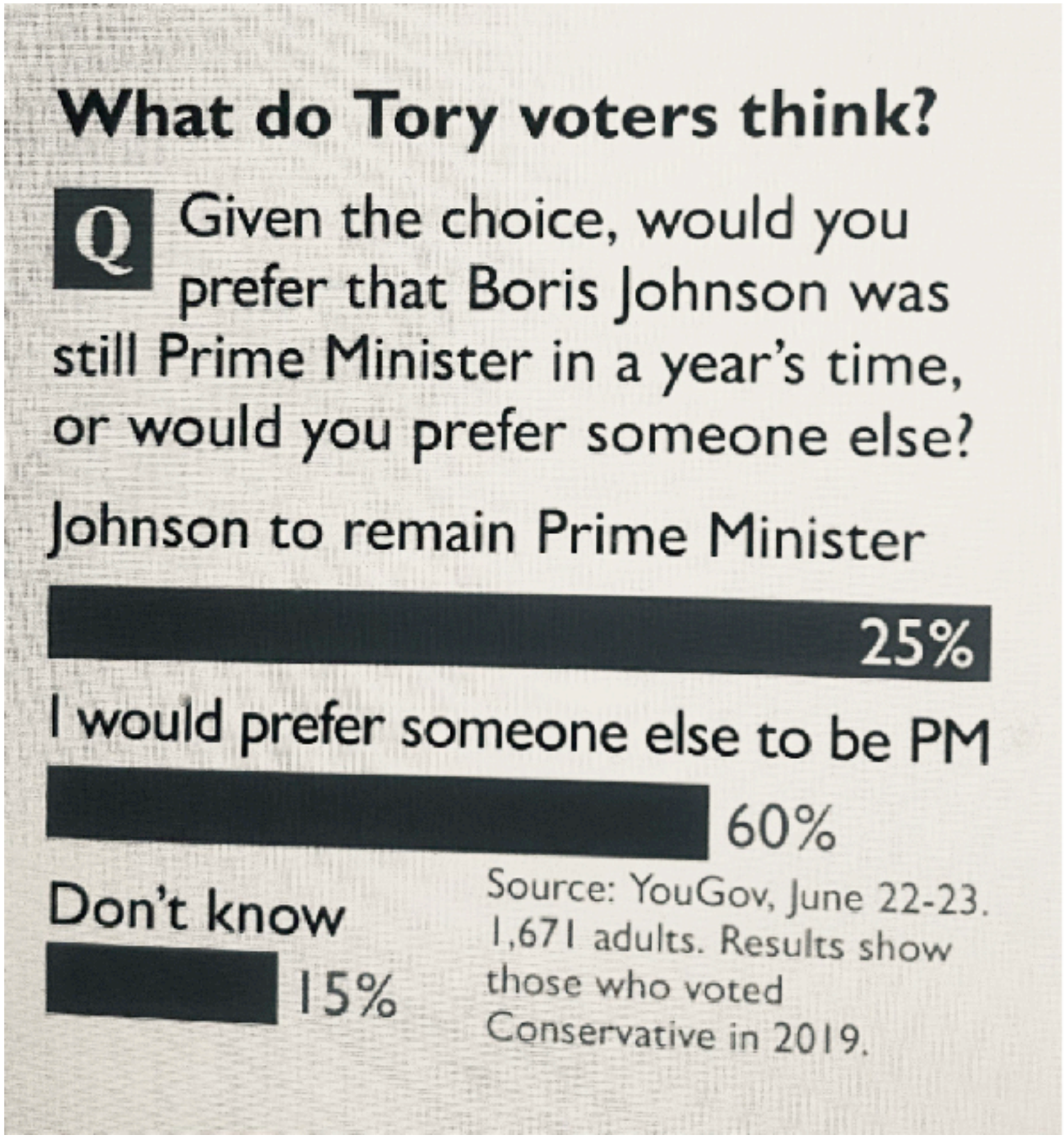
data visualization = communication

good data visualization = effective communication

bad data visualization = miscommunication

What do mediocre visualizations lead to?

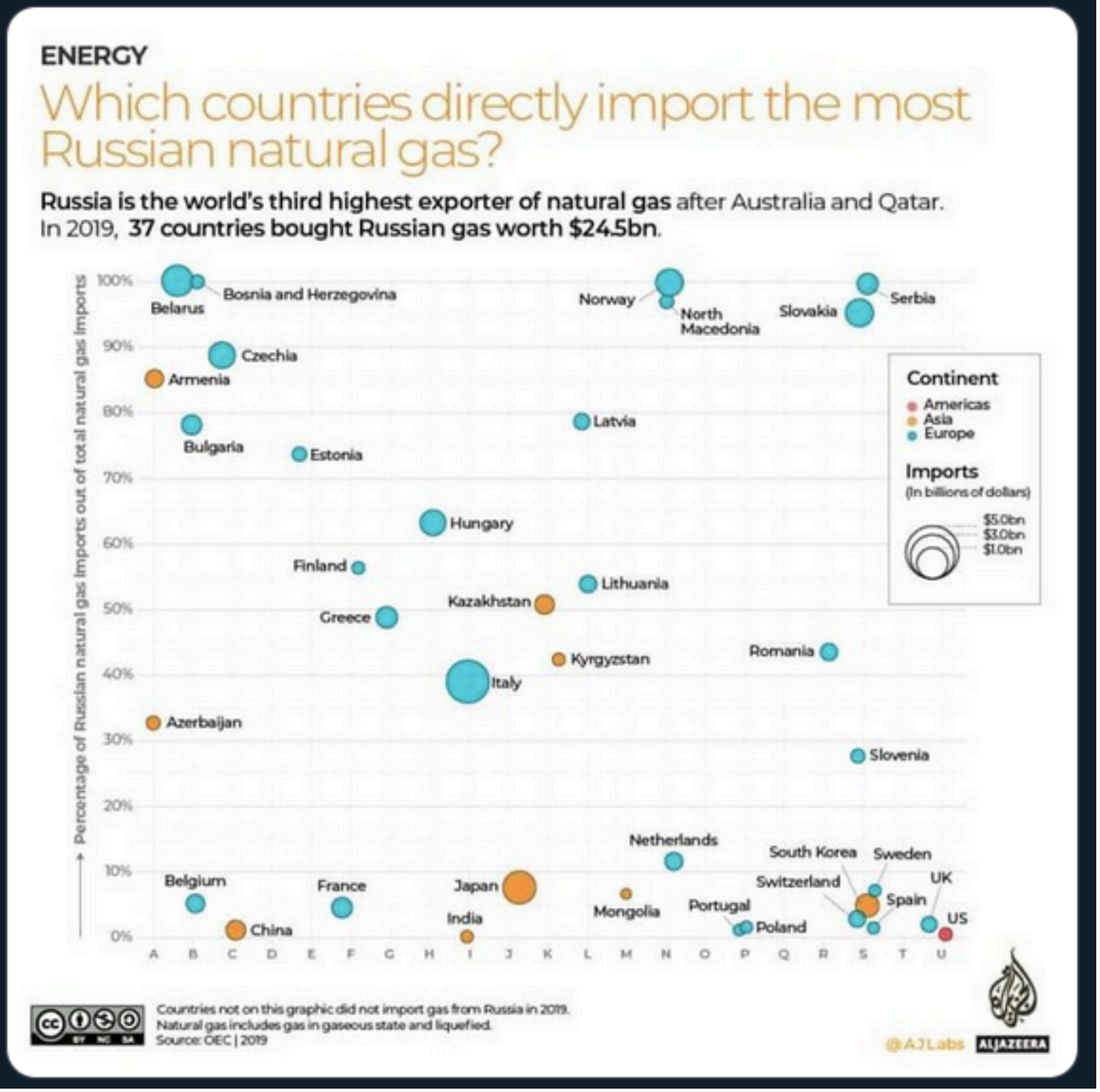
- confusion and frustration
- loss of attention
- poor impression of work quality
- misinterpretation of findings



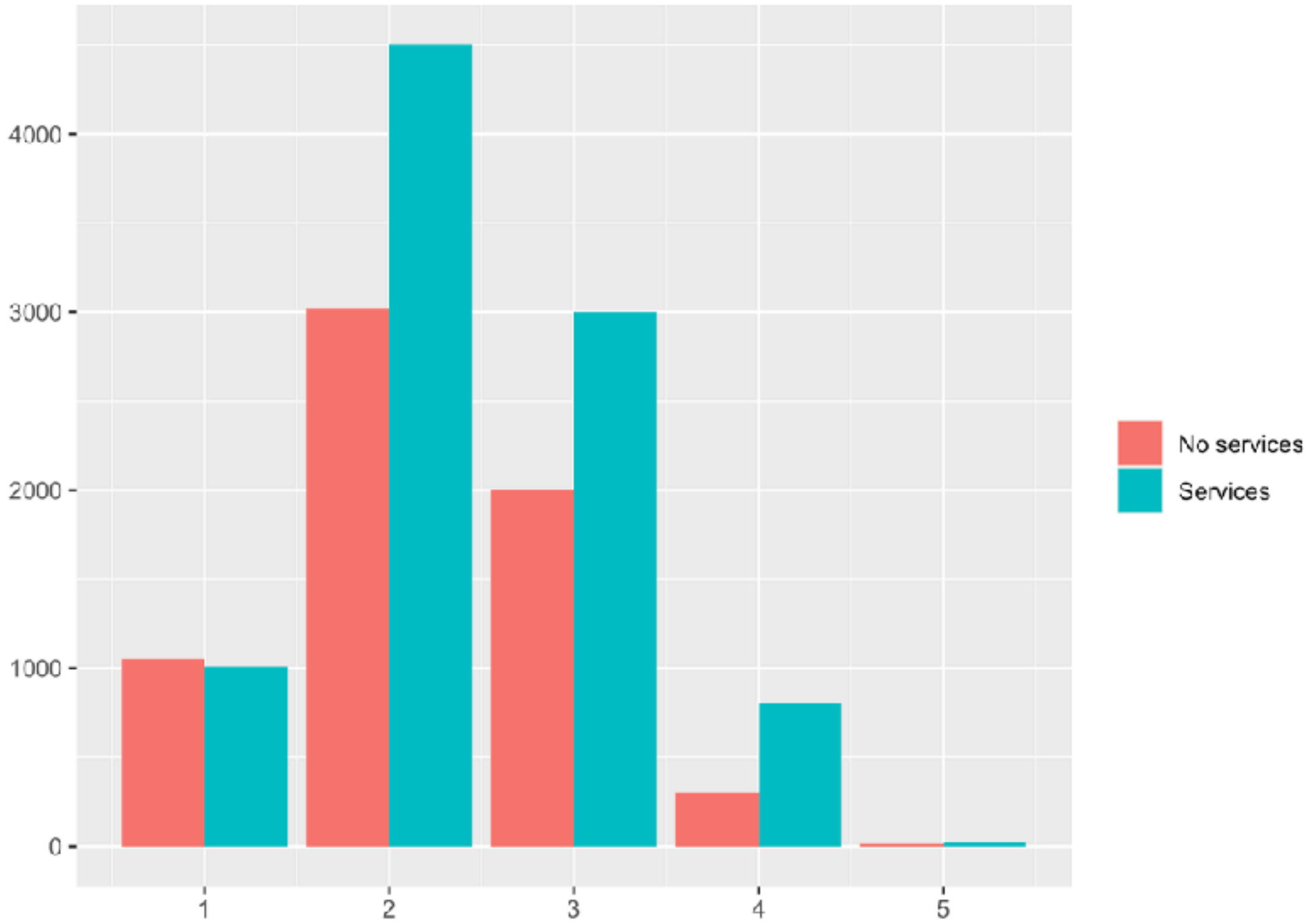
Source: Will Bailey-Watson on Twitter

Ben Stanley
@BDStanley

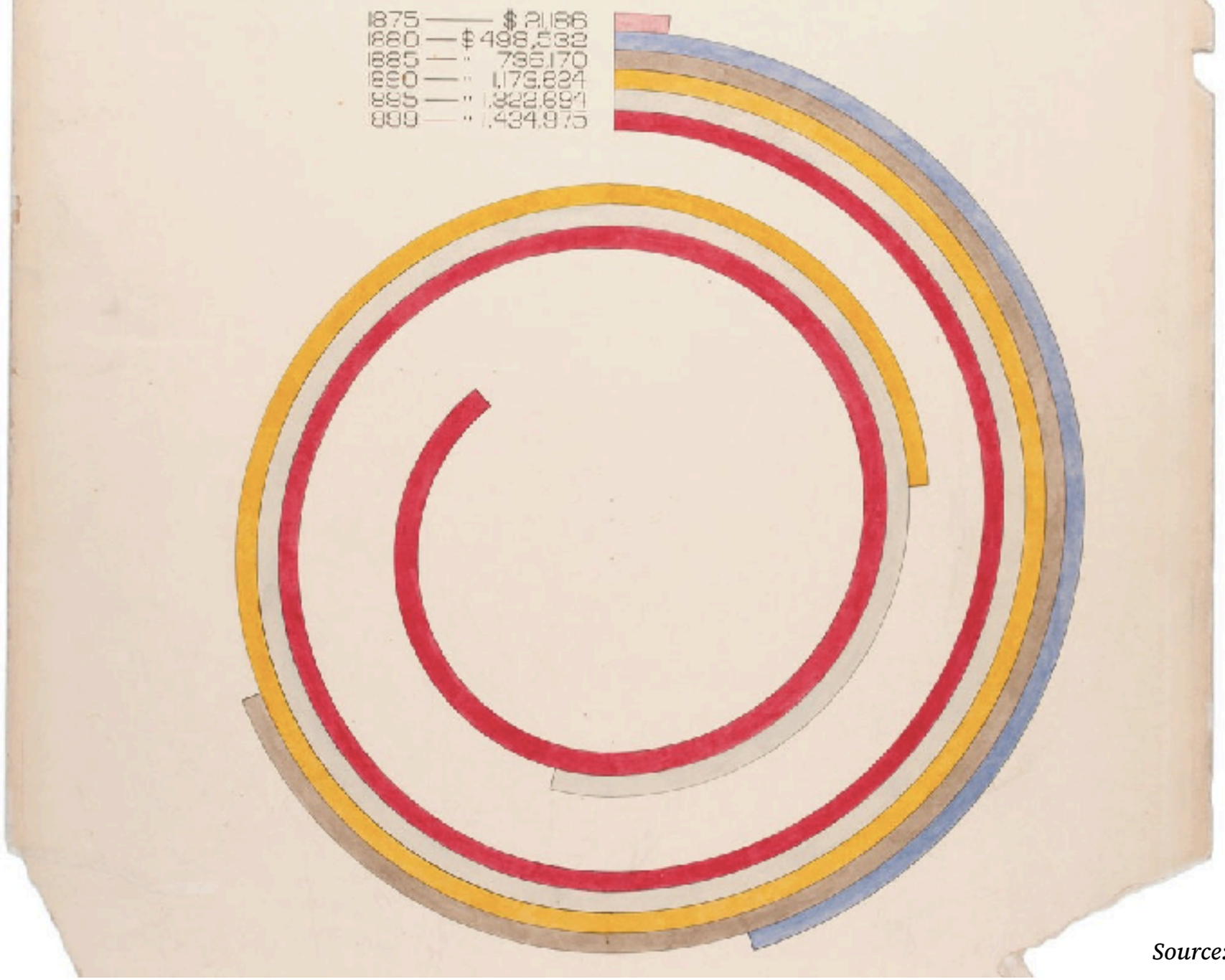
Incredible x-axis here.



Source: Ben Stanley on Twitter


4000
3000
2000
1000
0
1
2
3
4
5
No services
Services

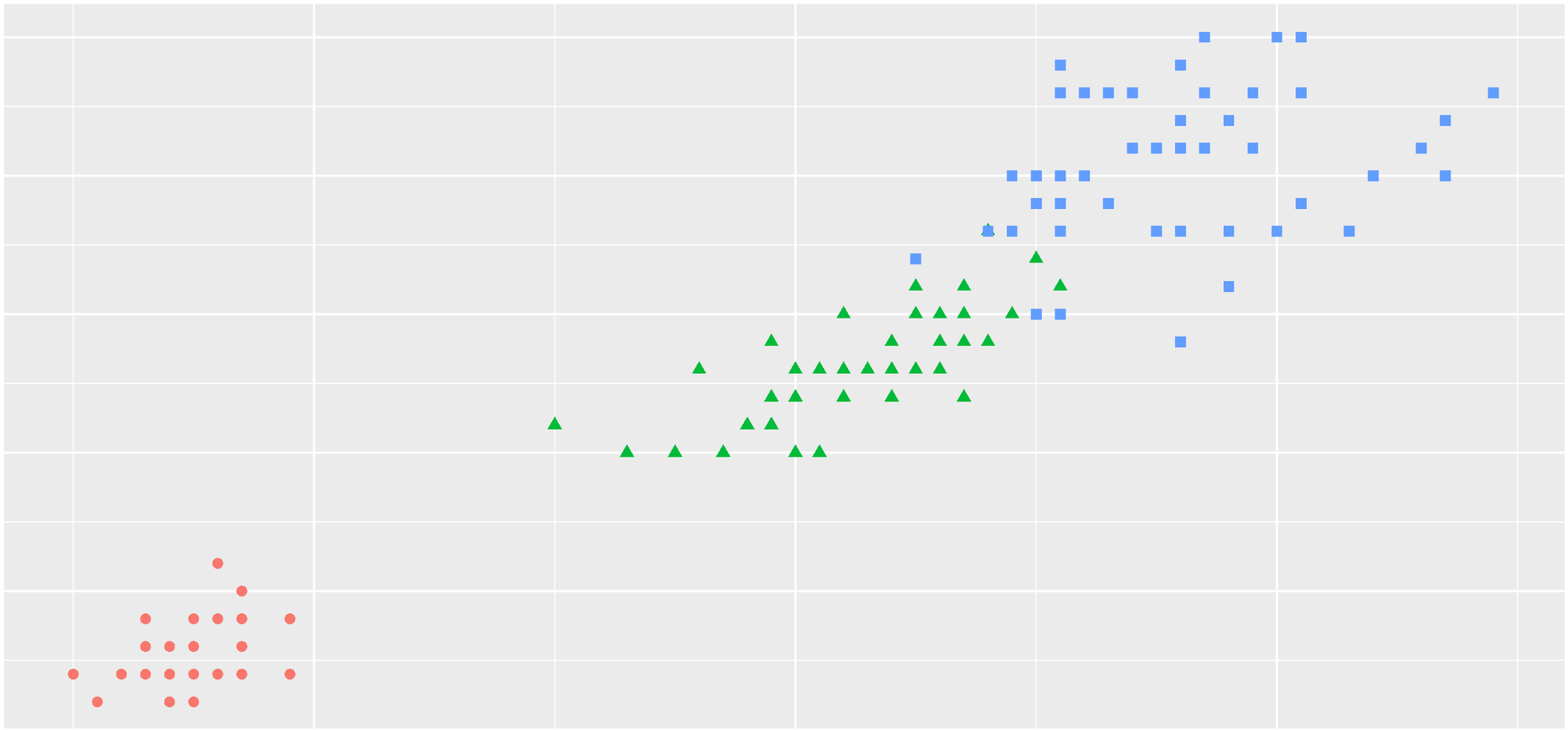
ASSESSED VALUE OF HOUSEHOLD AND KITCHEN FURNITURE OWNED BY GEORGIA NEGROES.



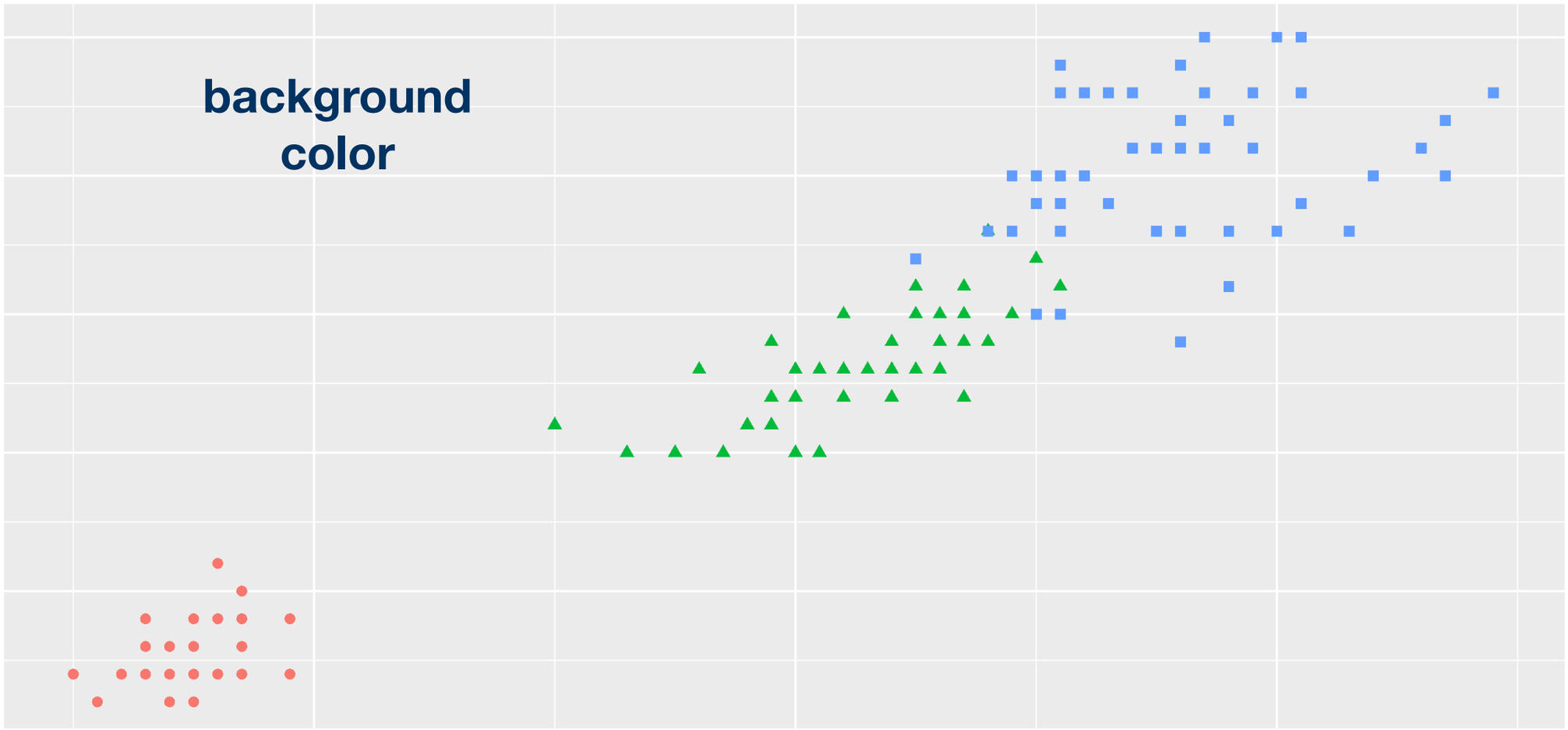
Source: W.E.B. Du Bois, ca. 1900

Anatomy of a data visualization

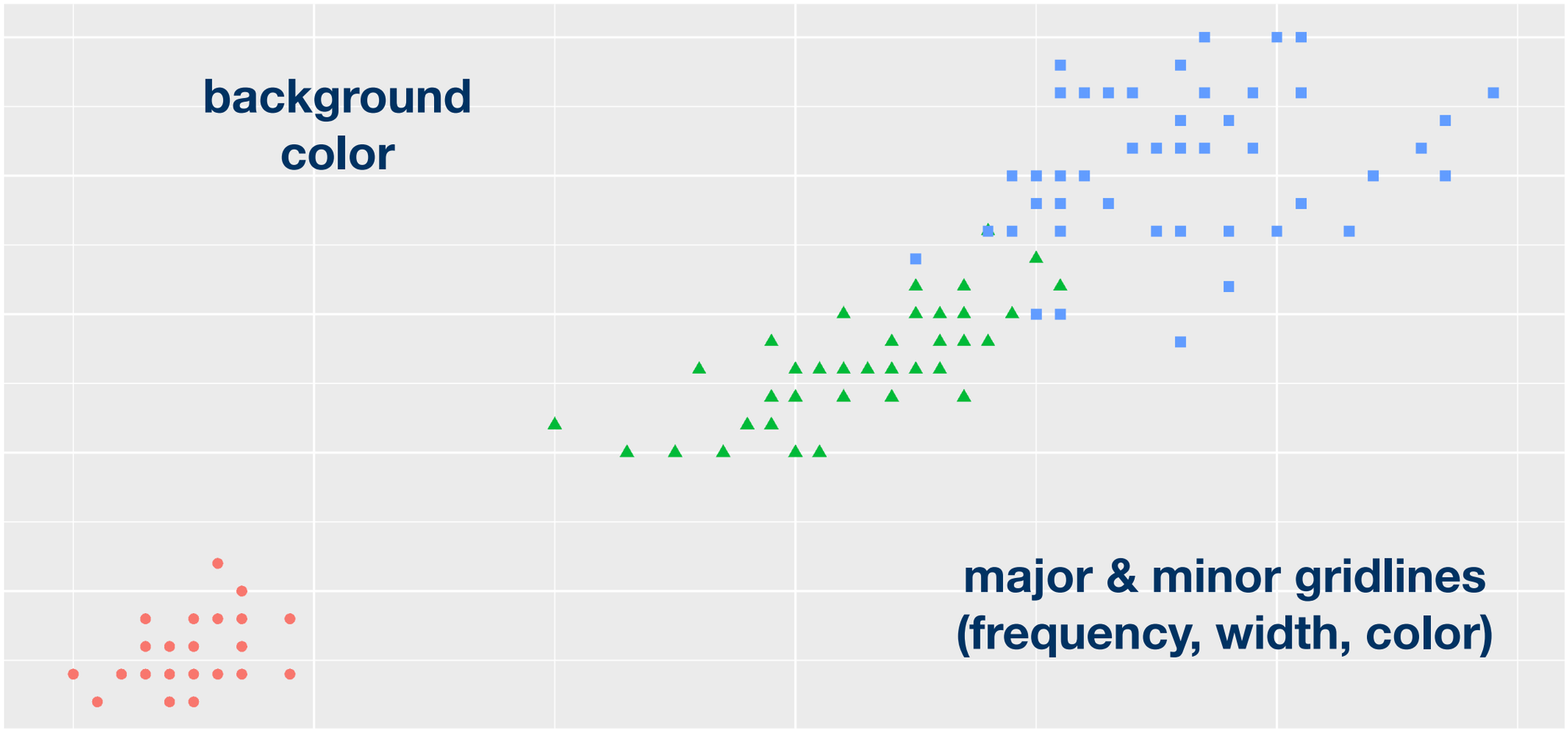
plot type & aesthetics
(colors, sizes, shapes, opacity)



plot type & aesthetics
(colors, sizes, shapes, opacity)

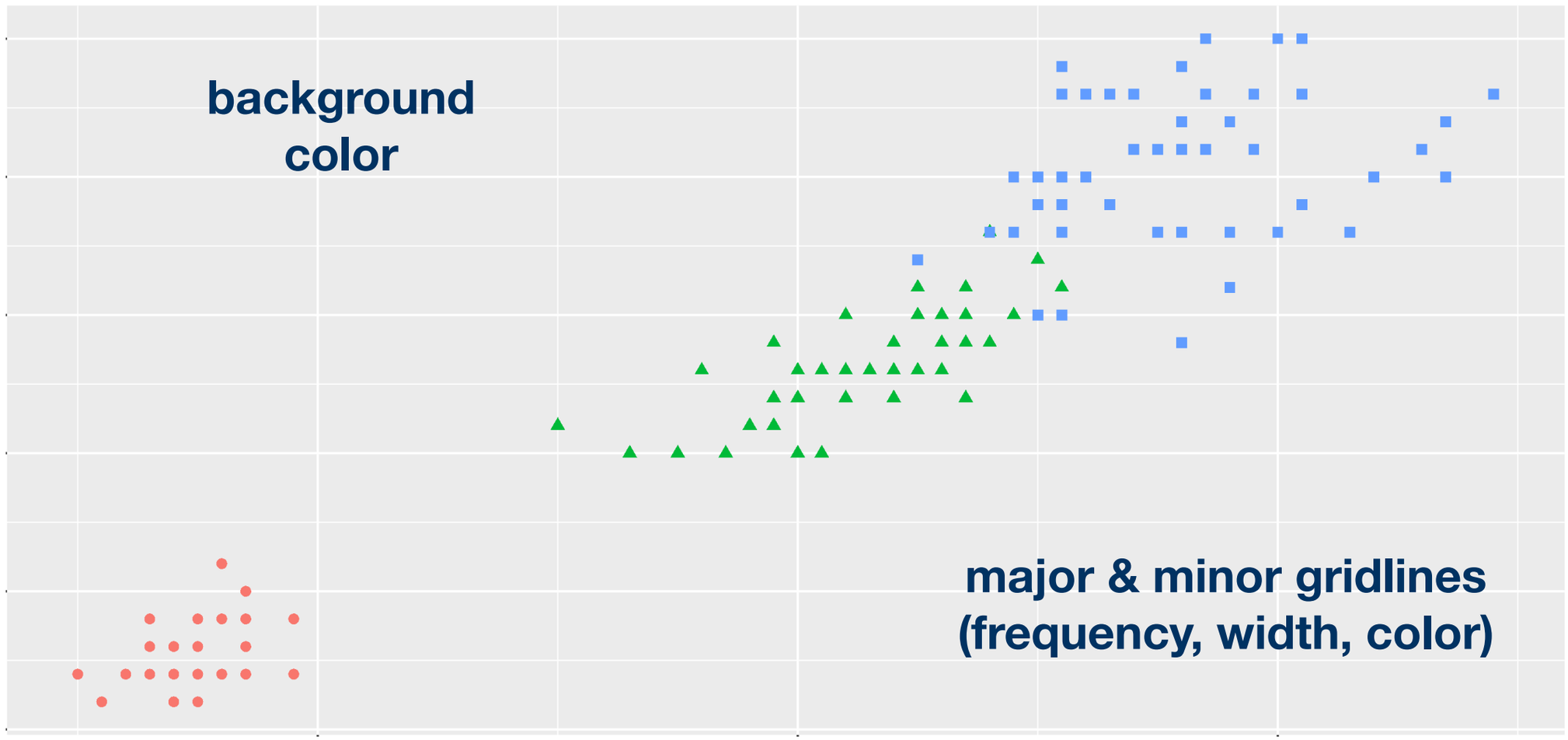


plot type & aesthetics
(colors, sizes, shapes, opacity)



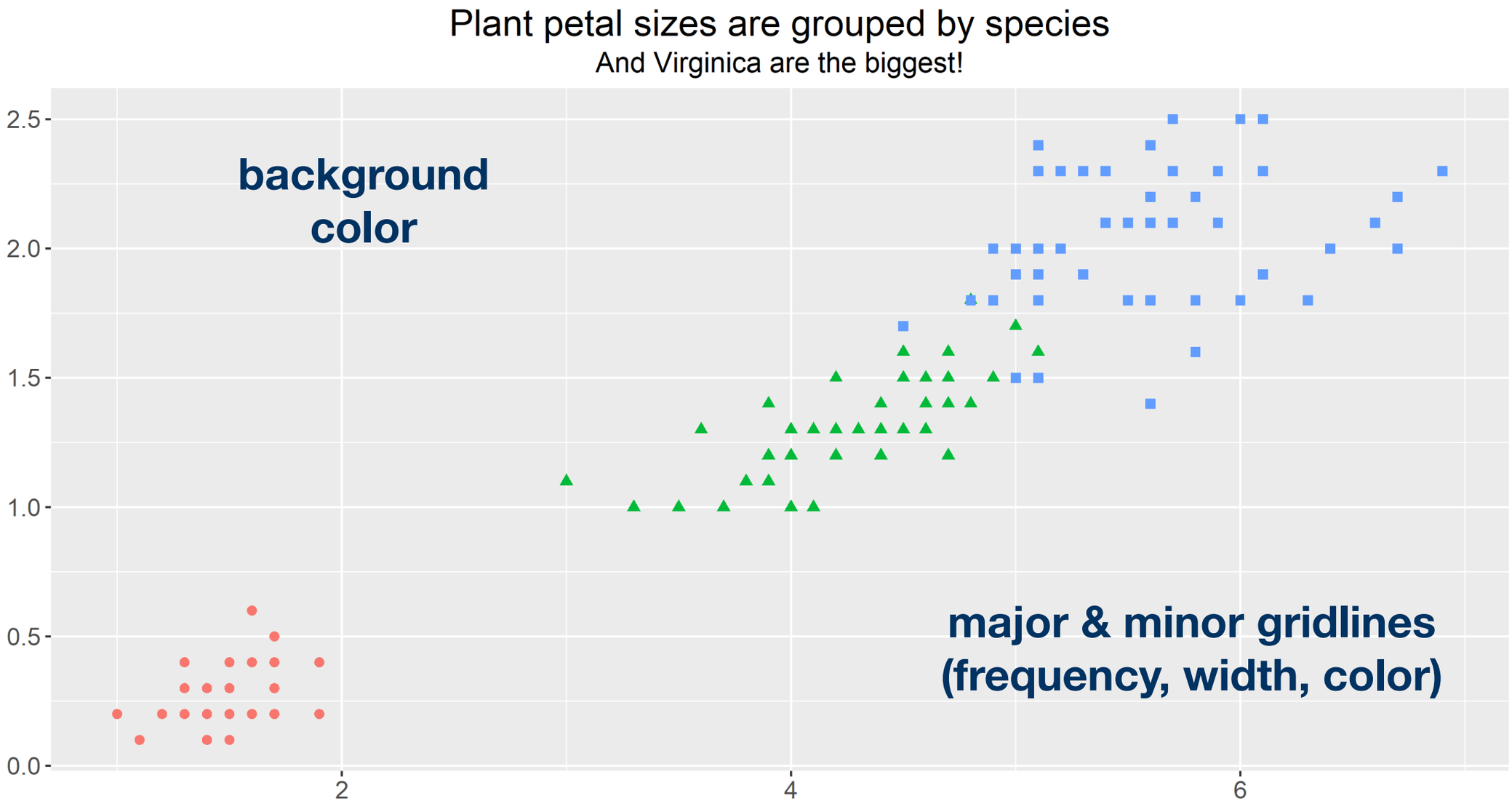
plot type & aesthetics (colors, sizes, shapes, opacity)

title & subtitle (size, font, color, location)



plot type & aesthetics
(colors, sizes, shapes, opacity)

title & subtitle
(size, font, color, location)

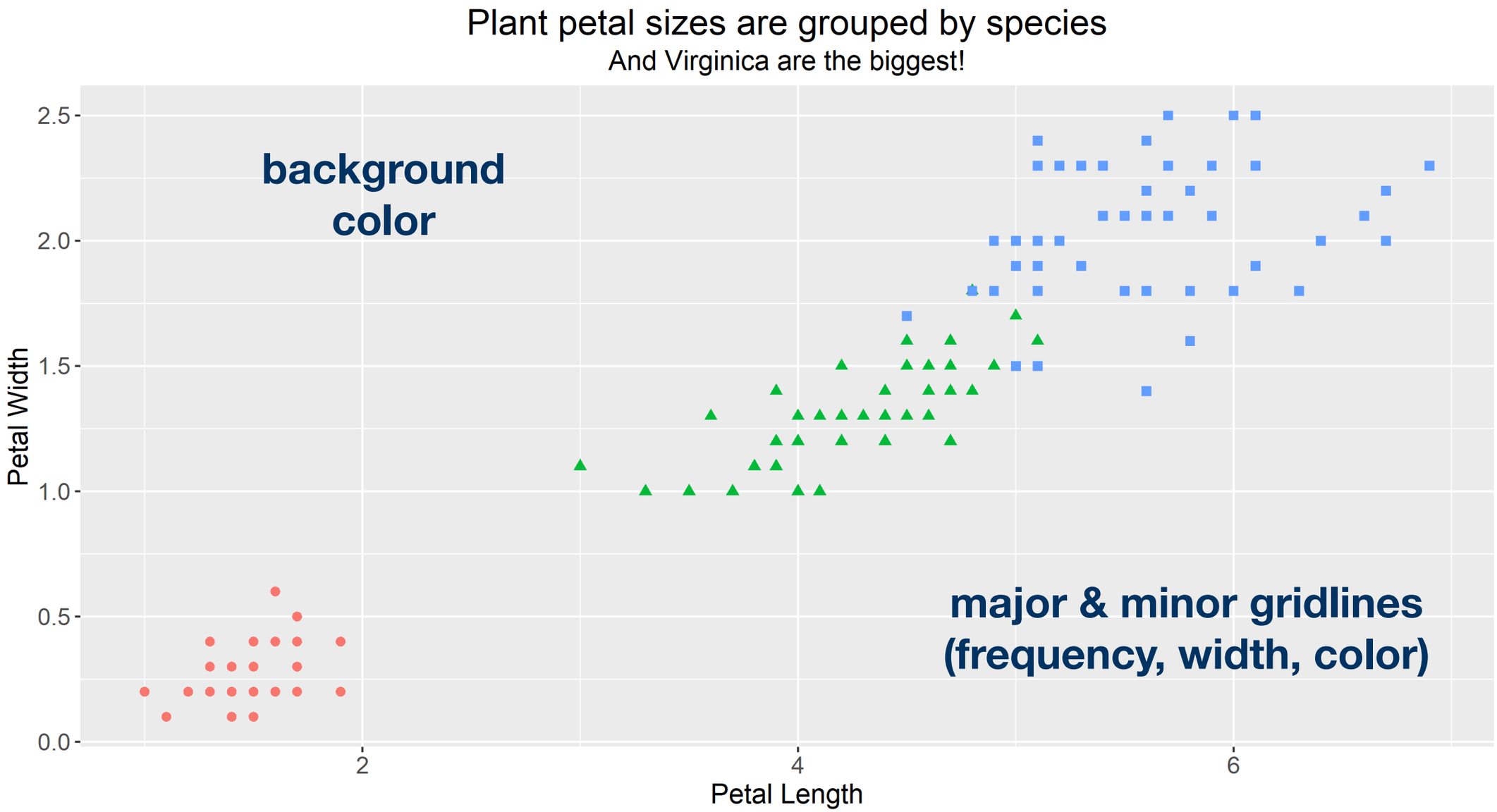


major & minor gridlines
(frequency, width, color)

X & Y tick marks
(frequency, size, font, color, angle)

plot type & aesthetics (colors, sizes, shapes, opacity)

title & subtitle (size, font, color, location)

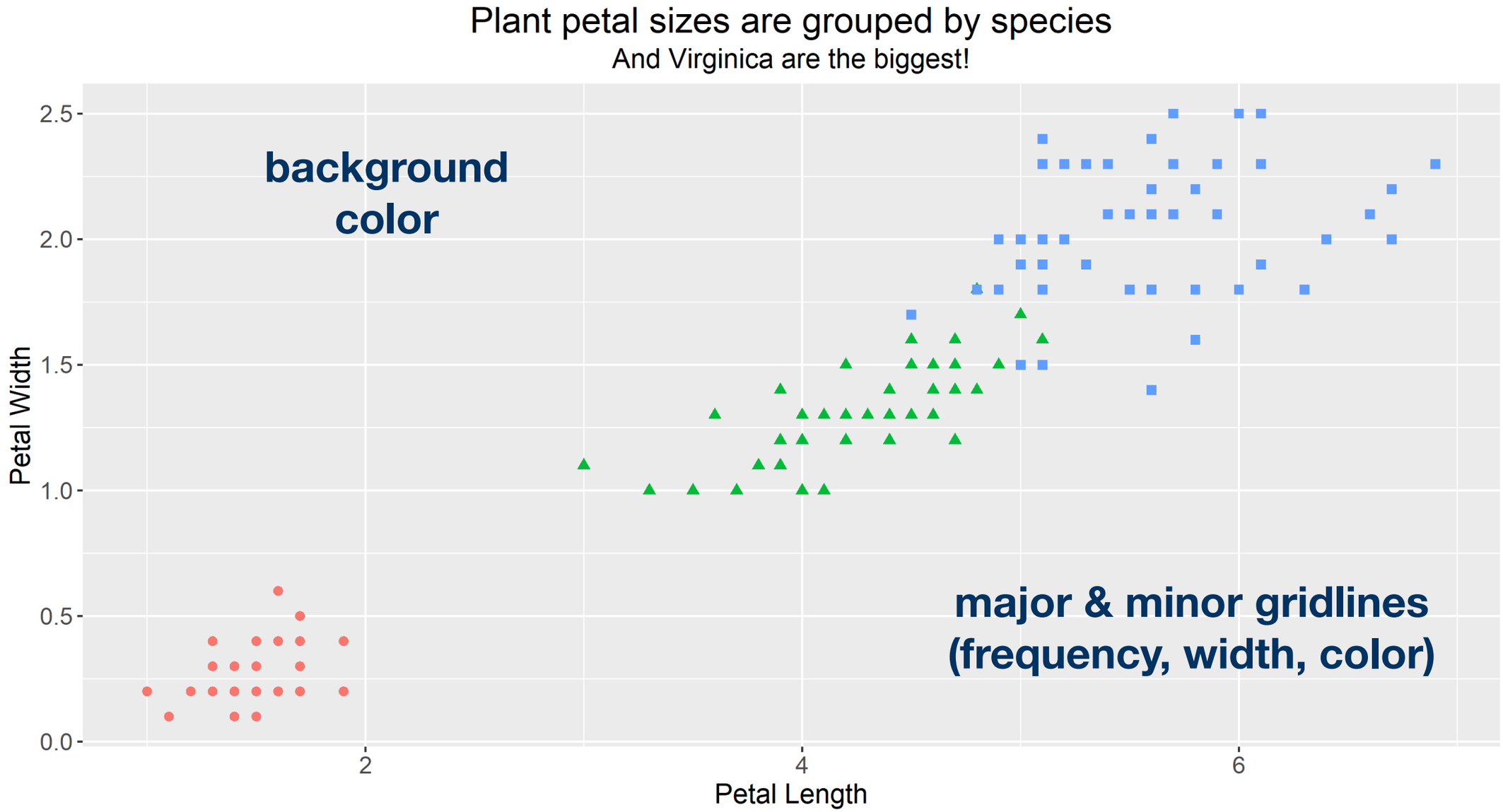


X & Y axis labels (size, font, color, angle)

X & Y tick marks (frequency, size, font, color, angle)

title & subtitle
(size, font, color, location)

plot type & aesthetics
(colors, sizes, shapes, opacity)



legend & legend title (location, size, font, color, existence)

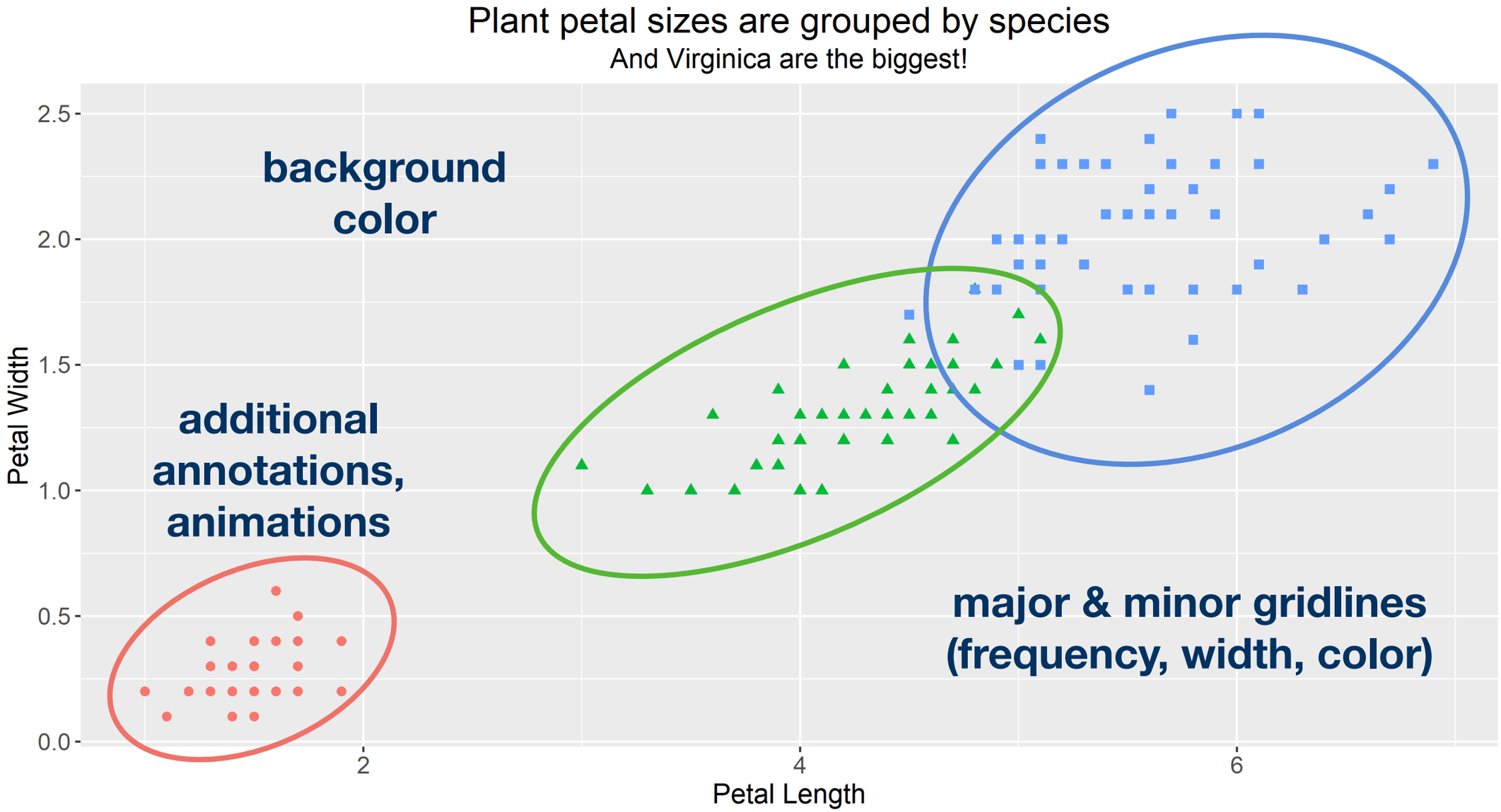
major & minor gridlines
(frequency, width, color)

X & Y axis labels
(size, font, color, angle)

X & Y tick marks
(frequency, size, font, color, angle)

title & subtitle
(size, font, color, location)

plot type & aesthetics
(colors, sizes, shapes, opacity)



background color

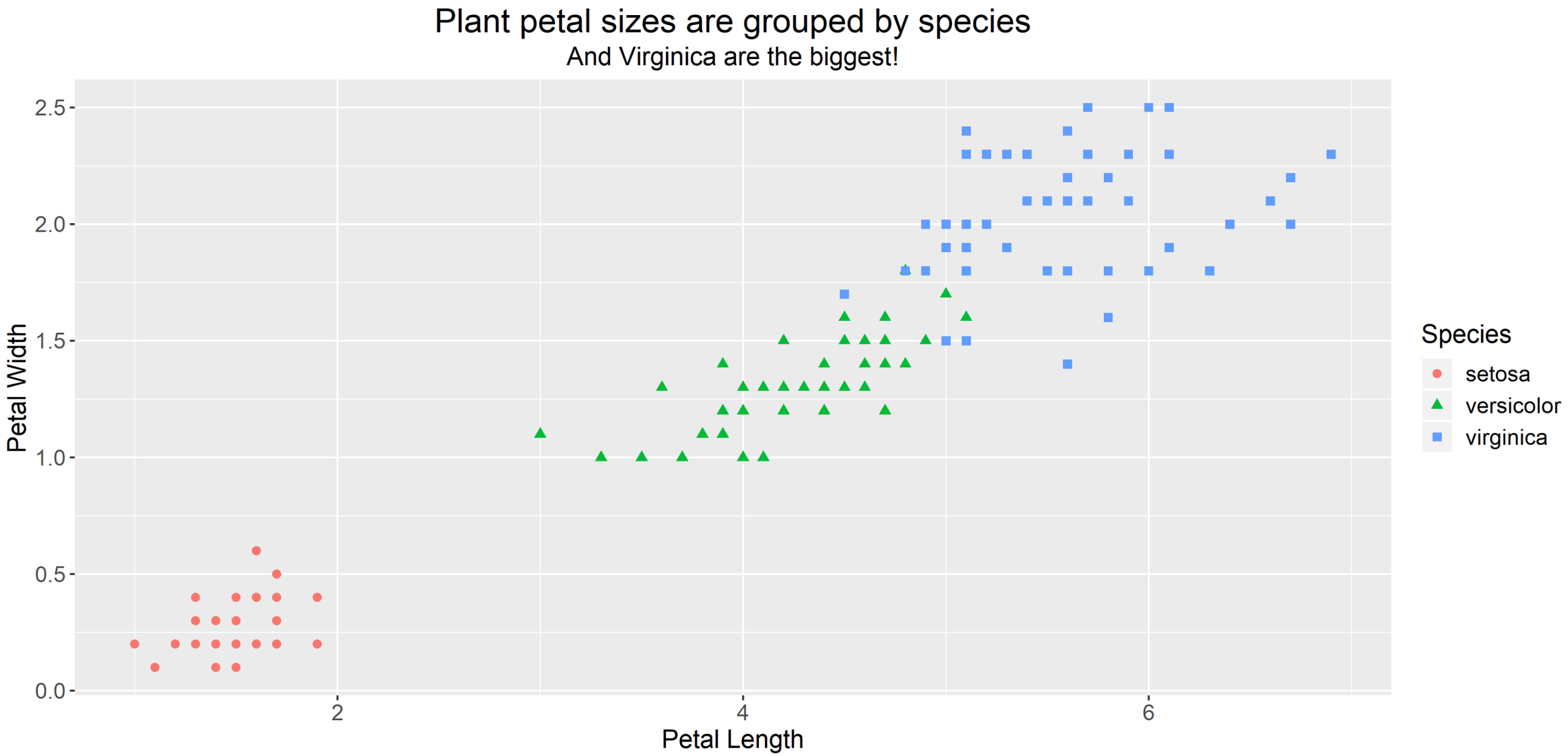
additional annotations, animations

legend & legend title (location, size, font, color, existence)

major & minor gridlines
(frequency, width, color)

X & Y axis labels
(size, font, color, angle)

X & Y tick marks
(frequency, size, font, color, angle)


Plant petal sizes are grouped by species
And Virginica are the biggest!
Petal Width
2.5
2.0
1.5
1.0
0.5
0.0
2
4
6
Petal Length
Species
setosa
versicolor
virginica

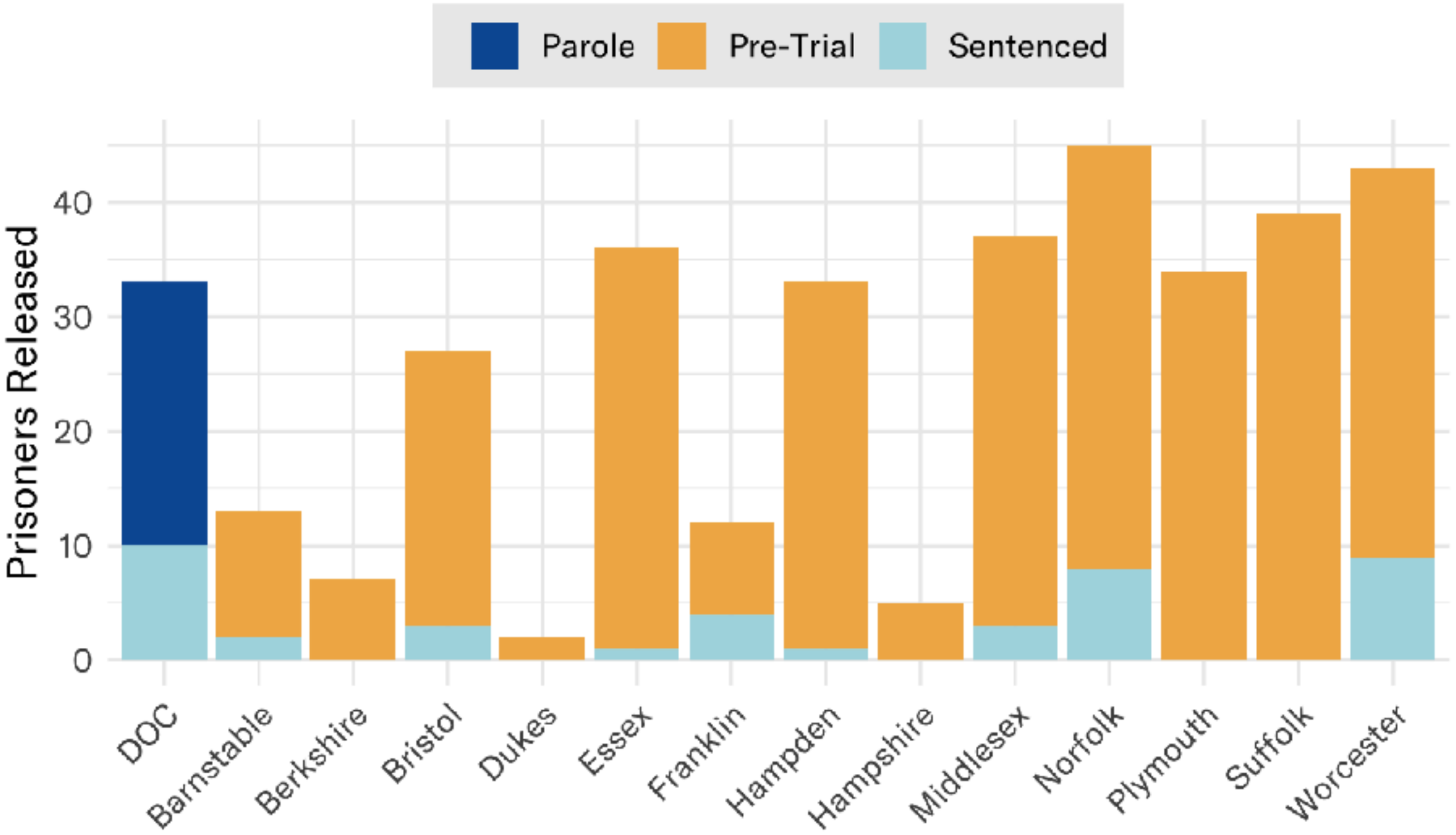
Types of data visualization

Tables

Rank	Boston Police Officer	FY20 Earnings	Change From FY19 Rank	Change From FY19 Earnings
1	Smith, Sean P.	$365,001	↑ +3	+$22,378
2	Lee, Waiman	$360,143	↑ +3	+$20,628
3	Brown, John M.	$345,758	↑ +12	+$28,388
4	Danilecki, John H.	$339,928	↓ -2	-$8,127
5	Barrett, Thomas E.	$339,353	↑ +2	+$7,275
6	McCormick, Kelley J.	$336,898	↑ +5	+$15,922
7	Connolly, Timothy	$334,077	↑ +18	+$31,109
8	Sweeney, Steve	$332,908	↑ +1	+$3,932
9	Cullity, Patrick J.	$329,101	↑ +3	+$9,032
10	Hughes, John C.	$322,637	↑ +31	+$31,203
11	Thomas, Terry J.	$319,001	↑ +9	+$11,134
12	Pusey, Daniel C.	$318,886	↑ +24	+$23,156
13	Webb, Keith A.	$318,141	↑ +1	-$200
14	Facey, Leighton B.	$317,506	↑ +29	+$28,844
15	Kervin, Timothy M.	$314,690	↓ -14	-$40,849
16	Harrington, Mark William	$314,443	↑ +29	+$27,098
17	Ciccolo, Robert W.	$314,026	↑ +9	+$12,894
18	Kozmiski, Therese M.	$313,656	↑ +53	+$35,552
19	Sullivan, Marc	$313,437	↓ -2	+$2,515
20	Gero, Jarrod A.	$311,210	↑ +19	+$17,573

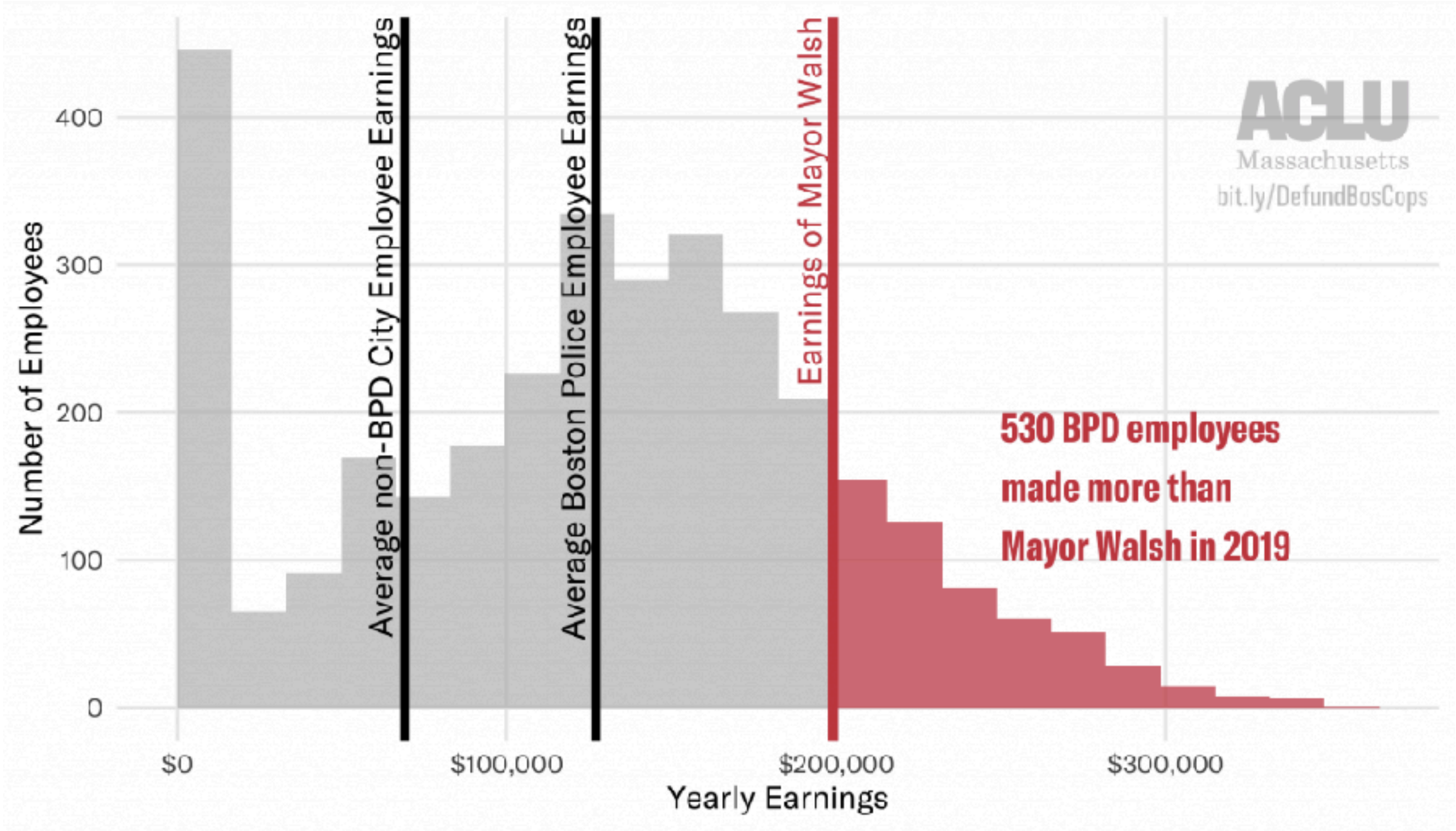
Source: "More of the Same: Unpacking the 2022 Boston Police Budget," ACLU of Massachusetts Data for Justice Project

Bar charts



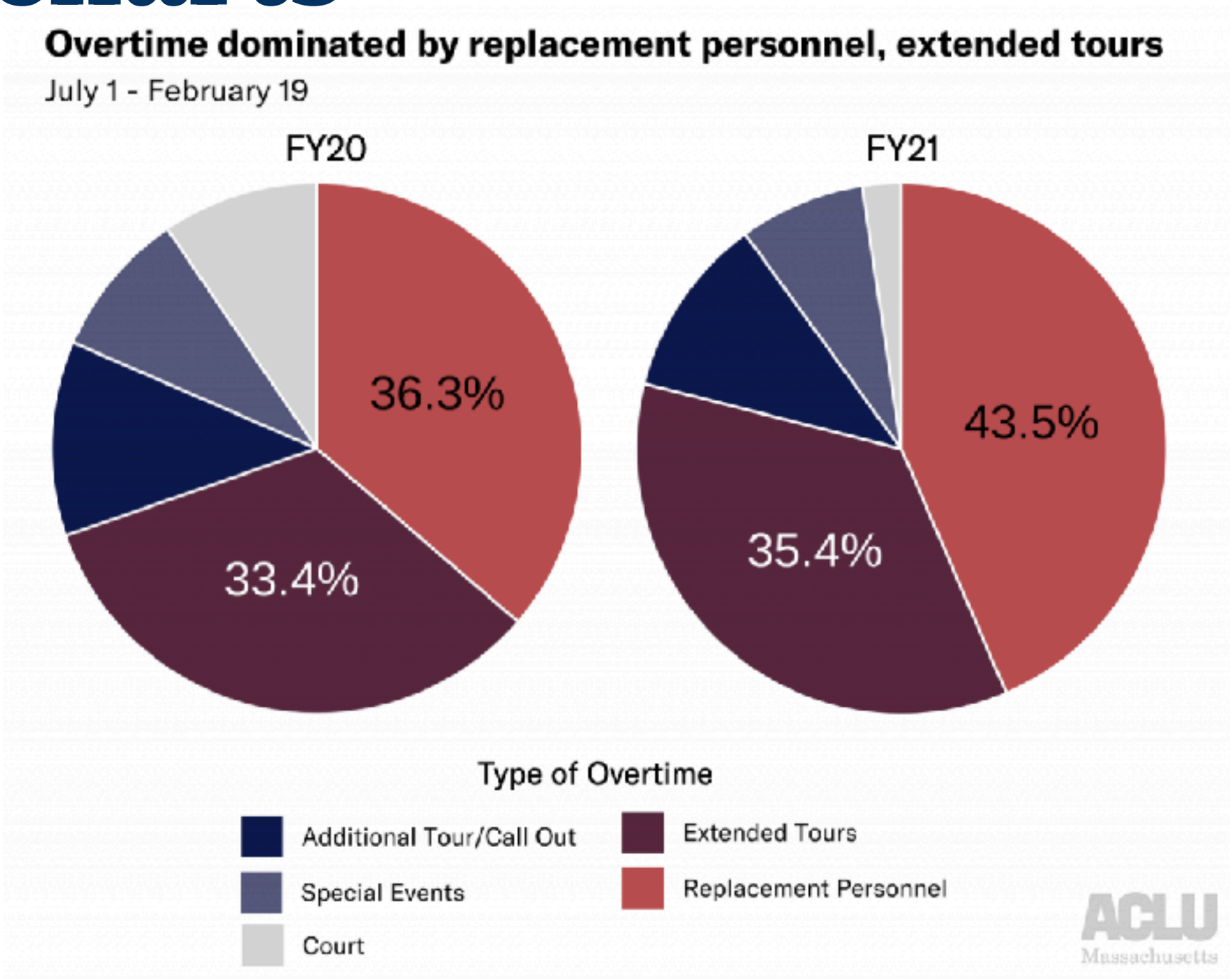
Source: "Incarcerated and In Danger: COVID-19 in Massachusetts Prisons and Jails,"ACLU of Massachusetts Data for Justice Project

Histogram



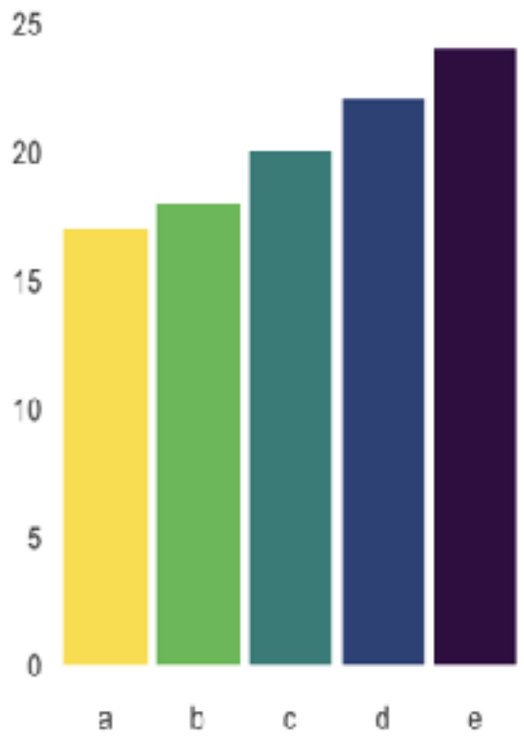
Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

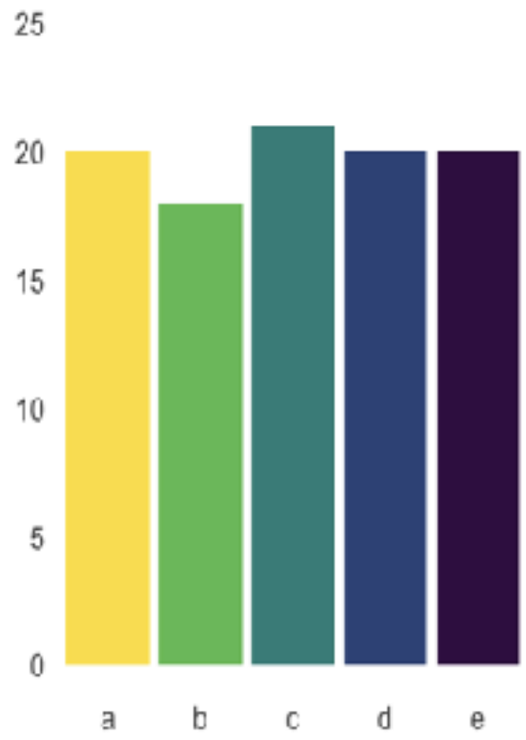
Pie charts

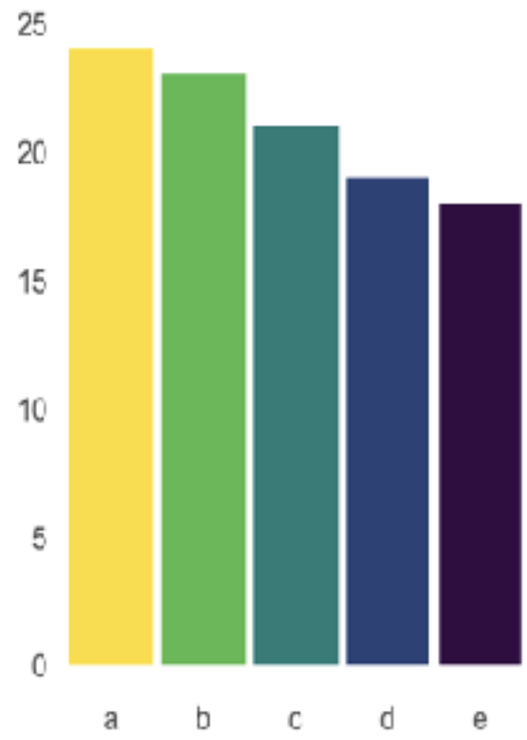


Source: "A Mayor's Roadmap to Curb Boston Police Overtime," ACLU of Massachusetts Data for Justice Project

Pie charts

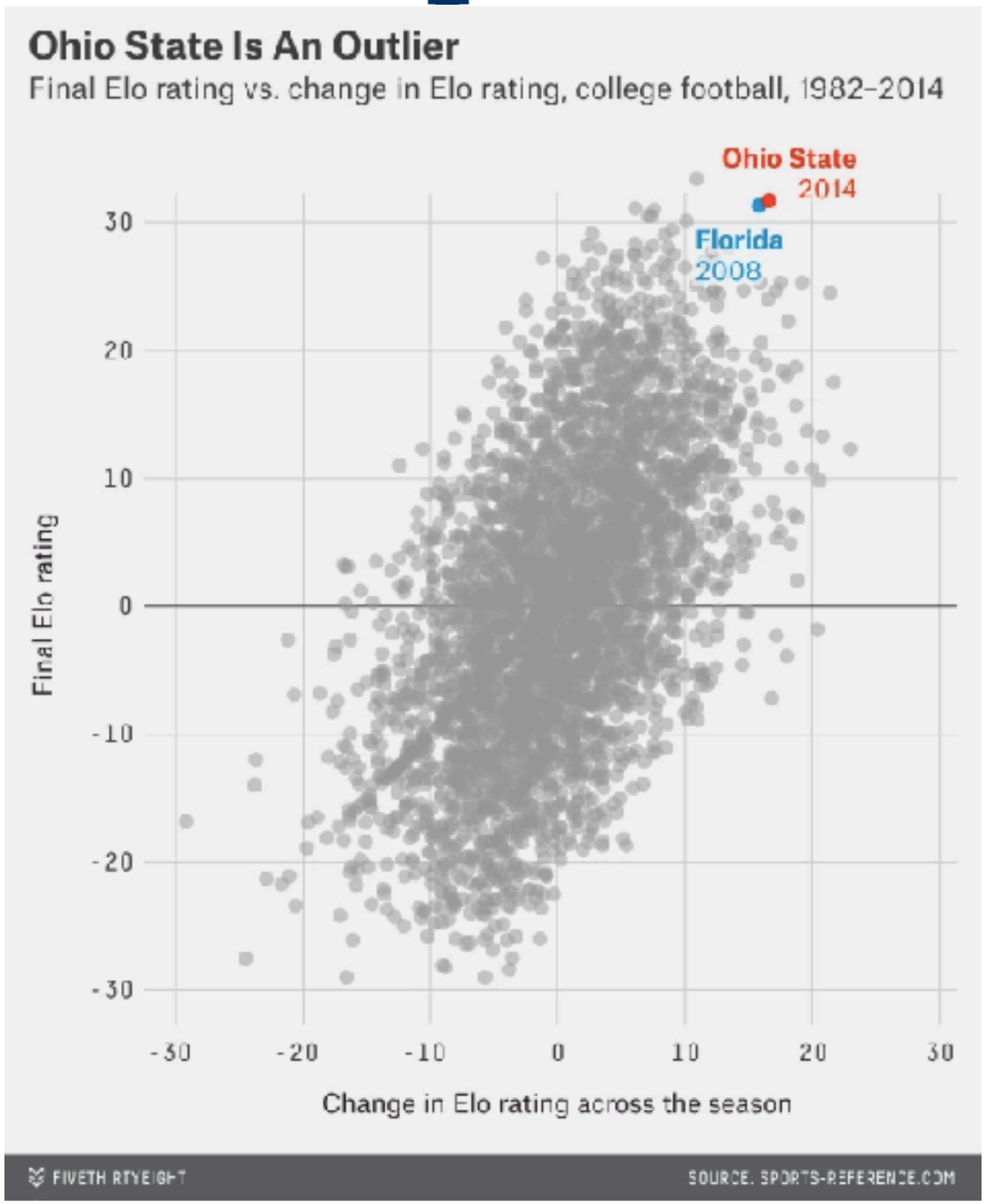




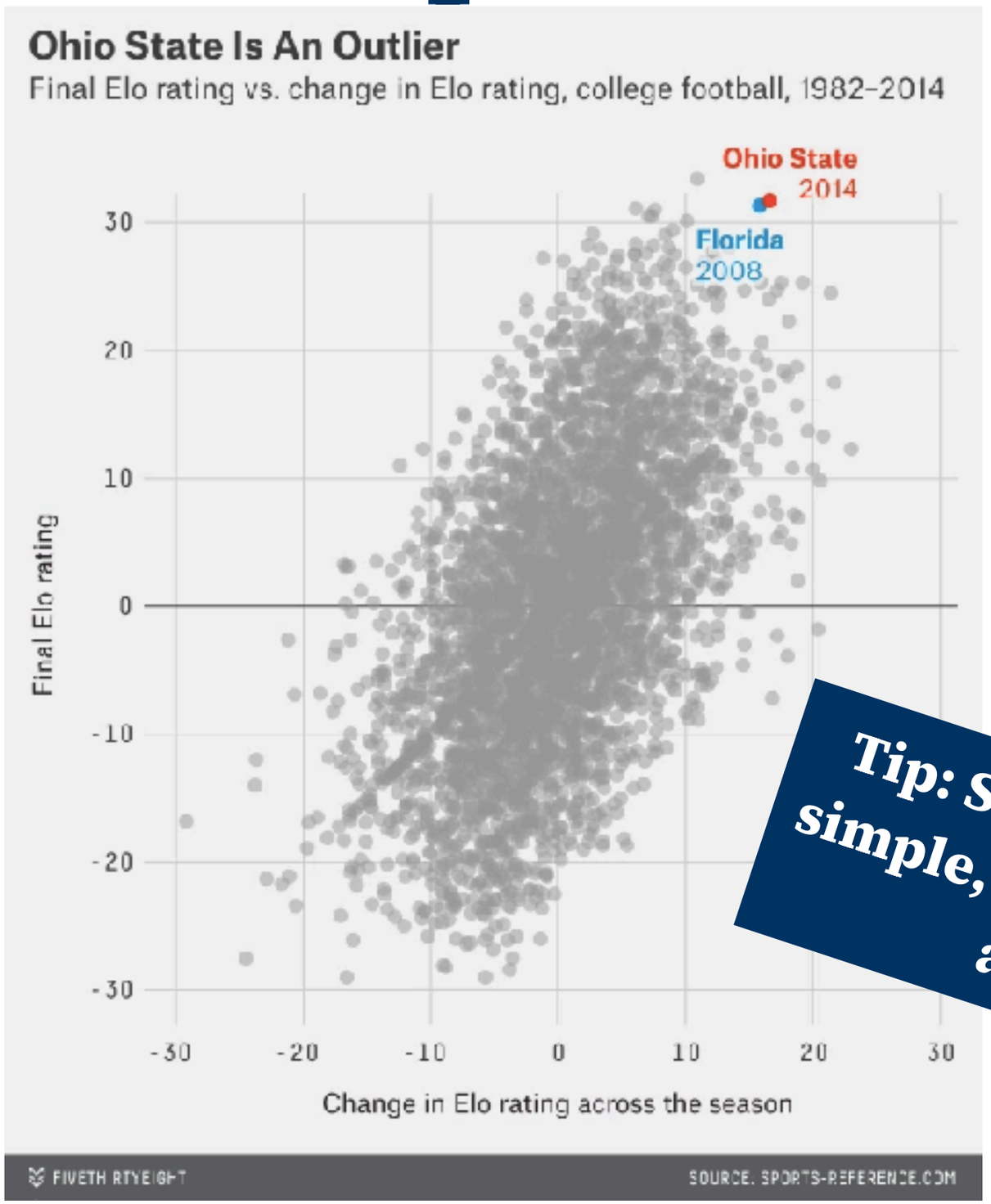


Source: "The Issue with Pie Chart," Data to Viz

Scatter/circle plots

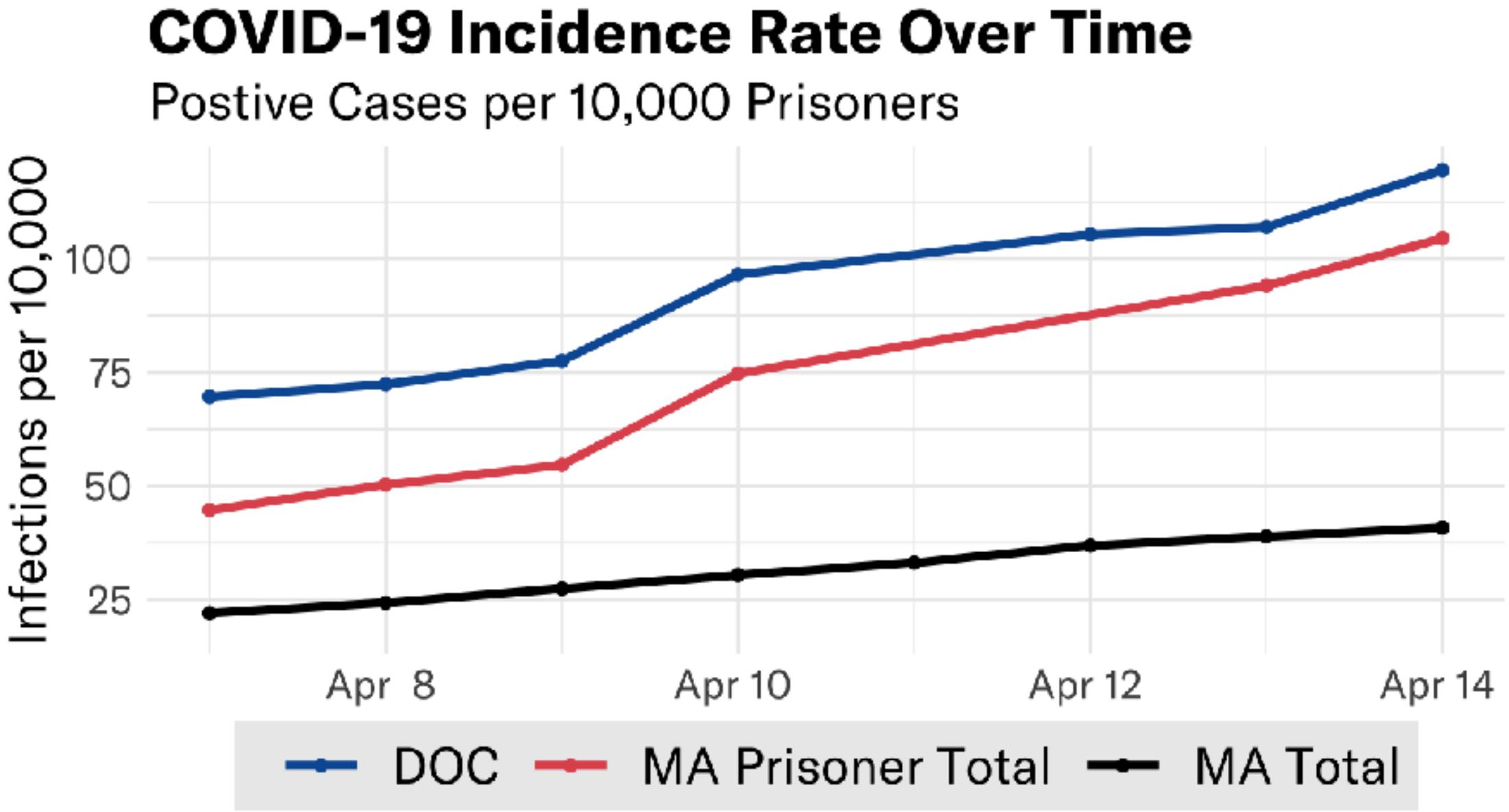


Scatter/circle plots



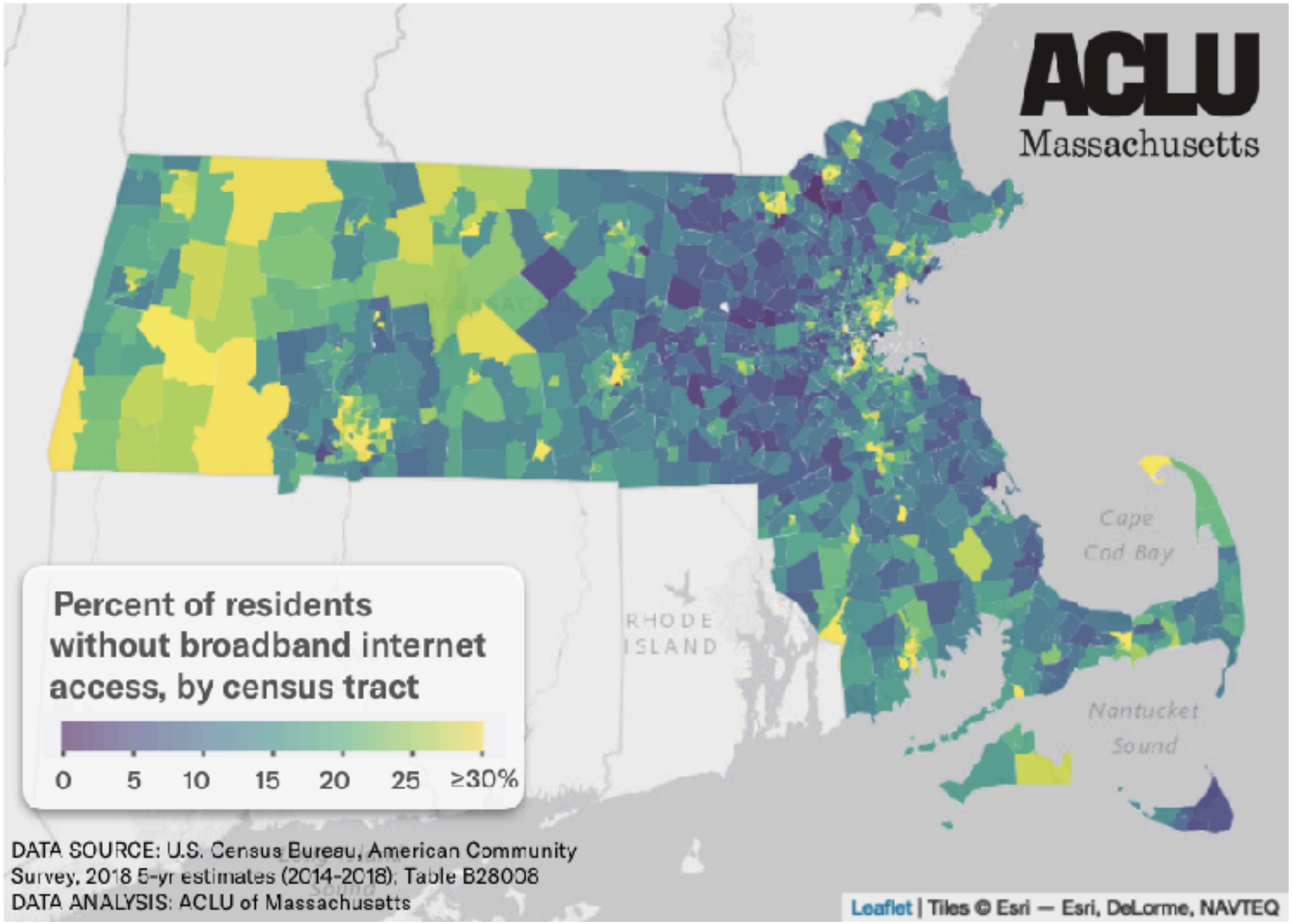
Tip: Scatter plots seem simple, but require extra attention

Line plots



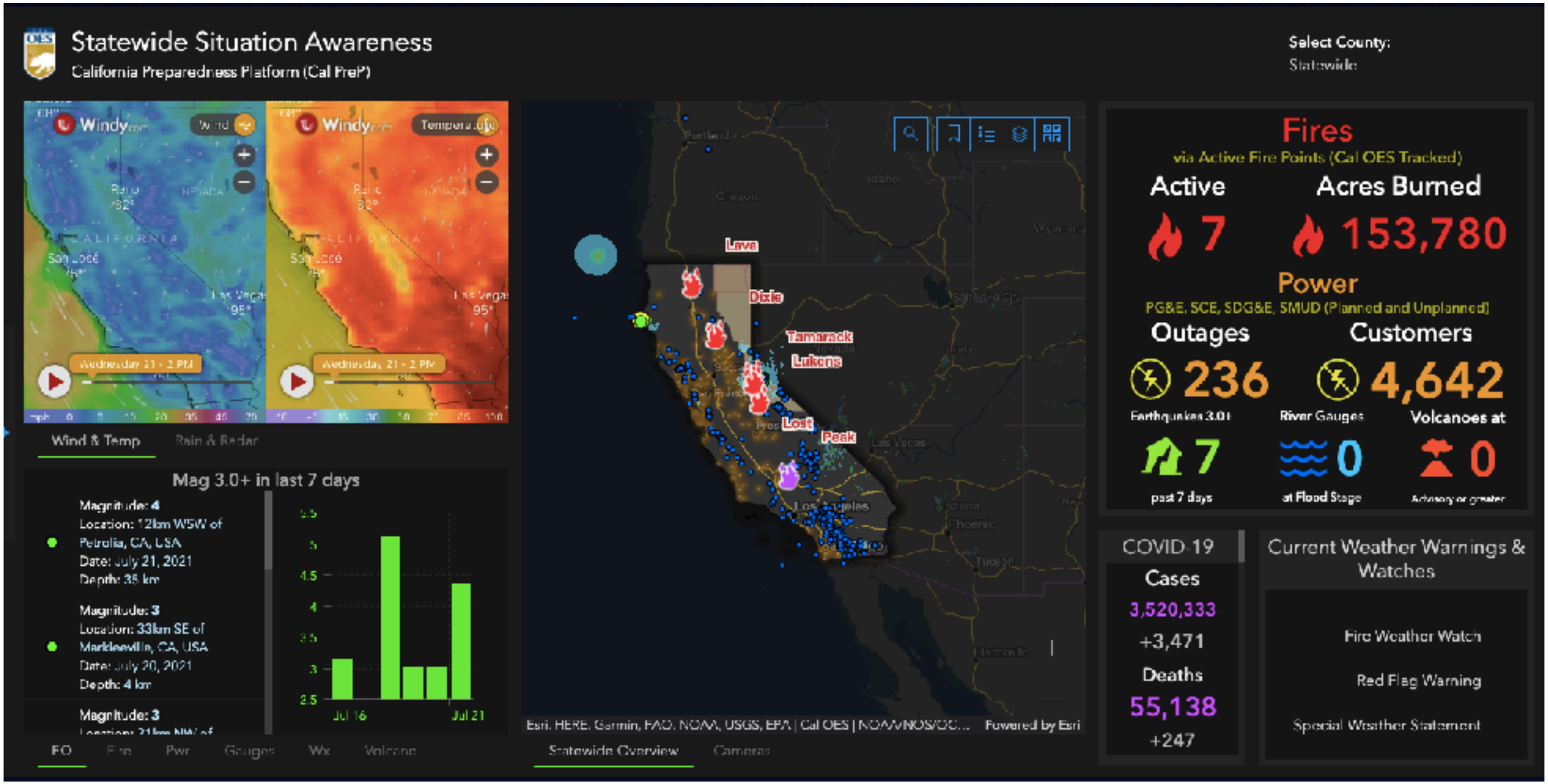
Source: "Incarcerated and In Danger: COVID-19 in Massachusetts Prisons and Jails,"ACLU of Massachusetts Data for Justice Project

Maps



Source: "Internet Deserts Prevent Remote Learning in COVID-19," ACLU of Massachusetts Data for Justice Project

Dashboards

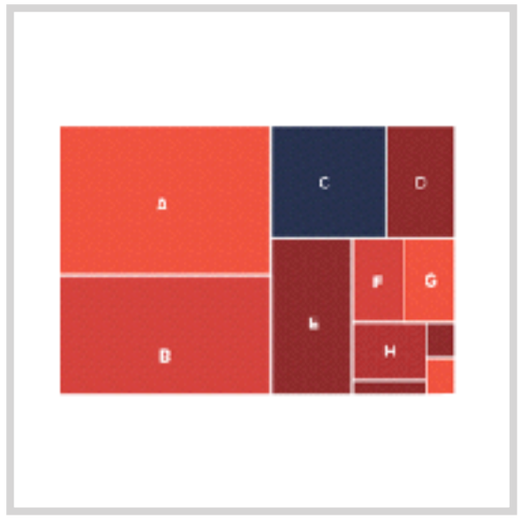


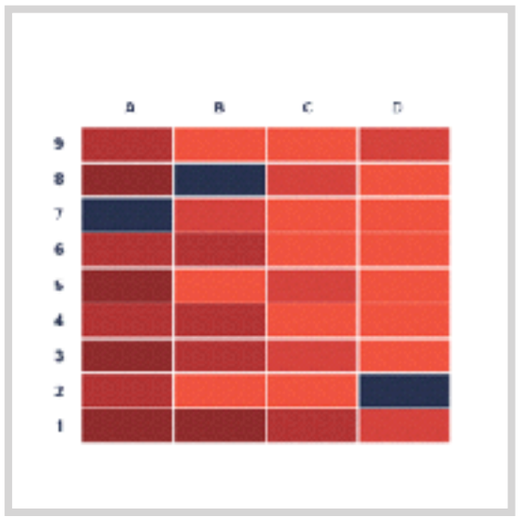
Source: California Governor's Office of Emergency Services Preparedness Platform

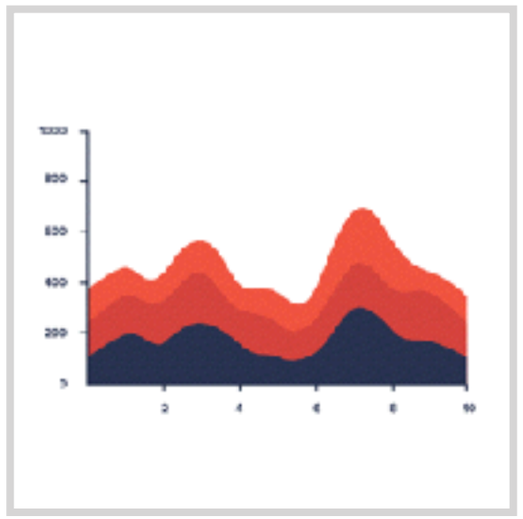
And more!

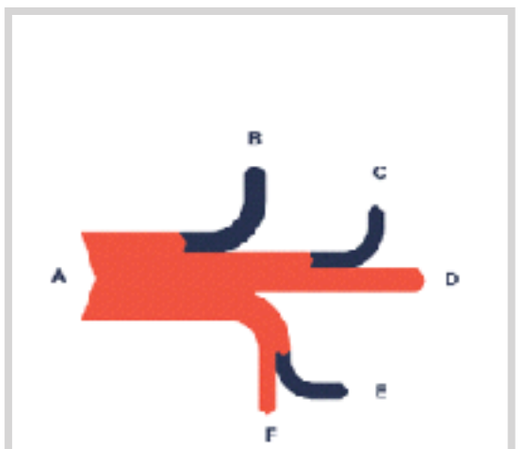
- Pictorial charts
- Treemaps
- 2-D histogram / heat map
- Area chart
- Alluvial/Sankey diagram
- Dumbbell plot

For more, see: ***datavizproject.com***









Source: Data Viz Project

matplotlib

Cheat sheet

Quick start

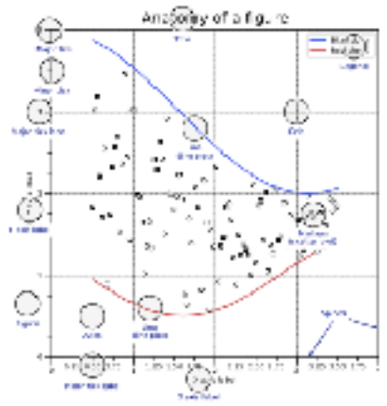
```
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt


X = np.linspace(0, 2*np.pi, 100)
Y = np.cos(X)

fig, ax = plt.subplots()
ax.plot(X,Y,color='C1')

fig.savefig("figure.pdf")
fig.show()
```

Anatomy of a figure



Subplots layout

```
subplot[s](nrows,ncols,..)
fig, axs = plt.subplots(3,3)

G = gridspec(rows,cols,..)
ax = G[3,:]

ax.inset_axes(extent)

d=make_axes_locatable(ax)
ax=d.new_horizontal('15%')
```

Getting help

- matplotlib.org
- github.com/matplotlib/matplotlib/issues
- discourse.matplotlib.org
- stackoverflow.com/questions/tagged/matplotlib
- gitter.im/matplotlib
- twitter.com/matplotlib
- Matplotlib users mailing list

Basic plots

```
plot([X],Y,[fmt],..)
```
X, Y, fmt, color, marker, linestyle

```
scatter(X,Y,..)
```
X, Y, sizes, colors, markers, cmap

```
bar[h](x,height,..)
```
x, height, width, bottom, align, color

```
imshow(Z,[cmap],..)
```
Z, cmap, interpolation, extent, origin

```
contour[f]([X],[Y],Z,..)
```
X, Y, **Z**, levels, colors, extent, origin

```
quiver([X],[Y],U,V,..)
```
X, Y, **U, V**, C, units, angles

```
pie(X,[explode],..)
```
X, explode, labels, colors, radius


```
text(x,y,text,..)
```
x, y, text, va, ha, size, weight, transform

```
fill[_between][x]( .. )
```
X, Y1, Y2, color, where

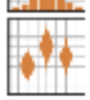
Advanced plots

```
step(X,Y,[fmt],..)
```
X, Y, fmt, color, marker, where

```
boxplot(X,..)
```
X, notch, sym, bootstrap, widths

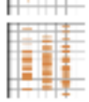
```
errorbar(X,Y,xerr,yerr,..)
```
X, Y, xerr, yerr, fmt

```
hist(X, bins, ..)
```
X, bins, range, density, weights

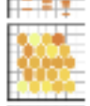


```
violinplot(D,..)
```
D, positions, widths, vert

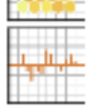
```
barbs([X],[Y],U,V,..)
```
X, Y, **U, V**, C, length, pivot, sizes



```
eventplot(positions,..)
```
positions, orientation, lineoffsets



```
hexbin(X,Y,C,..)
```
X, Y, C, gridsize, cmap

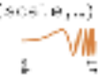


```
xcorr(X,Y,..)
```
X, Y, normed, detrend

Scales

```
ax.set_[xy]scale(scale,..)
```

linear
any values



log
values > 0

symlog
any values

logit
0 < values < 1

Projections

```
subplot(..,projection=mp)
```

p='polar'

p='3d'

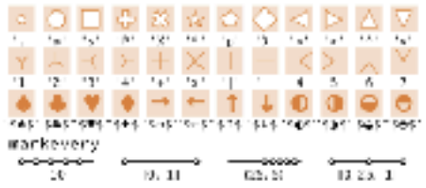
p=Orthographic() API
from cartopy.crs import Cartographic

Lines

```
linestyle or ls
```


```
capstyle or dash_capstyle
```

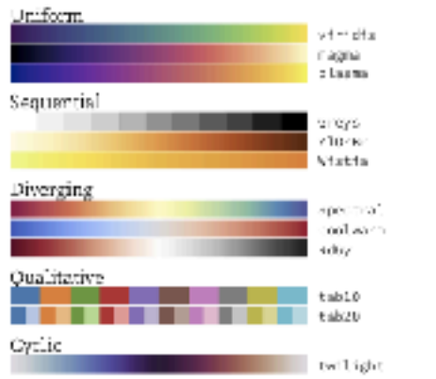
Markers



Colors

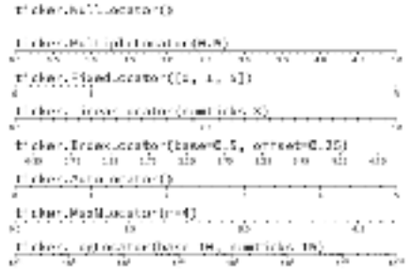
Colormaps

```
plt.get_cmap(name)
```



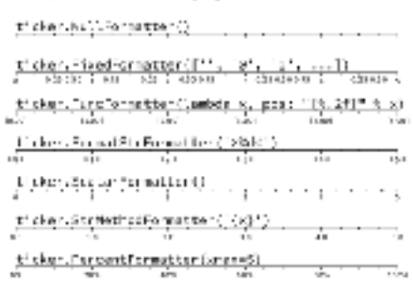
Tick locators

```
from matplotlib import ticker
ax.[xy]axis.set_[minor|major]_locator(locator)
```



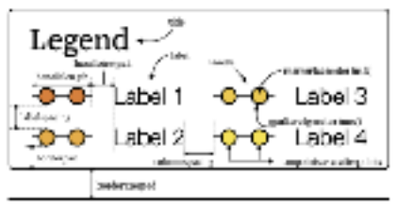
Tick formatters

```
from matplotlib import ticker
ax.[xy]axis.set_[minor|major]_formatter(formatter)
```

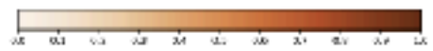


Ornaments

```
ax.legend(..)
```
handles, labels, loc, title, frameon




```
ax.colorbar(..)
```
mappable, ax, cax, orientation



```
ax.annotate(..)
```
text, xy, xytext, xycoords, textcoords, arrowprops

Event handling

```
fig, ax = plt.subplots()
def on_click(event):
  print(event)
fig.canvas.mpl_connect(
  'button_press_event', on_click)
```

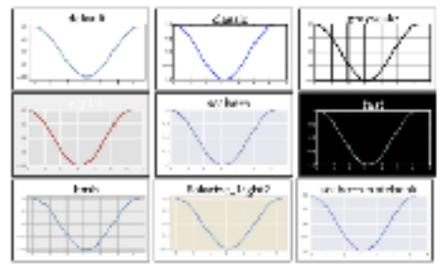
Animation

```
import matplotlib.animation as mpla

T = np.linspace(0,2*np.pi,100)
S = np.sin(T)
line, = plt.plot(T, S)
def animate(i):
  line.set_ydata(np.sin(T+i/50))
ani = mpla.FuncAnimation(
  plt.gcf(), animate, interval=5)
plt.show()
```

Styles

```
plt.style.use(style)
```



Quick reminder

```
ax.grid()
ax.patch.set_alpha(3)
ax.set_[xy]lim(vmin, vmax)
ax.set_[xy]label('Label')
ax.set_[xy]ticks(list)
ax.set_[xy]ticklabels(list)
ax.set_[sup]title(title)
ax.tick_params(width=10, ..)
ax.set_axis_[on|off]()

ax.tight_layout()
plt.gcf(), plt.gca()
mpl.rc('axes', linewidth=, ..)
fig.patch.set_alpha(8)
text='$\frac{-e^{i\pi}}{2^n}$'
```

Keyboard shortcuts

ctrl + s	Save	ctrl + w	Close plot
r	Reset view	f	Fullscreen 0/1
f	View forward	h	View back
p	Pan view	o	Zoom to rect.
x	X pan/zoom	y	Y pan/zoom
g	Minor grid 0/1	G	Major grid 0/1
l	X axis log/linear	L	Y axis log/linear

Ten Simple Rules

READ

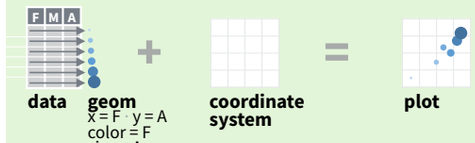
1. Know Your Audience
2. Identify Your Message
3. Adapt the Figure
4. Captions Are Not Optional
5. Do Not Trust the Defaults
6. Use Color Effectively
7. Do Not Mislead the Reader
8. Avoid 'Chartjunk'
9. Message Trumps Beauty
10. Get the Right Tool

Data Visualization with ggplot2 : : CHEAT SHEET

Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and geoms—visual marks that represent data points.

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.



Complete the template below to build a graph.

```
ggplot (data = <DATA>) +                          required
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),
 stat = <STAT>, position = <POSITION>) +          Not required, sensible defaults supplied
<COORDINATE_FUNCTION> +
<FACET_FUNCTION> +
<SCALE_FUNCTION> +
<THEME_FUNCTION>
```

ggplot(data = mpg, **aes(**x = cty, y = hwy**))** Begins a plot that you finish by adding layers to. Add one geom function per layer.

aesthetic mappings · data · geom

qplot(x = cty, y = hwy, data = mpg, geom = "point"**)** Creates a complete plot with given data, geom, and mappings. Supplies many useful defaults.

last_plot() Returns the last plot

ggsave("plot.png", width = 5, height = 5) Saves last plot as 5' x 5' file named "plot.png" in working directory. Matches file type to file extension.

Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

GRAPHICAL PRIMITIVES

a <- ggplot(economics, aes(date, unemploy))
b <- ggplot(seals, aes(x = long, y = lat))

a + geom_blank()
(Useful for expanding limits)

b + geom_curve(aes(yend = lat + 1, xend=long+1),curvature=1**)** - x, xend, y, yend, alpha, angle, color, curvature, linetype, size

a + geom_path(lineend="butt", linejoin="round", linemitre=1**)**
x, y, alpha, color, group, linetype, size

a + geom_polygon(aes(group = group)**)**
x, y, alpha, color, fill, group, linetype, size

b + geom_rect(aes(xmin = long, ymin=lat, xmax= long + 1, ymax = lat + 1)**)** - xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size

a + geom_ribbon(aes(ymin=unemploy - 900, ymax=unemploy + 900)**)** - x, ymax, ymin, alpha, color, fill, group, linetype, size

LINE SEGMENTS

common aesthetics: x, y, alpha, color, linetype, size

b + geom_abline(aes(intercept=0, slope=1)**)**
b + geom_hline(aes(yintercept = lat)**)**
b + geom_vline(aes(xintercept = long)**)**

b + geom_segment(aes(yend=lat+1, xend=long+1)**)**
b + geom_spoke(aes(angle = 1:1155, radius = 1)**)**

ONE VARIABLE continuous

c <- ggplot(mpg, aes(hwy)); c2 <- ggplot(mpg)

c + geom_area(stat = "bin")
x, y, alpha, color, fill, linetype, size

c + geom_density(kernel = "gaussian"**)**
x, y, alpha, color, fill, group, linetype, size, weight

c + geom_dotplot()
x, y, alpha, color, fill

c + geom_freqpoly() x, y, alpha, color, group, linetype, size

c + geom_histogram(binwidth = 5**)** x, y, alpha, color, fill, linetype, size, weight

c2 + geom_qq(aes(sample = hwy)**)** x, y, alpha, color, fill, linetype, size, weight

discrete

d <- ggplot(mpg, aes(fl))

d + geom_bar()
x, alpha, color, fill, linetype, size, weight

TWO VARIABLES

continuous x , continuous y

e <- ggplot(mpg, aes(cty, hwy))

e + geom_label(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)** x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

e + geom_jitter(height = 2, width = 2**)**
x, y, alpha, color, fill, shape, size

e + geom_point(), x, y, alpha, color, fill, shape, size, stroke

e + geom_quantile(), x, y, alpha, color, group, linetype, size, weight

e + geom_rug(sides = "bl"**)**, x, y, alpha, color, linetype, size

e + geom_smooth(method = lm**)**, x, y, alpha, color, fill, group, linetype, size, weight

e + geom_text(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)**, x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

discrete x , continuous y

f <- ggplot(mpg, aes(class, hwy))

f + geom_col(), x, y, alpha, color, fill, group, linetype, size

f + geom_boxplot(), x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight

f + geom_dotplot(binaxis = "y", stackdir = "center"**)**, x, y, alpha, color, fill, group

f + geom_violin(scale = "area"**)**, x, y, alpha, color, fill, group, linetype, size, weight

discrete x , discrete y

g <- ggplot(diamonds, aes(cut, color))

g + geom_count(), x, y, alpha, color, fill, shape, size, stroke

continuous bivariate distribution

h <- ggplot(diamonds, aes(carat, price))

h + geom_bin2d(binwidth = c(0.25, 500)**)**
x, y, alpha, color, fill, linetype, size, weight

h + geom_density2d()
x, y, alpha, colour, group, linetype, size

h + geom_hex()
x, y, alpha, colour, fill, size

continuous function

i <- ggplot(economics, aes(date, unemploy))

i + geom_area()
x, y, alpha, color, fill, linetype, size

i + geom_line()
x, y, alpha, color, group, linetype, size

i + geom_step(direction = "hv")
x, y, alpha, color, group, linetype, size

visualizing error

df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
j <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))

j + geom_crossbar(fatten = 2**)**
x, y, ymax, ymin, alpha, color, fill, group, linetype, size

j + geom_errorbar(), x, ymax, ymin, alpha, color, group, linetype, size, width (also **geom_errorbarh()**)

j + geom_linerange()
x, ymin, ymax, alpha, color, group, linetype, size

j + geom_pointrange()
x, y, ymin, ymax, alpha, color, fill, group, linetype, shape, size

maps

data <- data.frame(murder = USArrests$Murder, state = tolower(rownames(USArrests)))
map <- map_data("state")
k <- ggplot(data, aes(fill = murder))

k + geom_map(aes(map_id = state), map = map**)**
+ expand_limits(x = map$long, y = map$lat**)**,
map_id, alpha, color, fill, linetype, size

THREE VARIABLES

seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2)); l <- ggplot(seals, aes(long, lat))

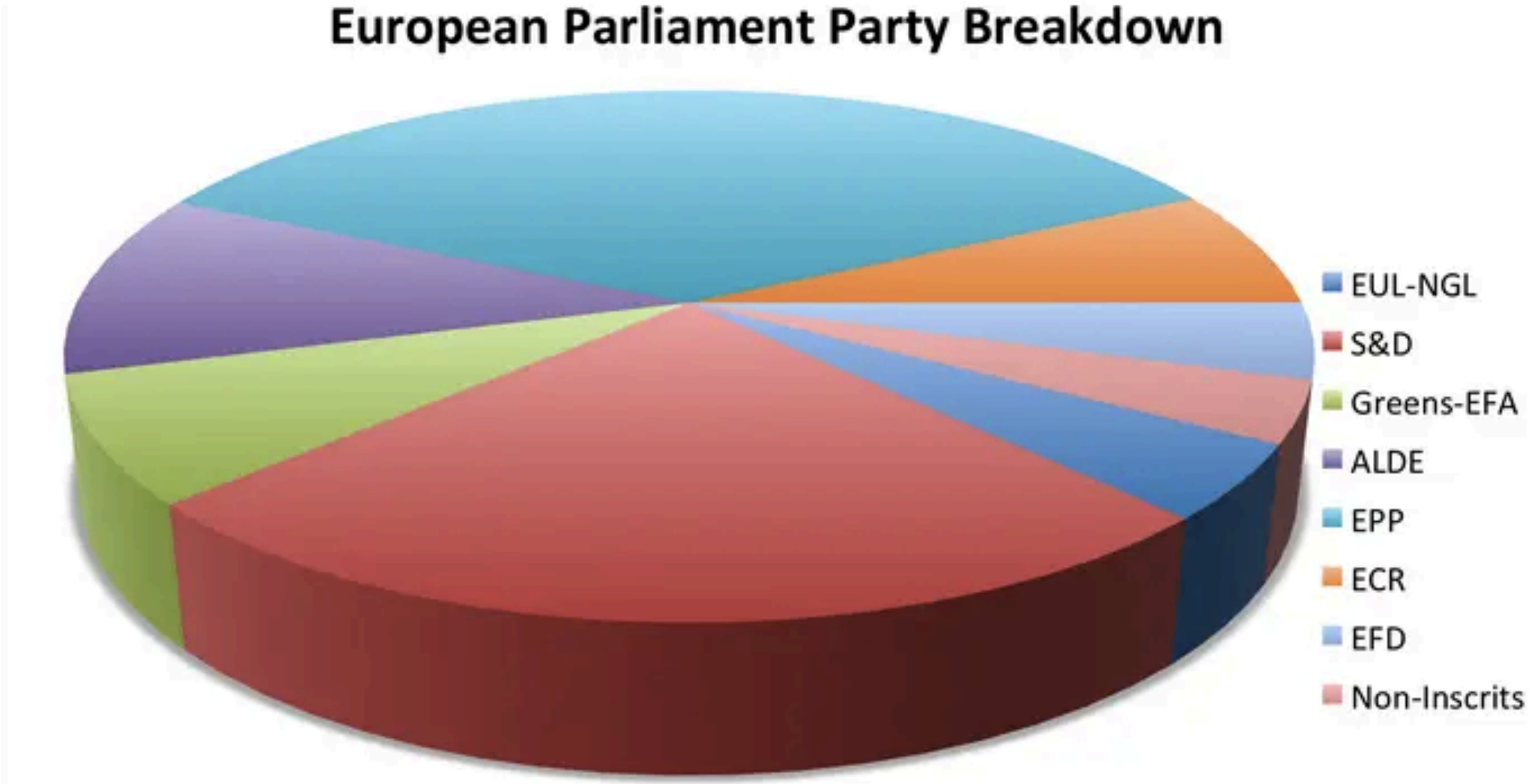
l + geom_contour(aes(z = z))
x, y, z, alpha, colour, group, linetype, size, weight

l + geom_raster(aes(fill = z), hjust=0.5, vjust=0.5, interpolate=FALSE**)**
x, y, alpha, fill

l + geom_tile(aes(fill = z)**)**, x, y, alpha, color, fill, linetype, size, width

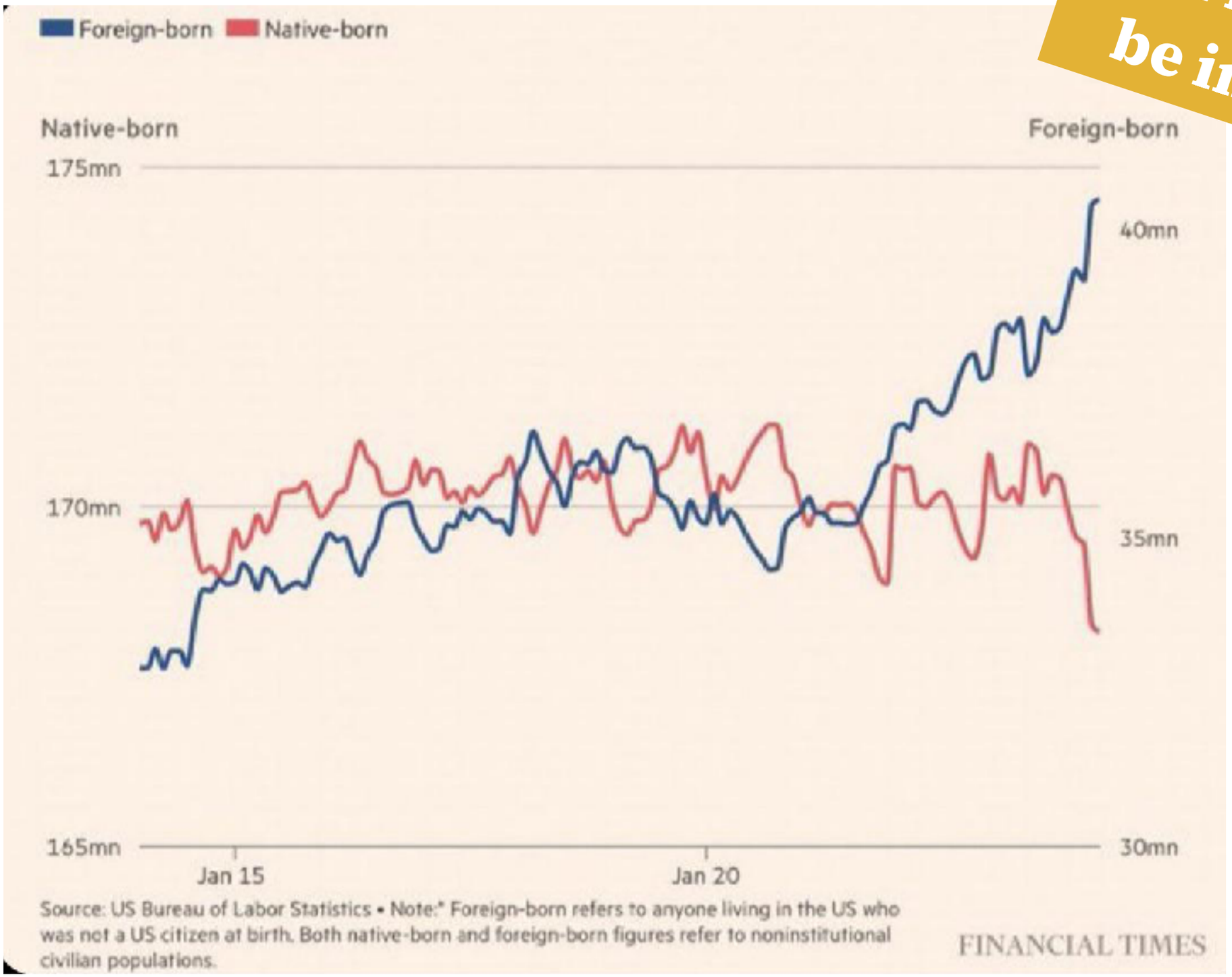
Roast! This! Viz!

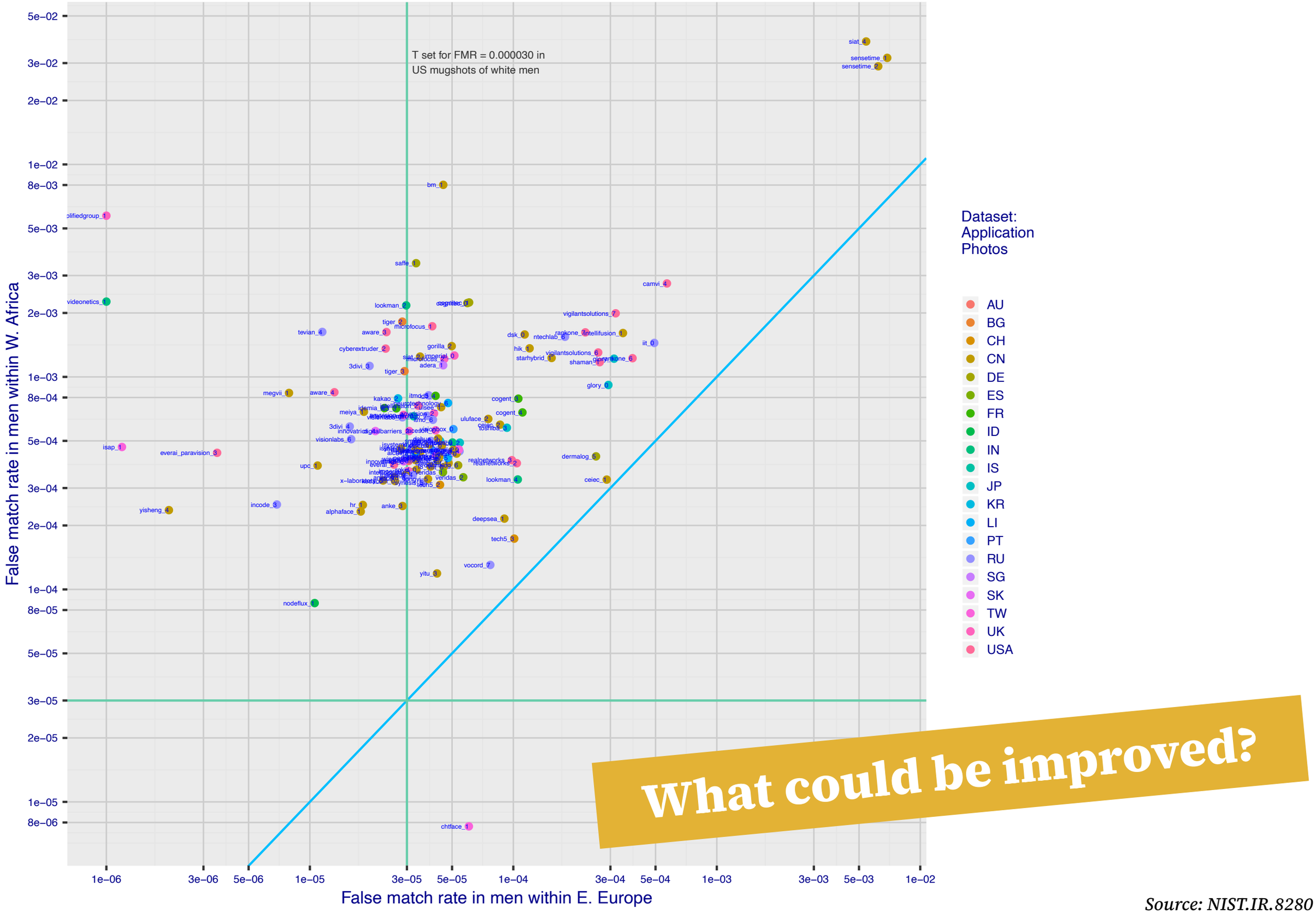
What could be improved?



Source: Walter Hickey, Business Insider

What could be improved?





What could be improved?

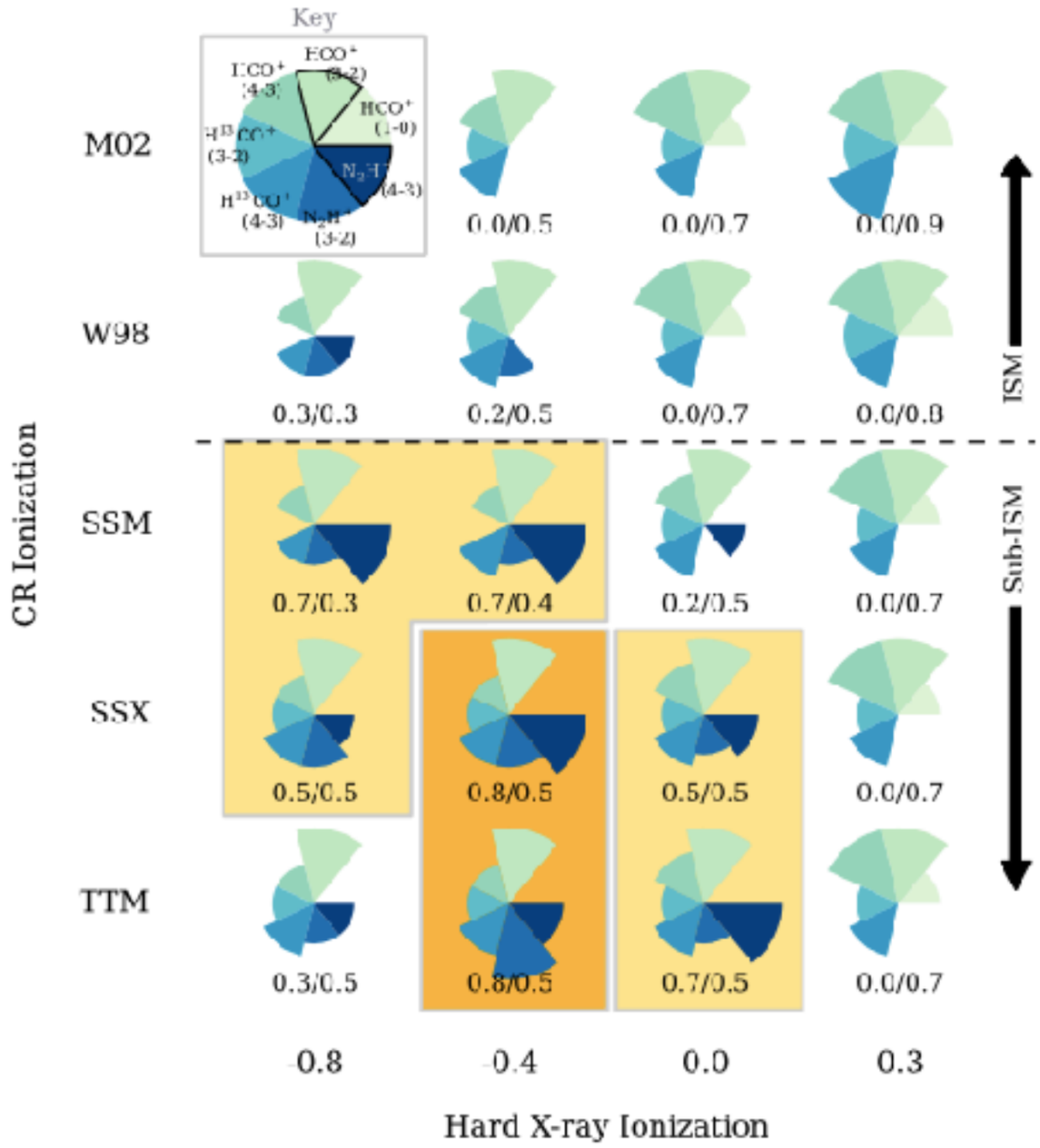
Source: NIST.IR.8280

10 Best Practices for Effective Data Visualization

1. **Know your audience!**
   - For public vizzes, this usually requires bar, line, and scatter plots.
   - Basically never use undefined acronyms or symbols.
2. **Question the defaults.** This means fonts, colors, sizes, shapes, and more.
3. **Use descriptive titles:** *tell* your viewer what they should see, and don't be redundant.
4. **Be selective in data and design.** Consider your data-to-ink ratio and remove distractions.
5. **Use color sparingly, intentionally, and consistently.**
   - Keep branding in mind
   - Remember colorblindness
6. **Use reasonable axis tick** formats and frequencies, not just the default.
7. **Consider creative alternatives to legends.**
8. **Don't be afraid of annotations.** Use whatever software you prefer to add them.
9. **Make your text bigger** than you'd think. Nobody likes to squint!
10. **Use the blink test.**

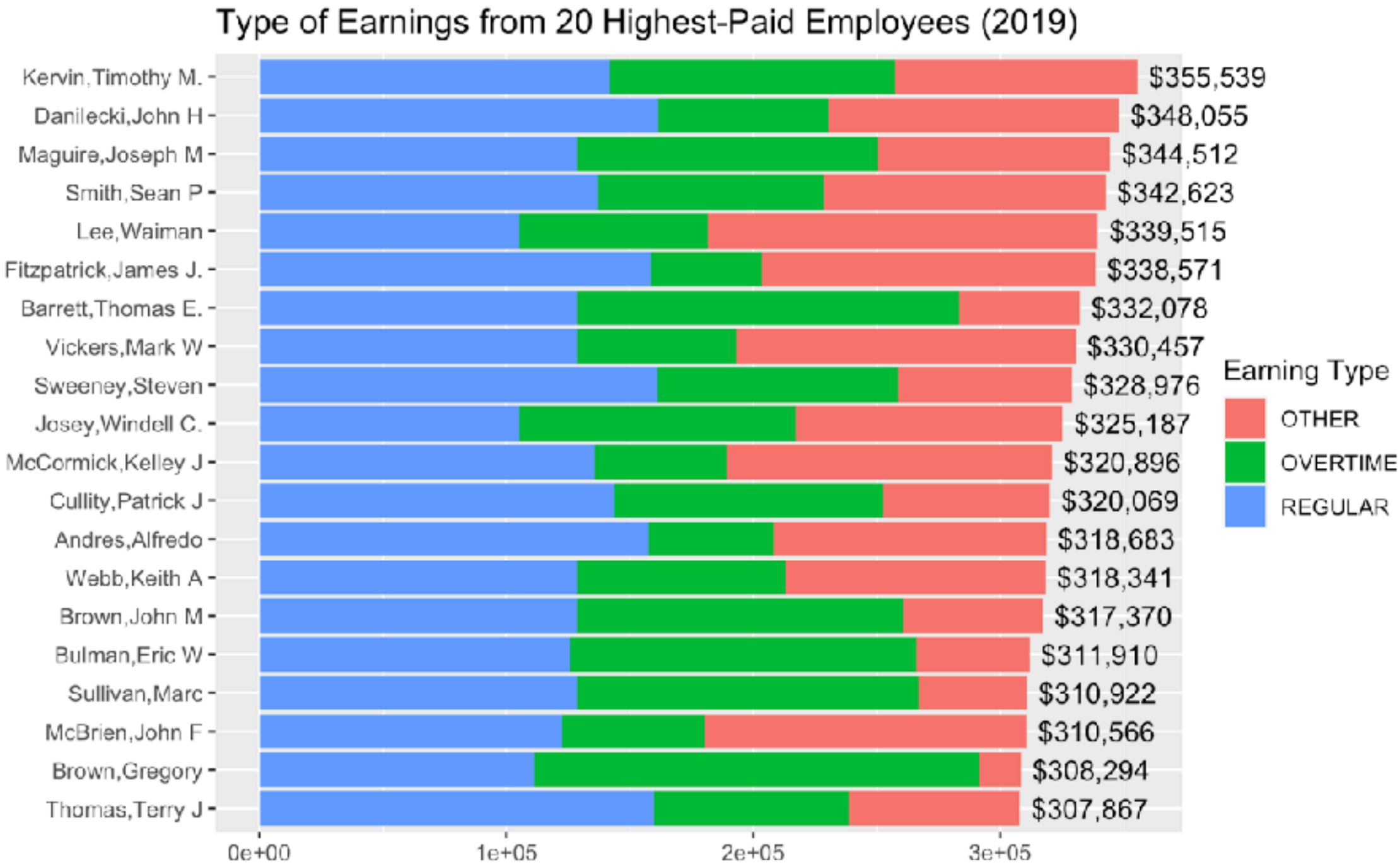
1. Know your audience!

- For public vizzes, this usually requires bar, line, and scatter plots.
- Basically never use undefined acronyms or symbols.



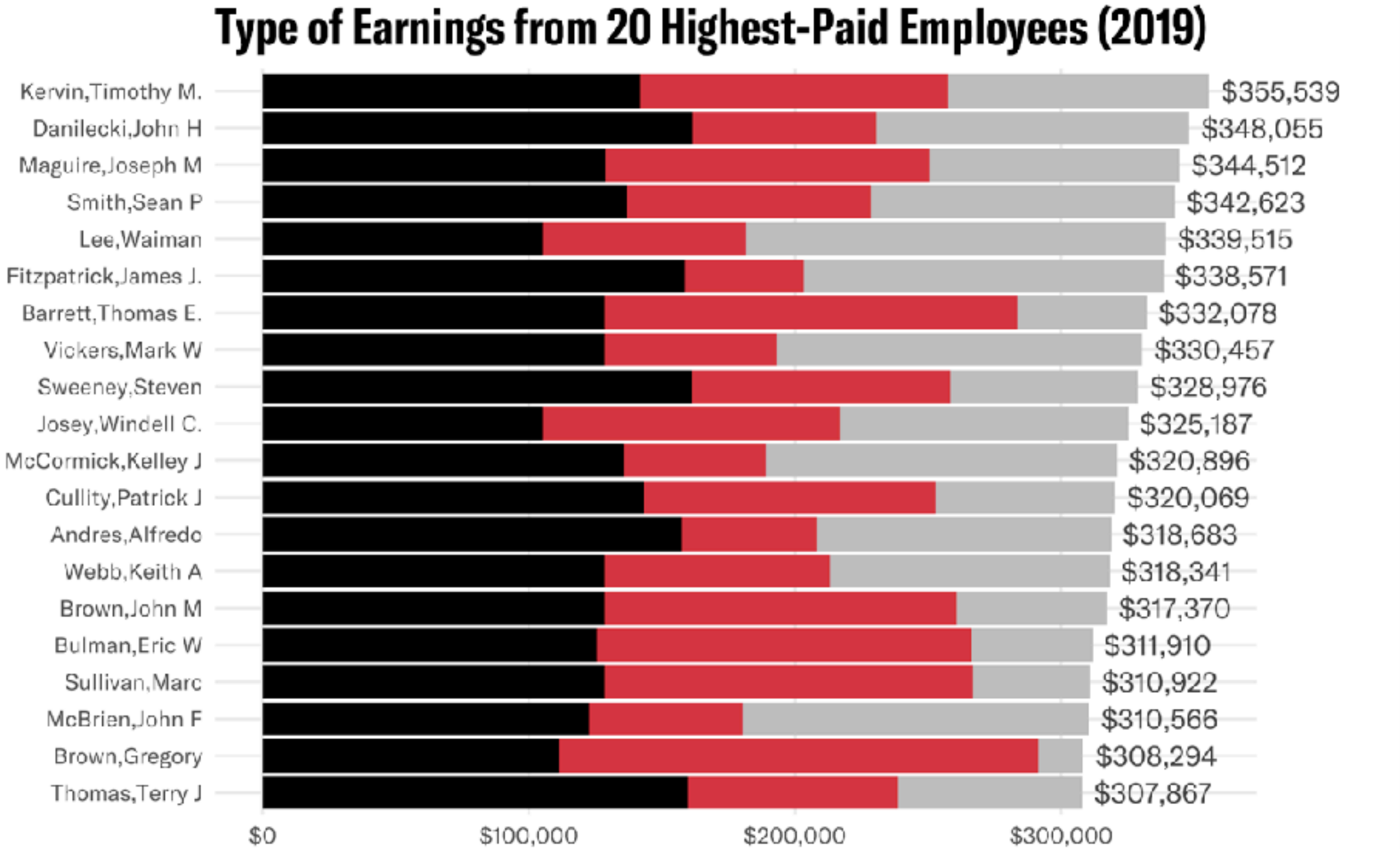
Source: "Constraining the X-ray and Cosmic Ray Ionization Chemistry of the TW Hya Protoplanetary Disk: Evidence for a Sub-interstellar Cosmic Ray Rate," Cleeves et al., 2015

2. **Question the defaults.** This means fonts, colors, sizes, shapes, and more.



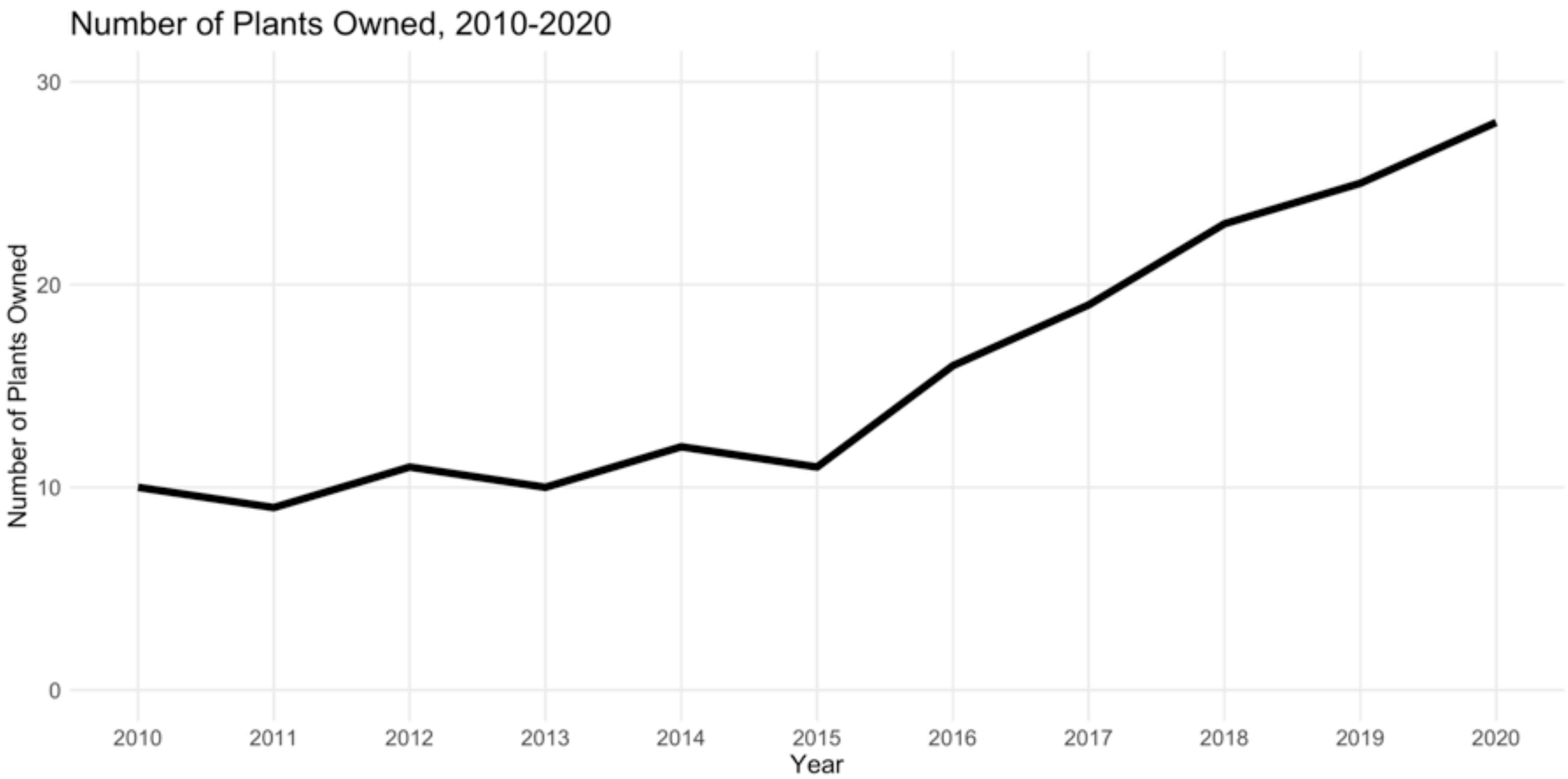
Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

2. **Question the defaults.** This means fonts, colors, sizes, shapes, and more.

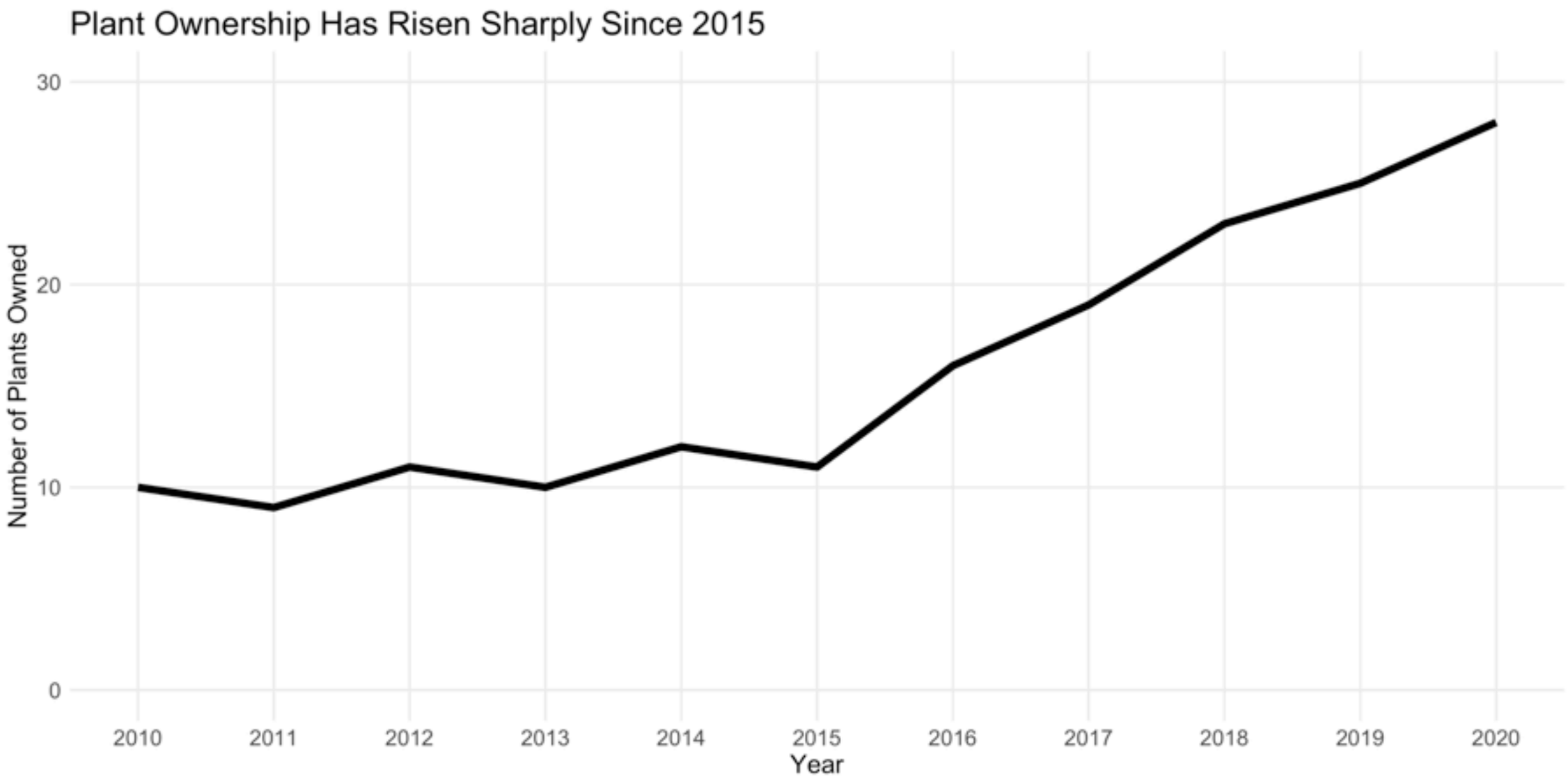


Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

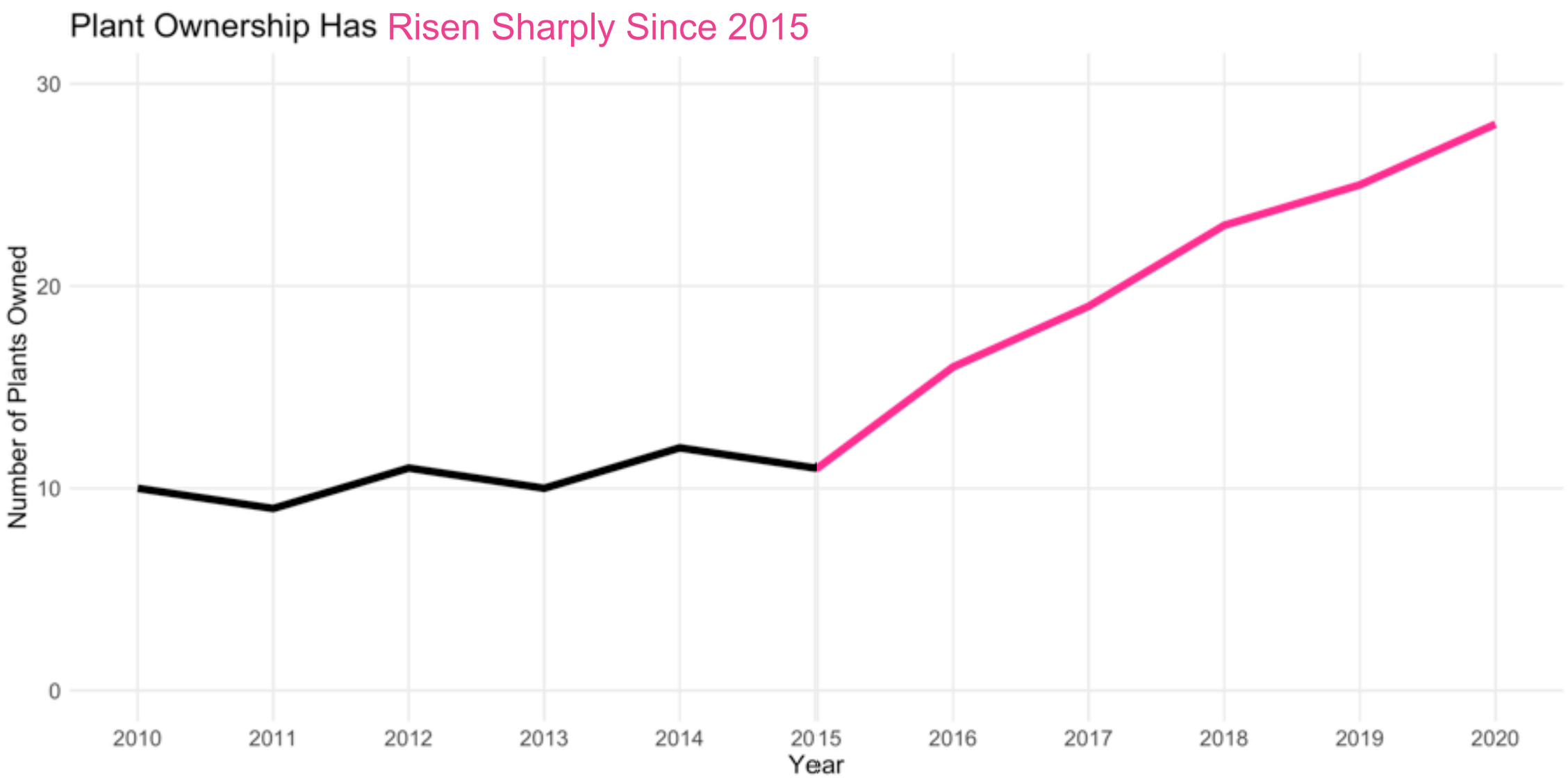
3. **Use descriptive titles:** *tell* your viewer what they should see, and don't be redundant.



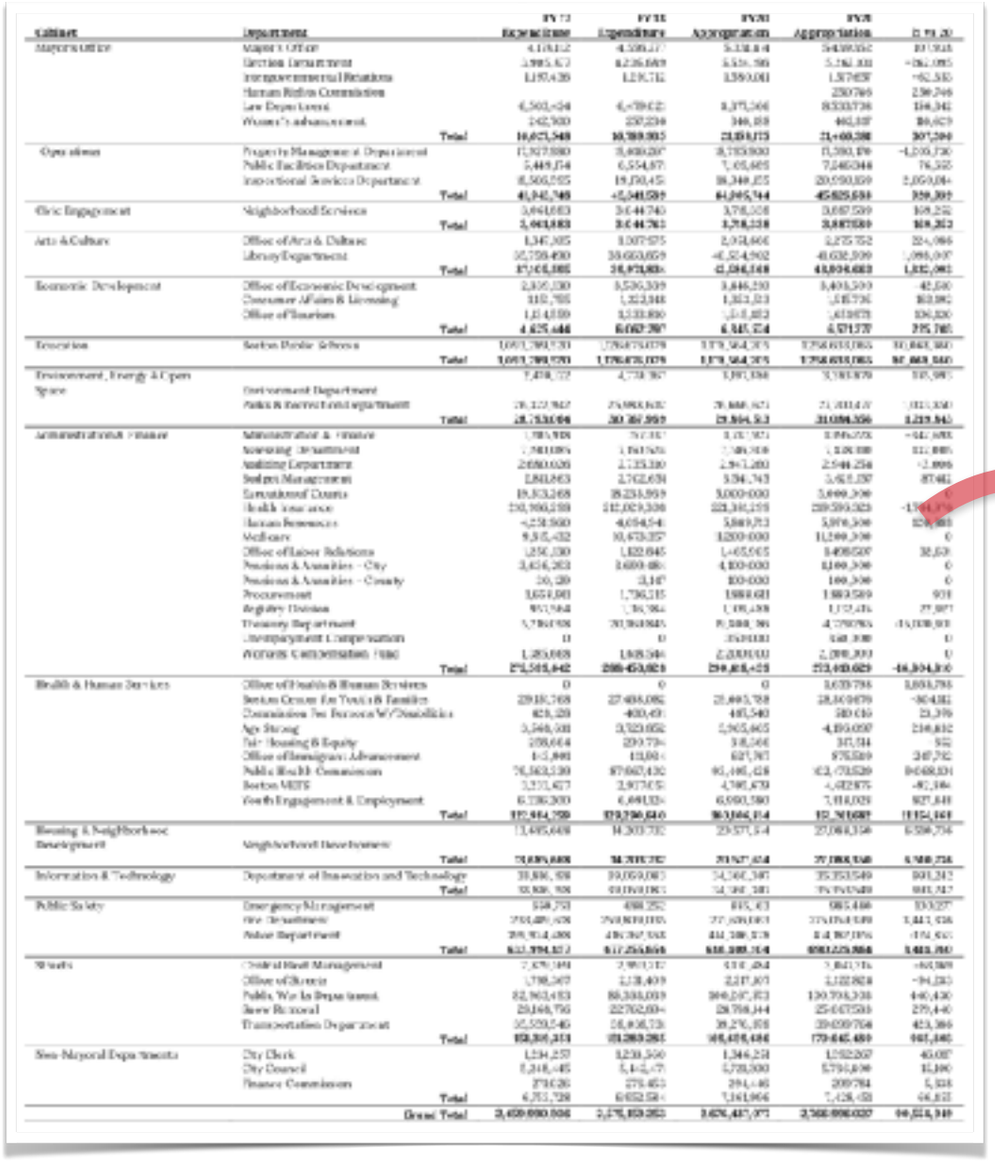
3. **Use descriptive titles:** *tell* your viewer what they should see, and don't be redundant.

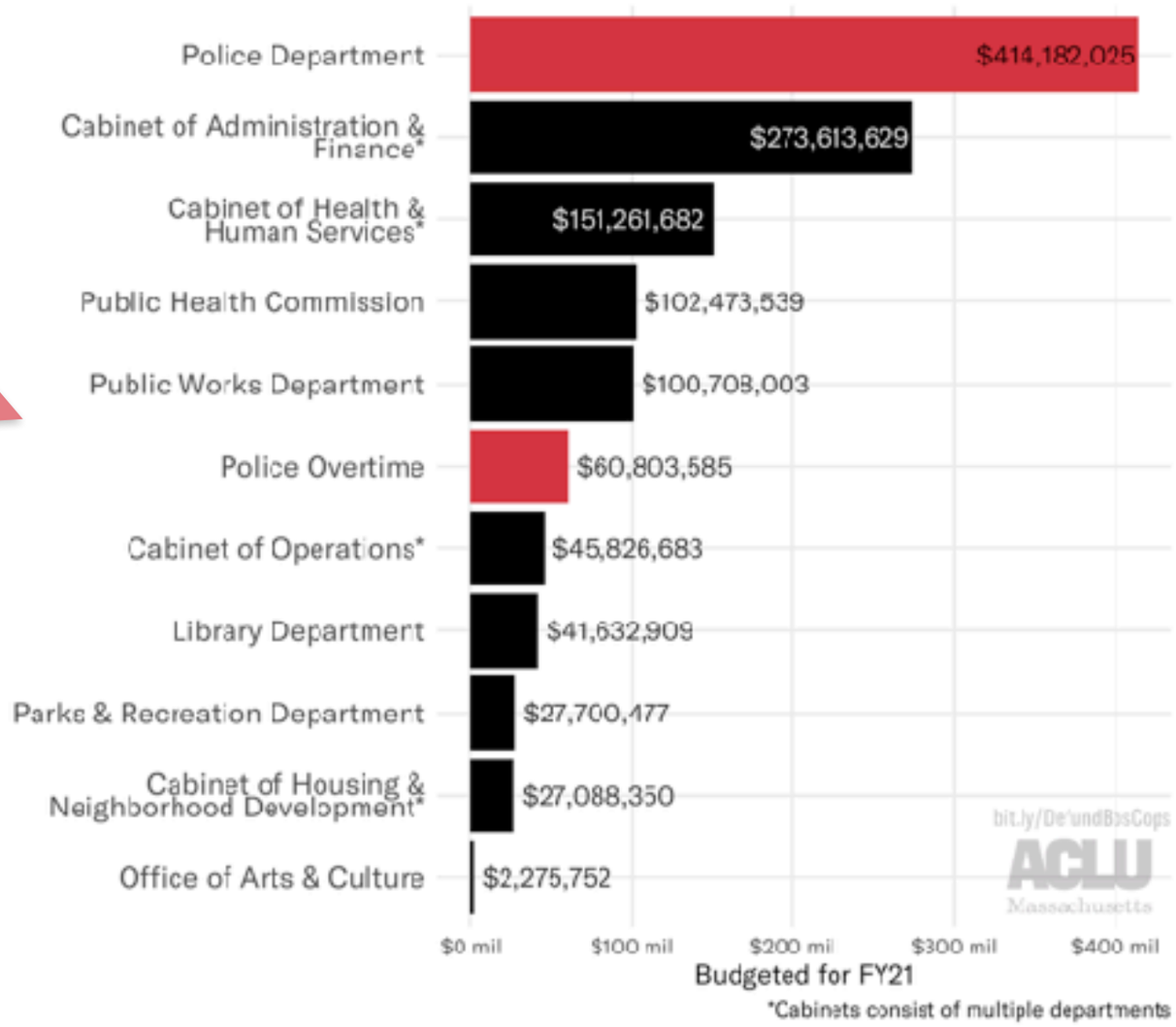


3. **Use descriptive titles:** *tell* your viewer what they should see, and don't be redundant.



4. **Be selective in data and design.** Consider your data-to-ink ratio and remove distractions.

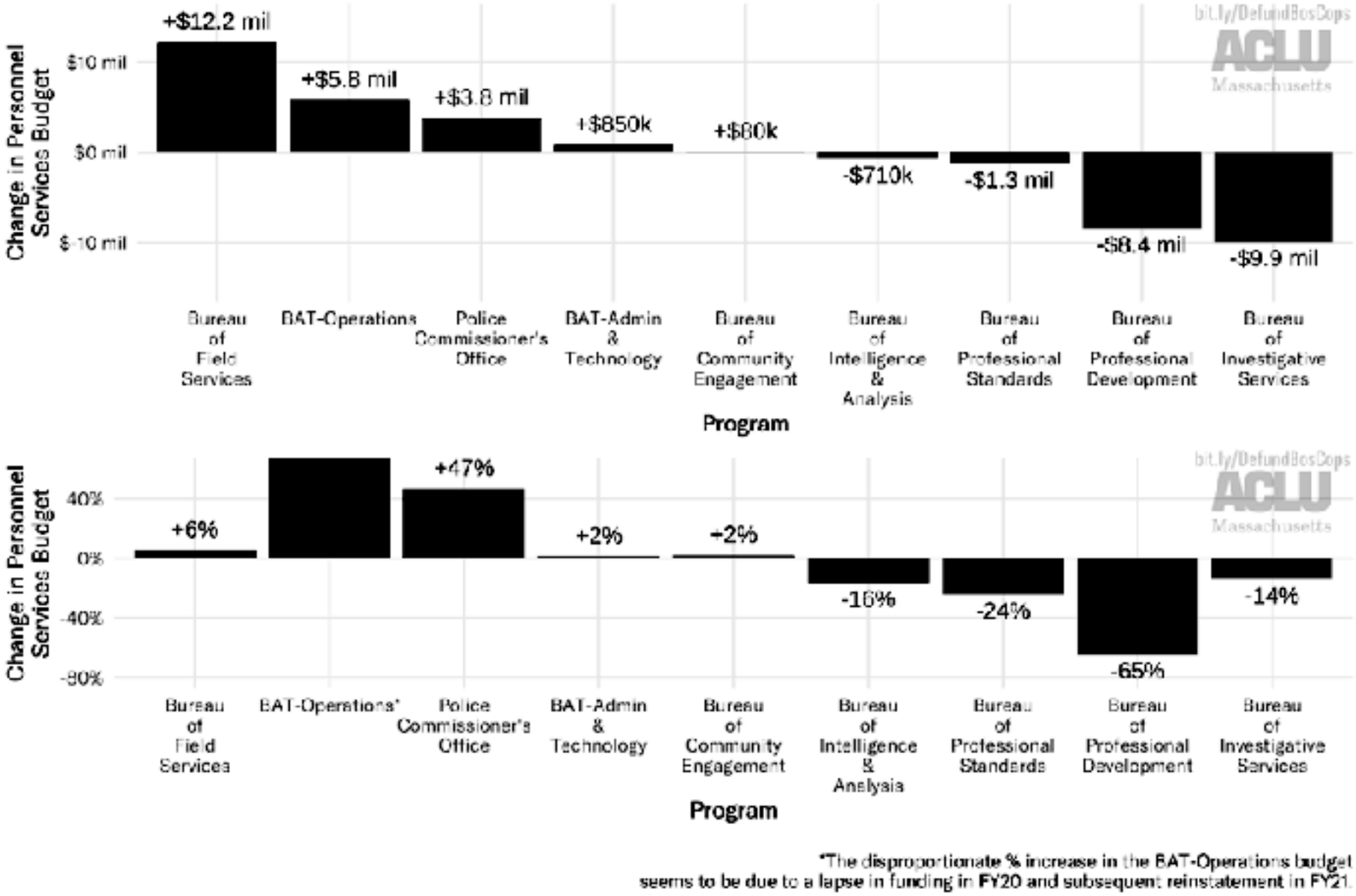




Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

5. Use color sparingly, intentionally, and consistently.

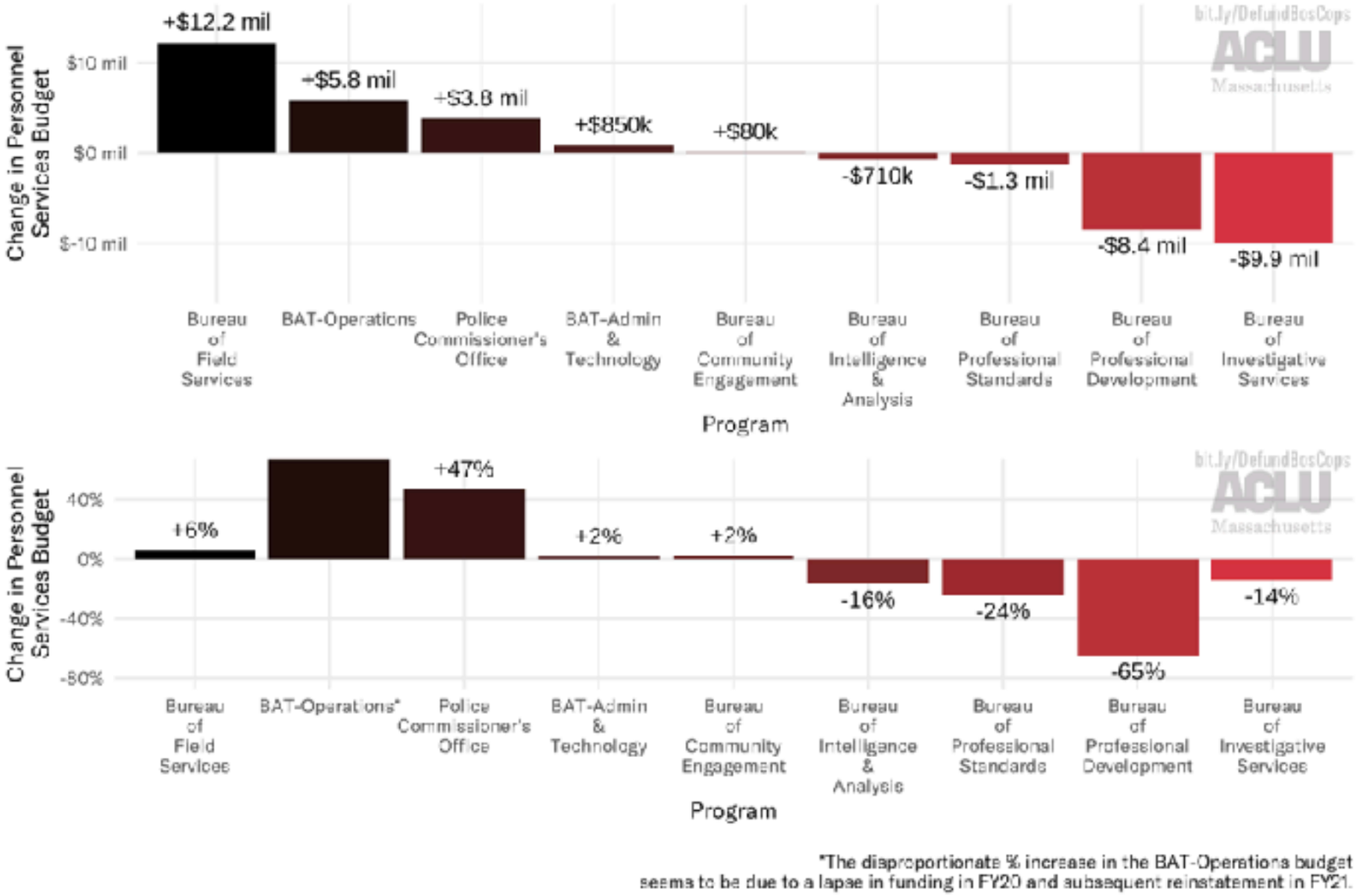
- Keep branding in mind
- Remember colorblindness



Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

5. Use color sparingly, intentionally, and consistently.

- Keep branding in mind
- Remember colorblindness



Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

5. Use color sparingly, intentionally, and consistently.

- Keep branding in mind
- Remember colorblindness

STScI | SPACE TELESCOPE SCIENCE INSTITUTE

OSWALD

Overpass

ACLU Massachusetts

GT America

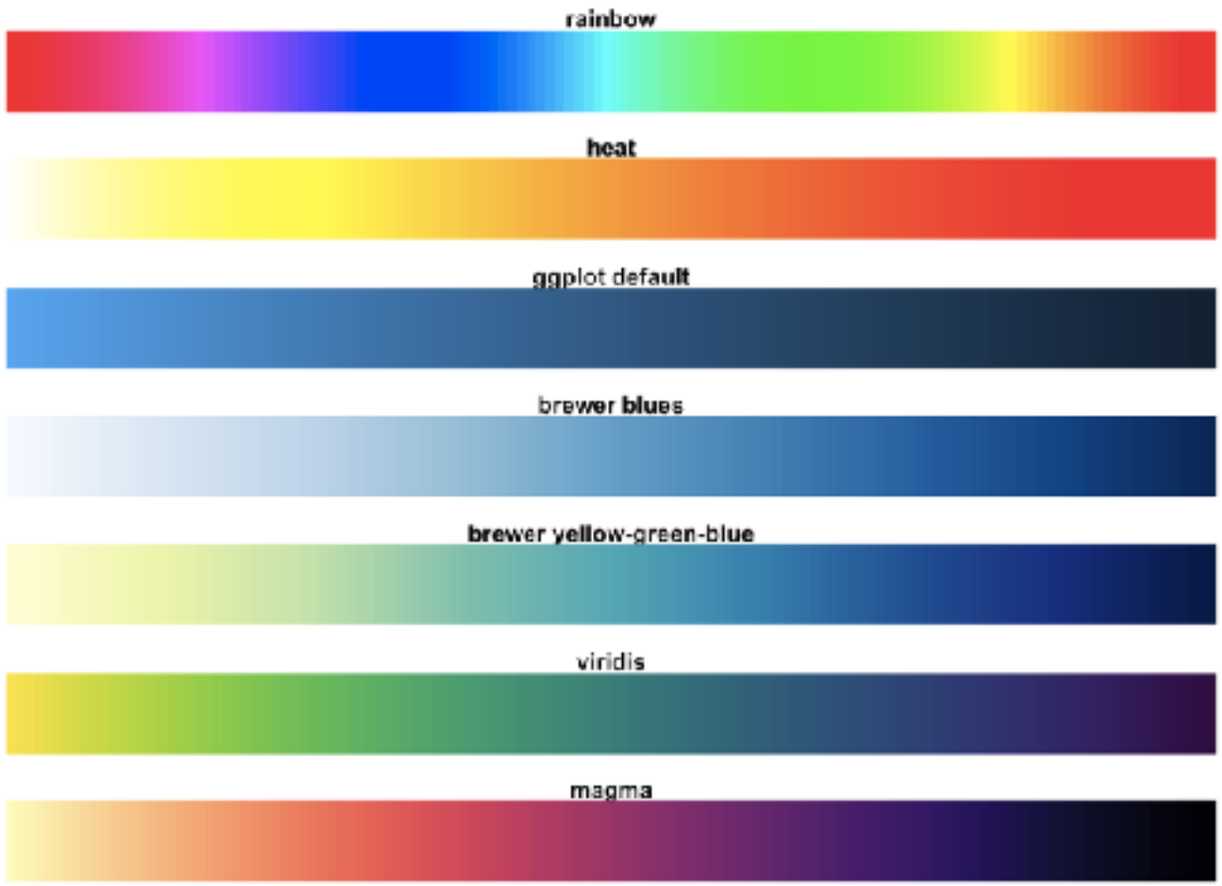
Century Schoolbook

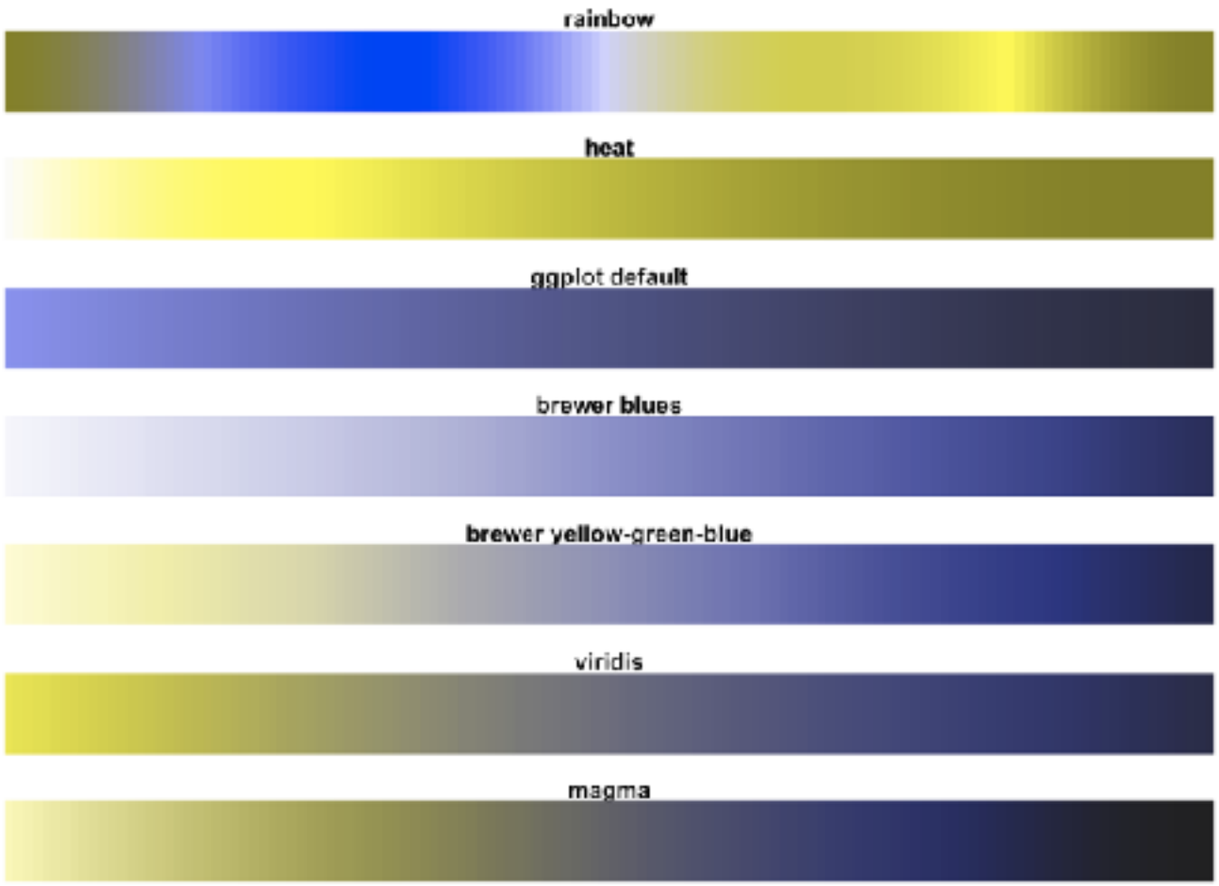
Source Serif

Inter

5. Use color sparingly, intentionally, and consistently.

- Keep branding in mind
- Remember colorblindness



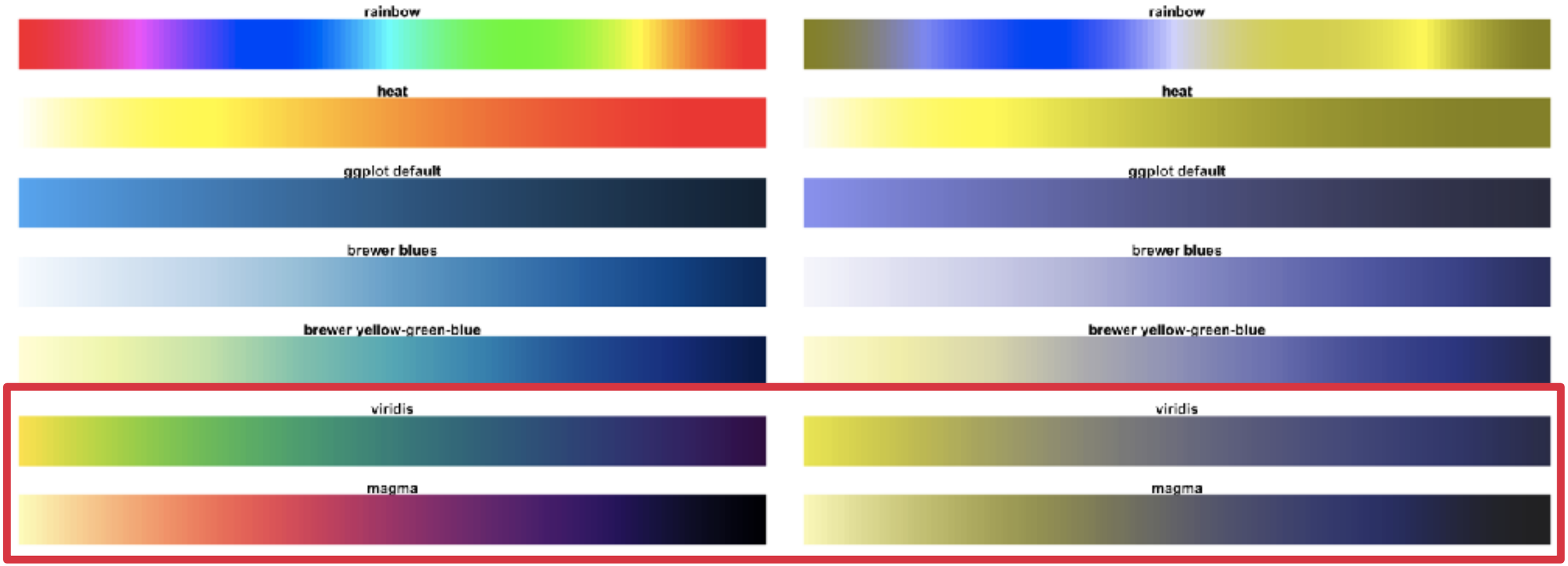


Deuteranopia

Source: "Introduction to the viridis color maps," Bob Rudis, Noam Ross and Simon Garnier, CRAN

5. Use color sparingly, intentionally, and consistently.

- Keep branding in mind
- Remember colorblindness

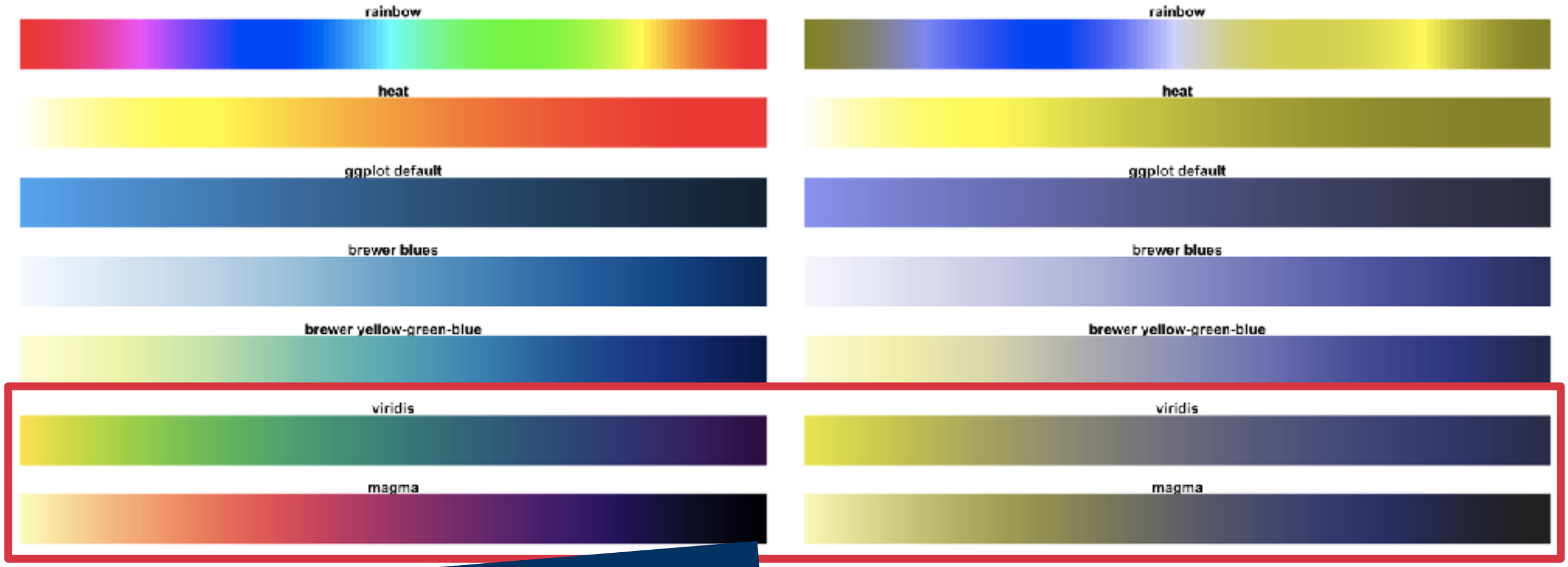


Deuteranopia

Source: "Introduction to the viridis color maps," Bob Rudis, Noam Ross and Simon Garnier, CRAN

5. Use color sparingly, intentionally, and consistently.

- Keep branding in mind
- Remember colorblindness

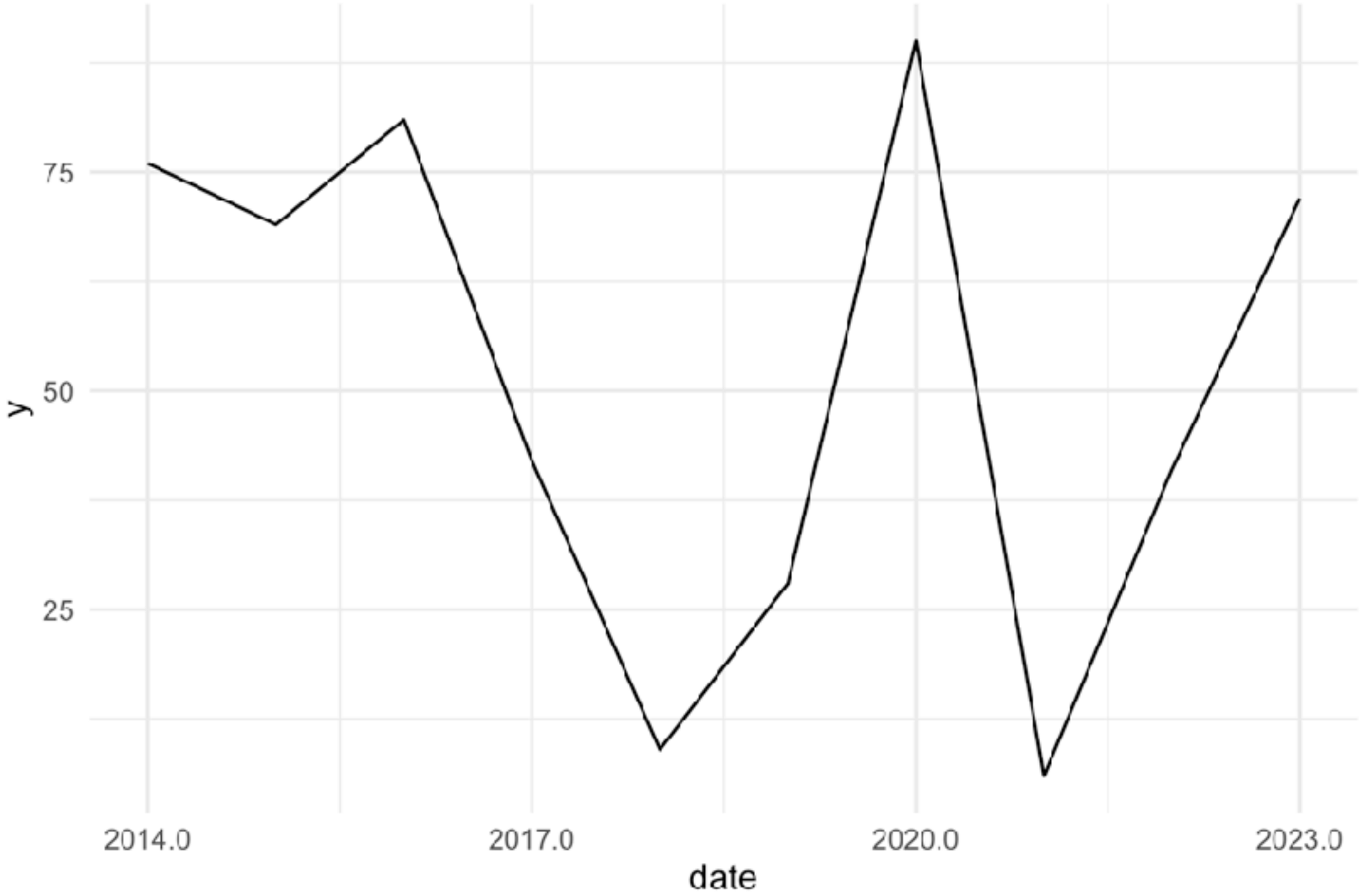


Paul Tol's Palettes

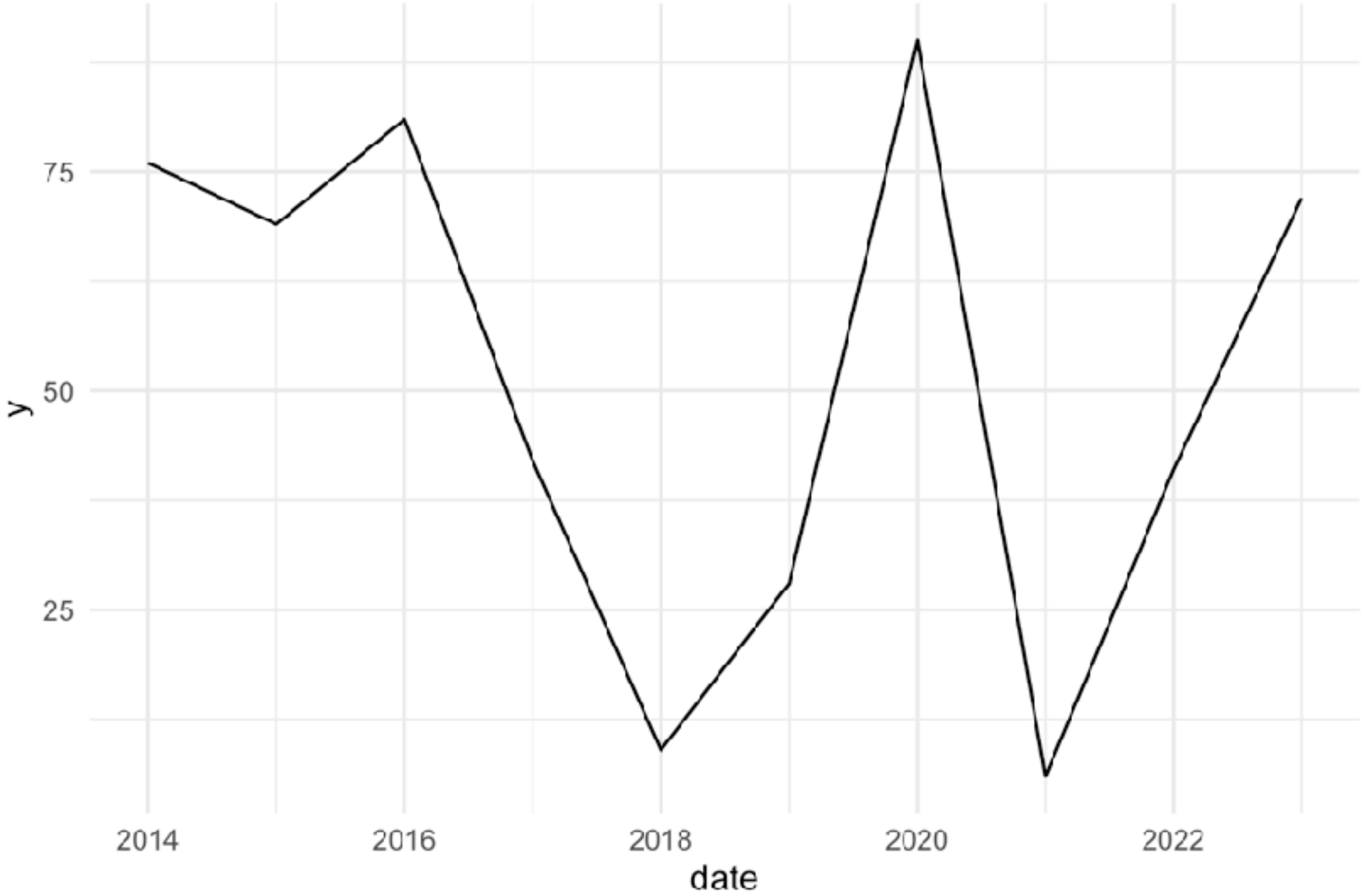
Deuteranopia

Source: "Introduction to the viridis color maps," Bob Rudis, Noam Ross and Simon Garnier, CRAN

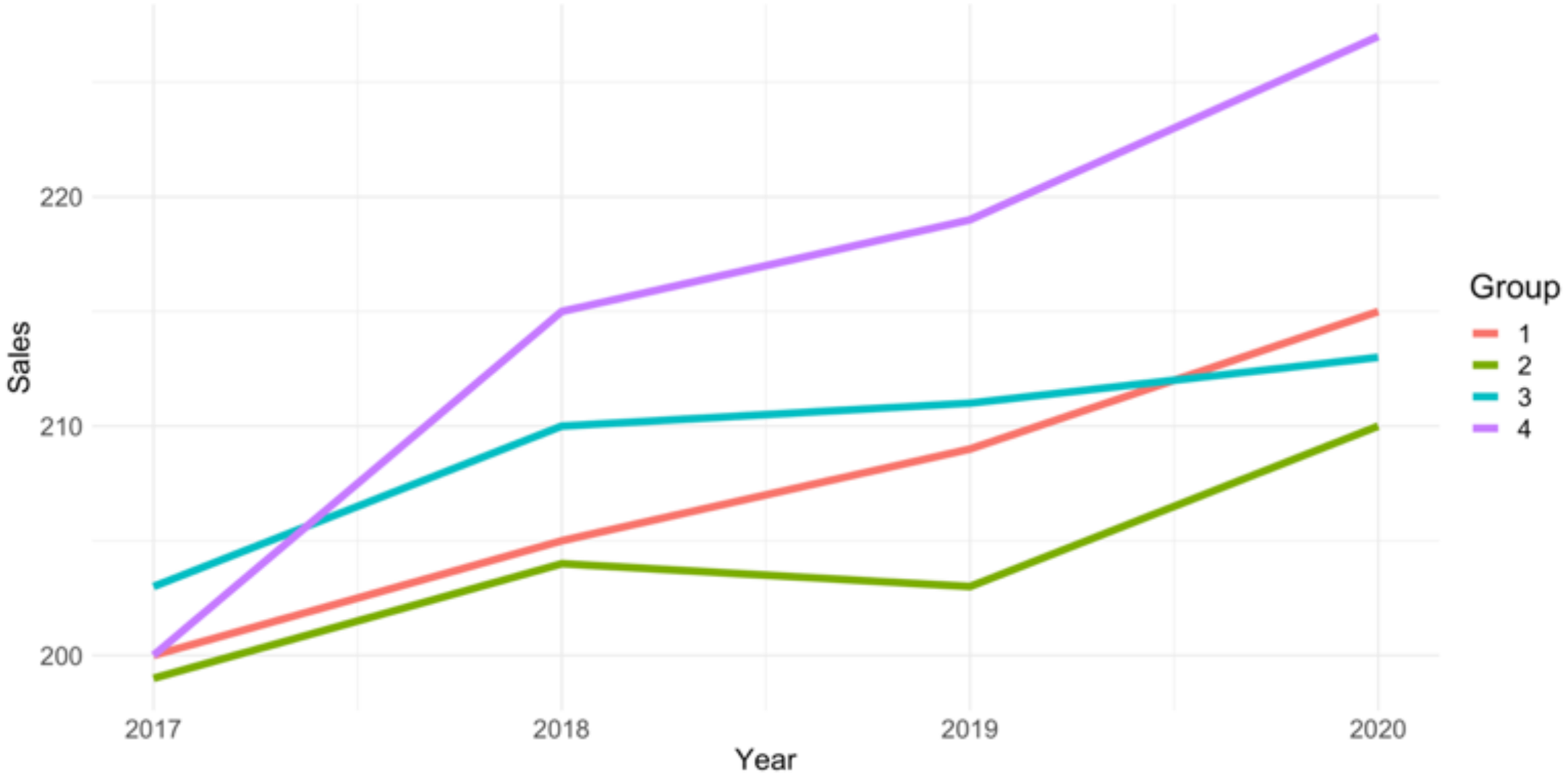
6. **Use reasonable axis tick** formats and frequencies, not just the default.



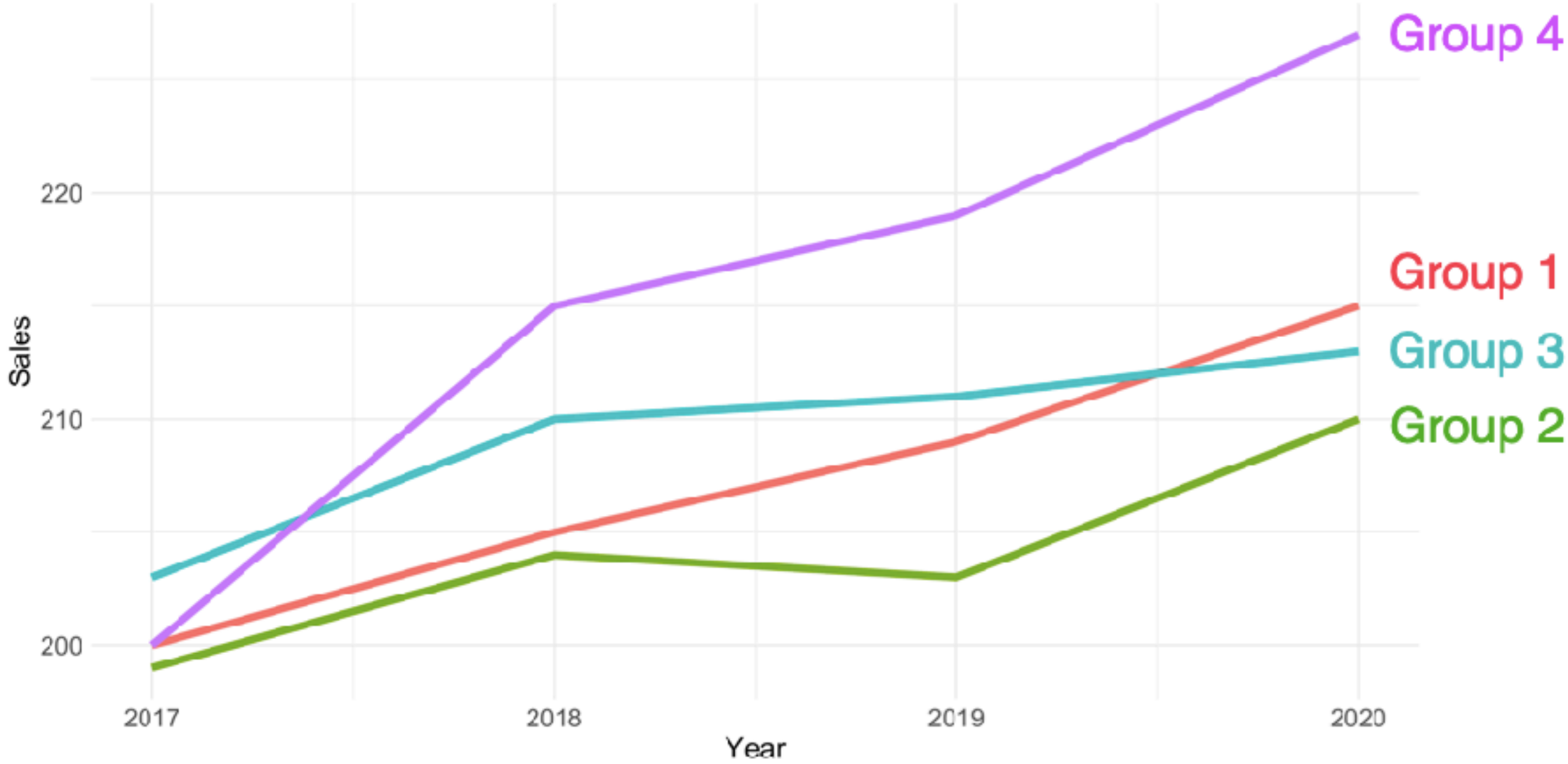
6. **Use reasonable axis tick** formats and frequencies, not just the default.



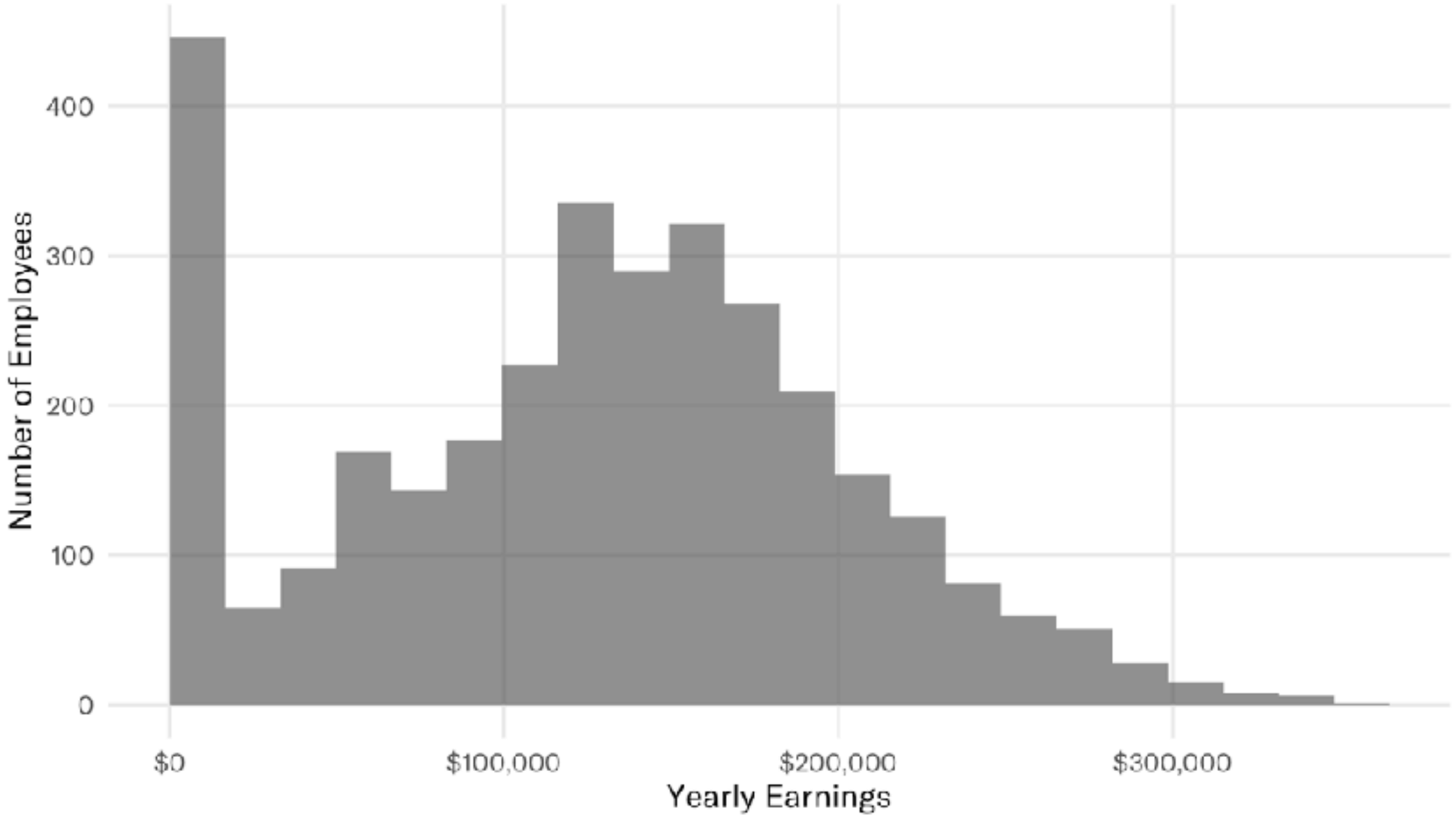
7. Consider creative alternatives to legends.



7. Consider creative alternatives to legends.

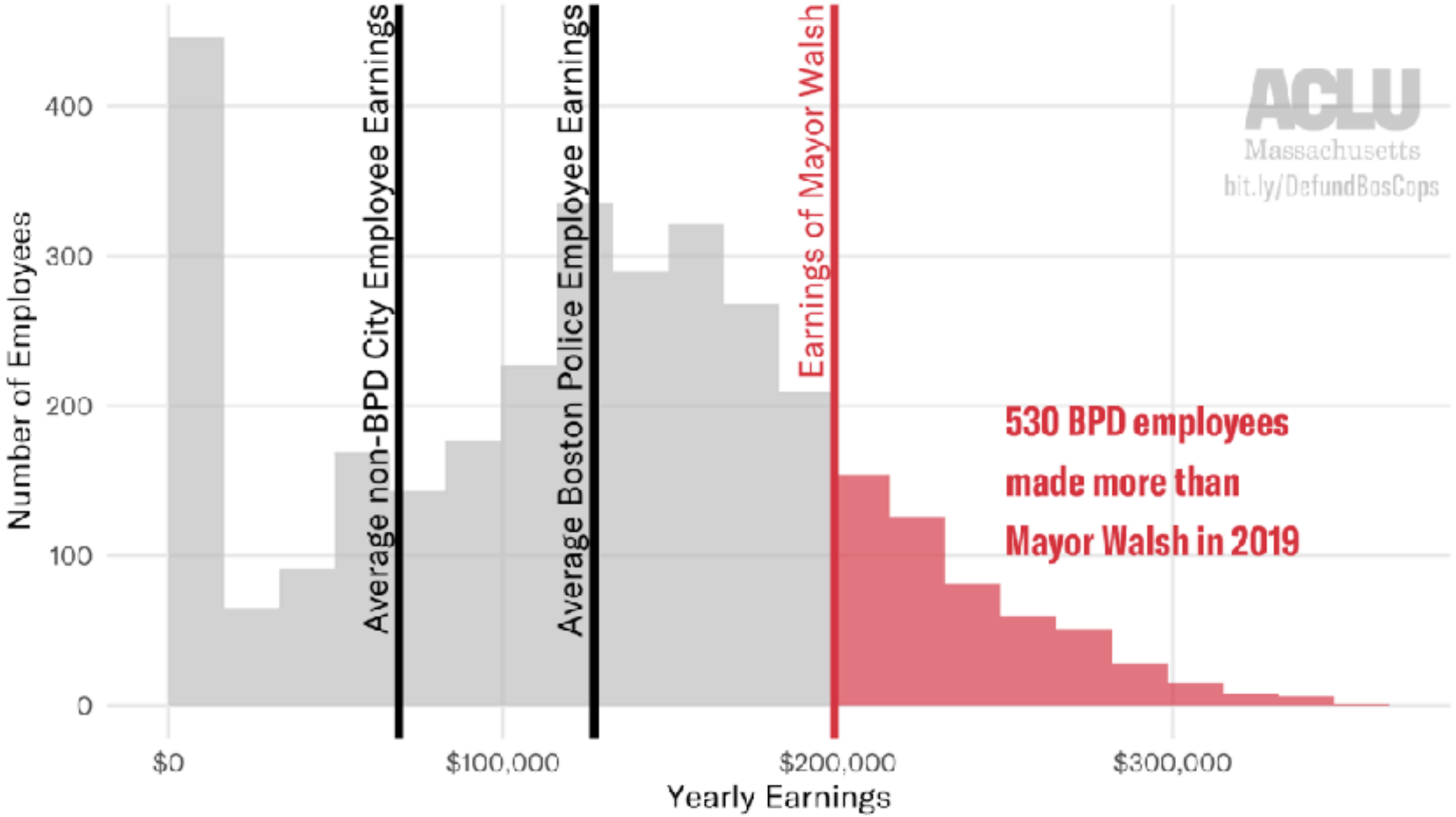


8. **Don't be afraid of annotations.** Use whatever software you prefer to add them.



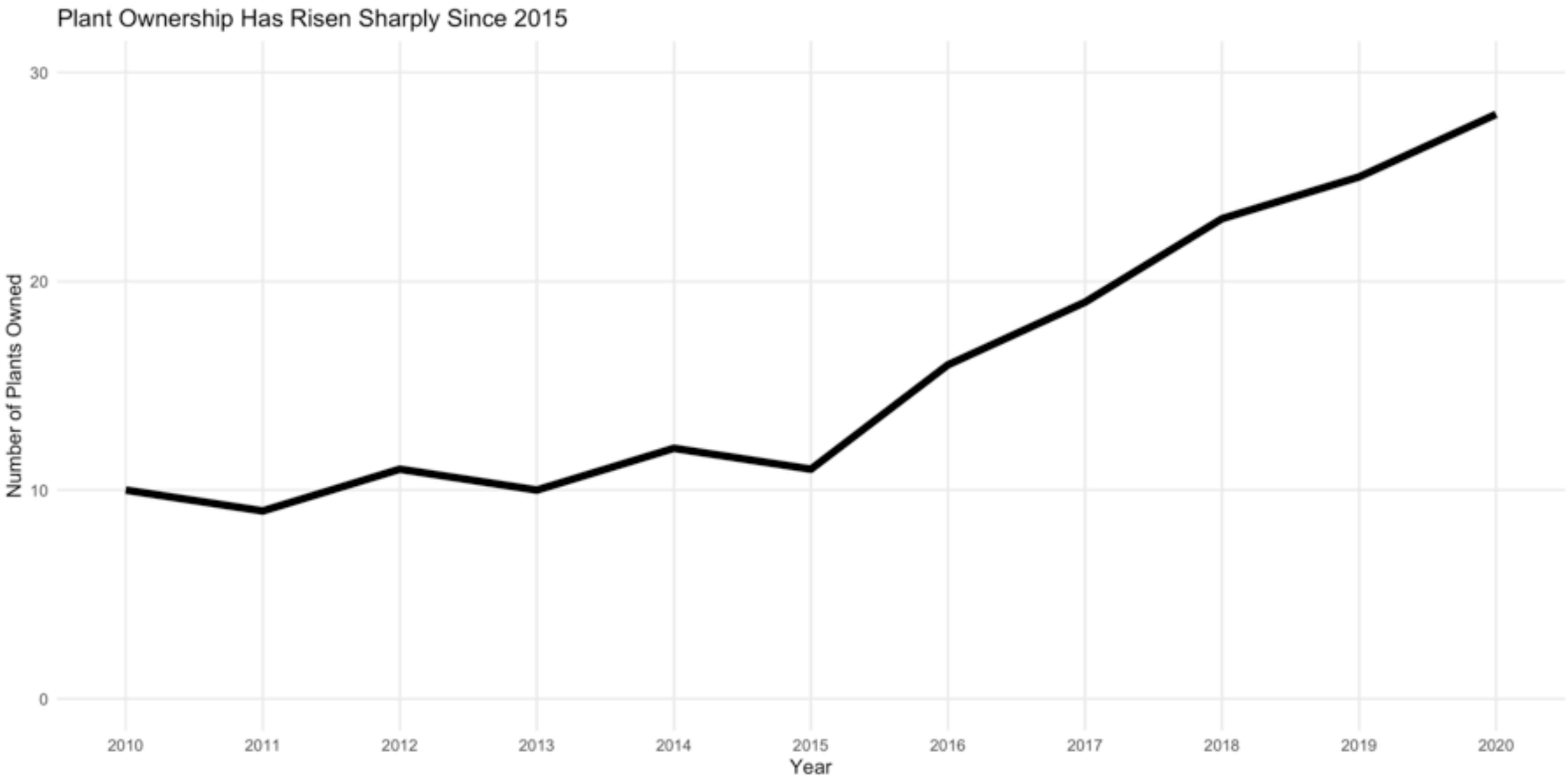
Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

8. **Don't be afraid of annotations.** Use whatever software you prefer to add them.

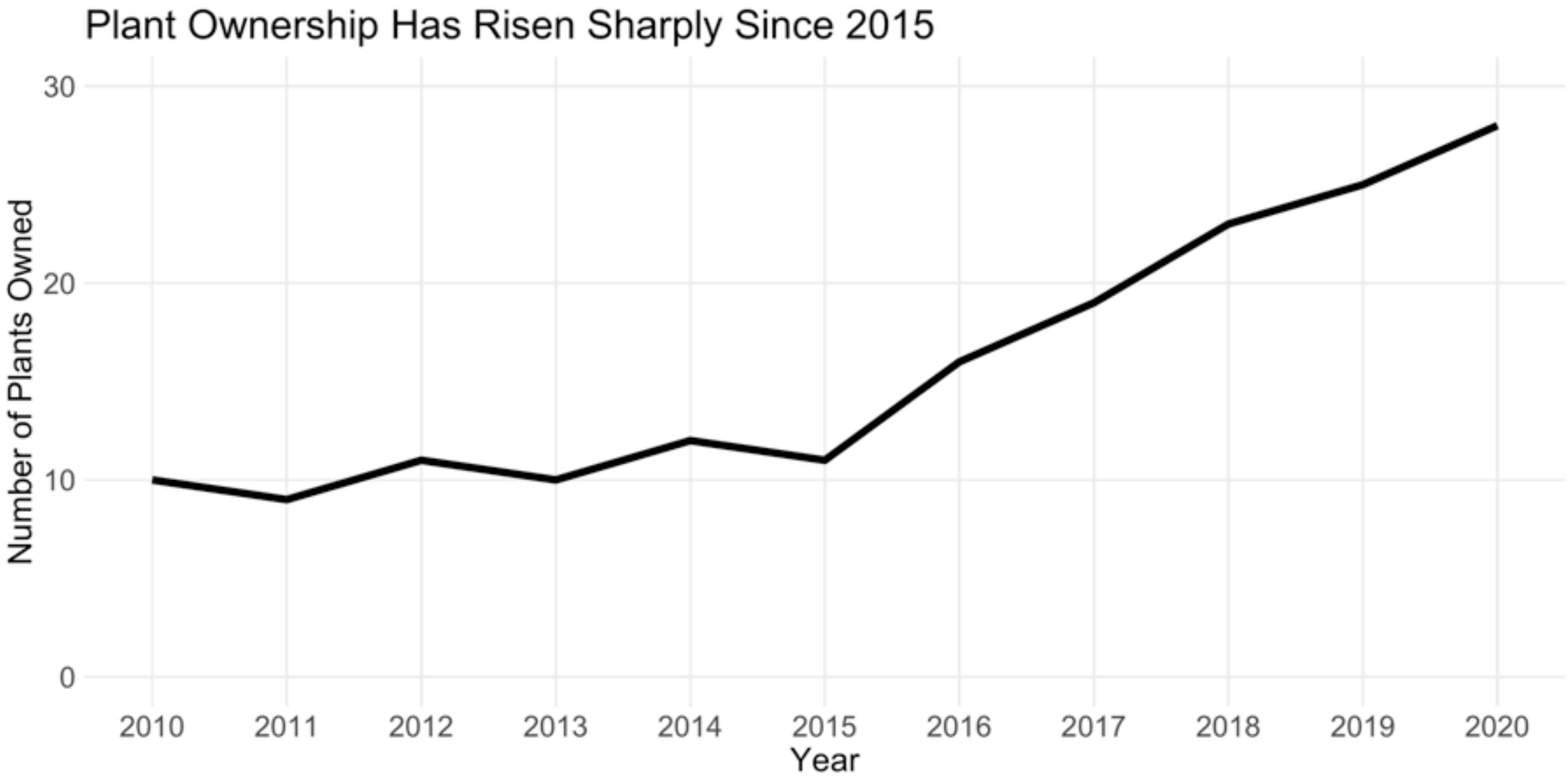


Source: "Unpacking the Boston Police Budget," ACLU of Massachusetts Data for Justice Project

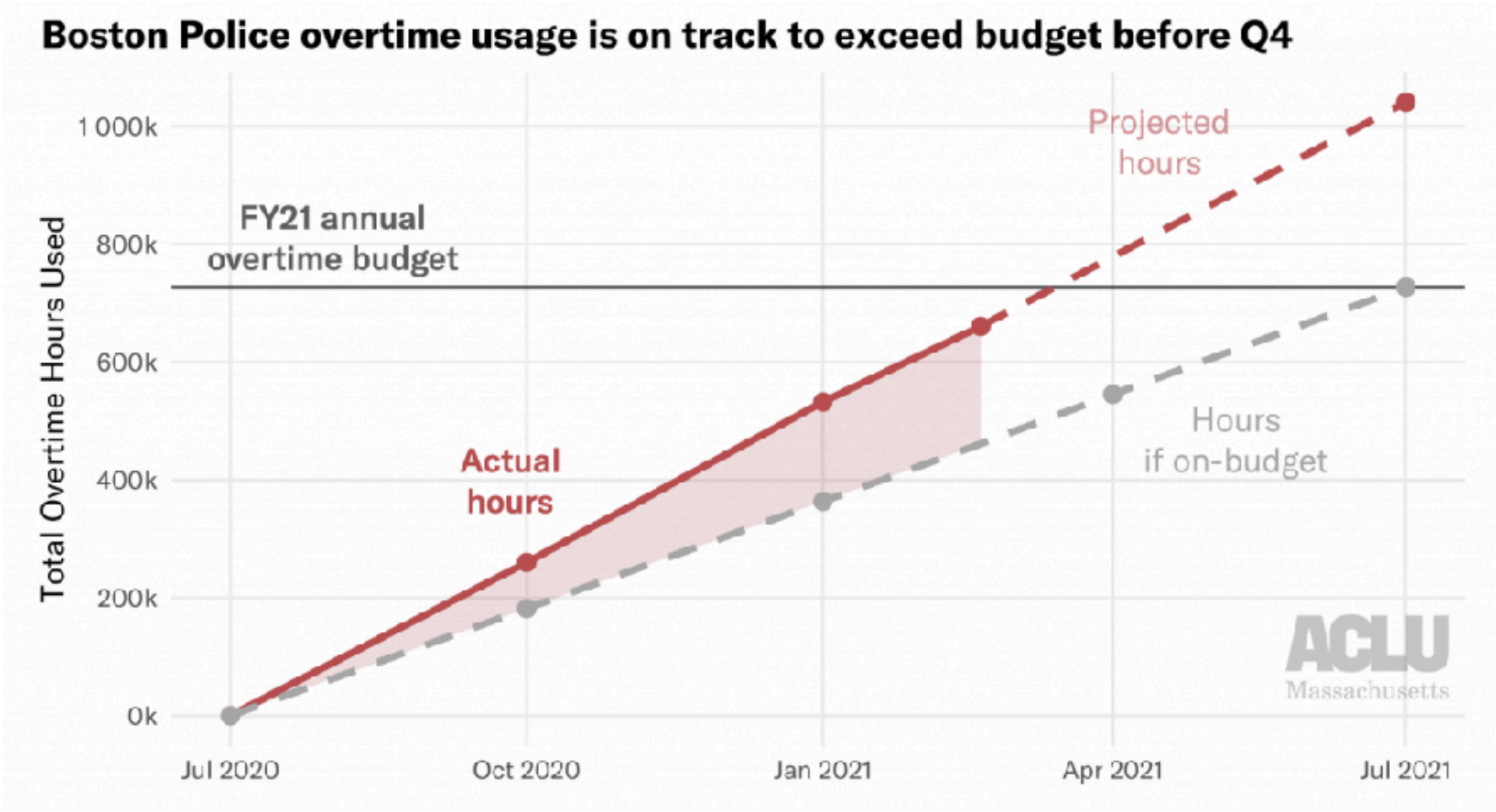
9. **Make your text bigger** than you'd think. Nobody likes to squint!



9. Make your text bigger than you'd think. Nobody likes to squint!

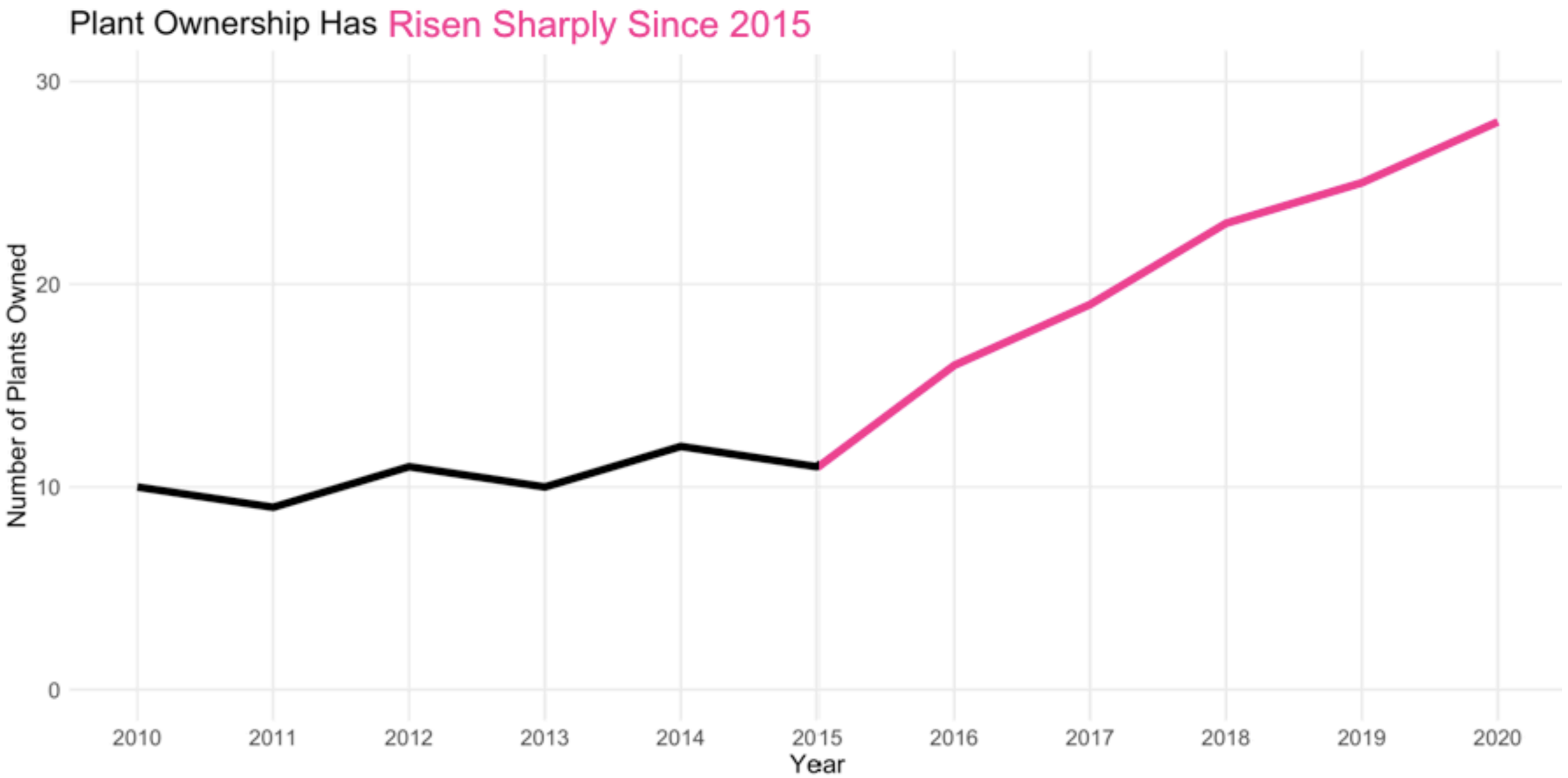


10. Use the blink test.



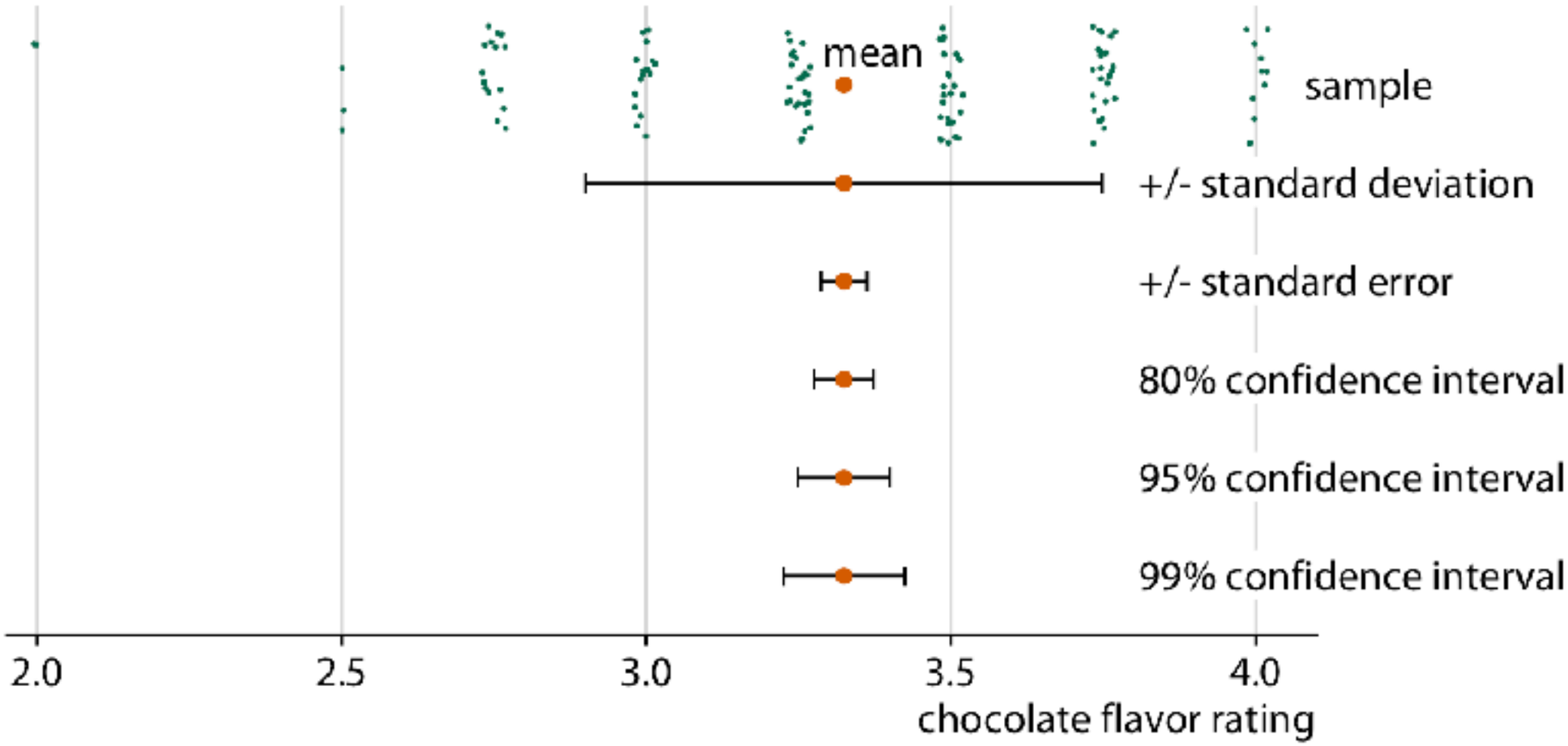
Source: "A Mayor's Roadmap to Curb Boston Police Overtime," ACLU of Massachusetts Data for Justice Project

10. Use the blink test.



Bonus: Visualizing error

- Understand the difference between standard deviation, standard error, and confidence intervals. Use the appropriate error! (And know when there is no need to visualize error.)



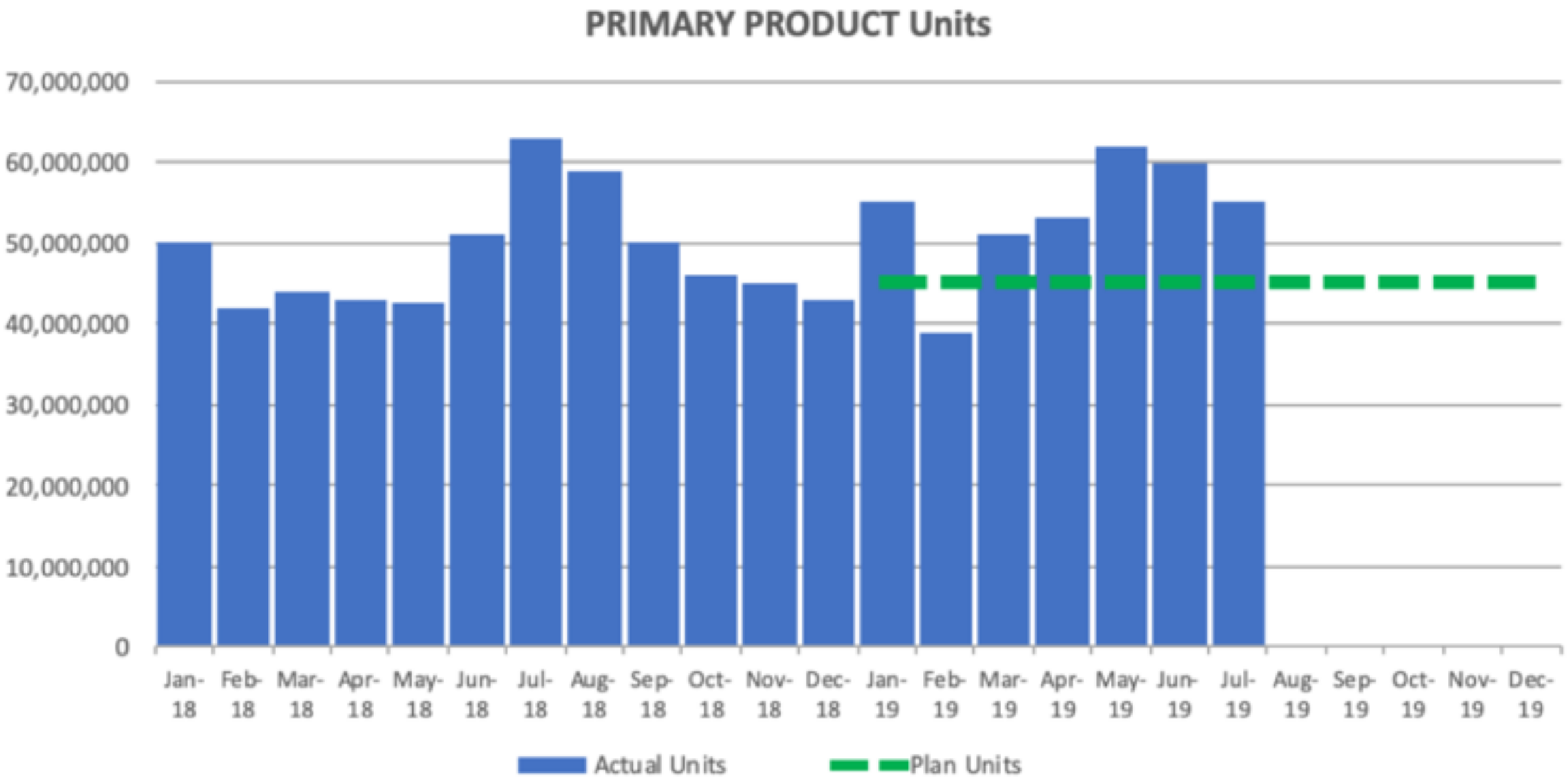
Source: "Visualizing Uncertainty," Claus Wilde

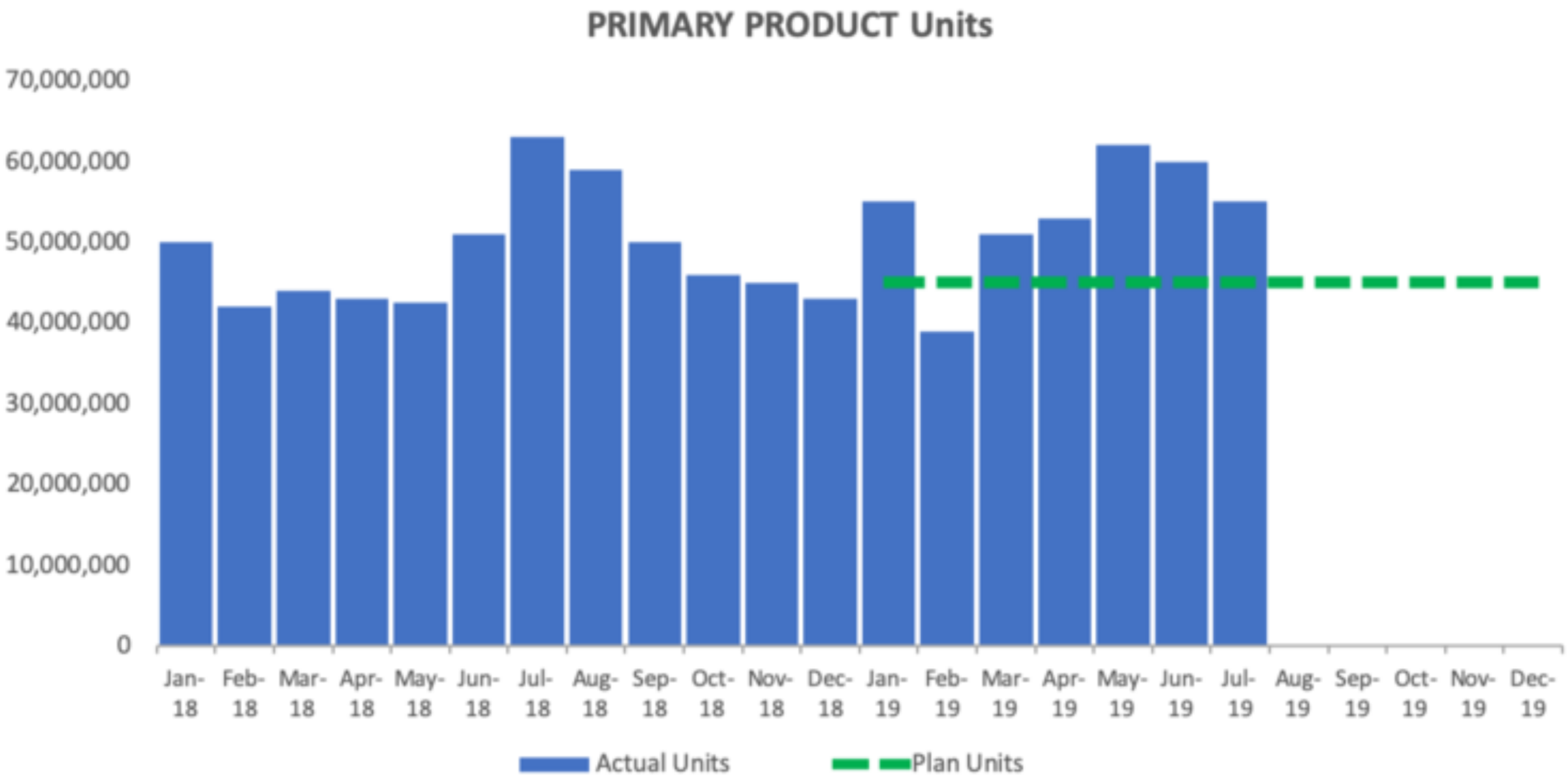
Bonus: Visualizing error

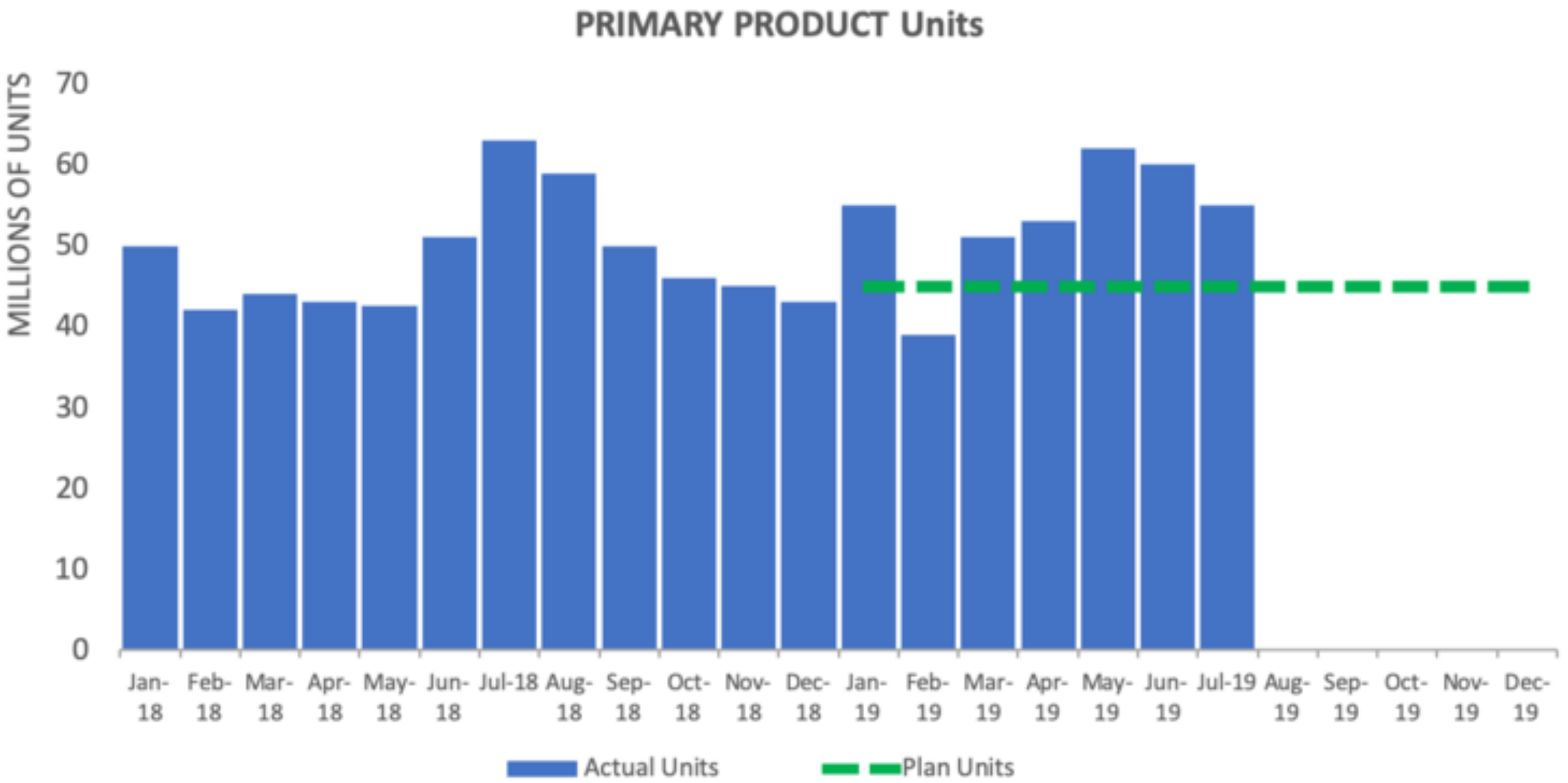
- Understand the difference between standard deviation, standard error, and confidence intervals. Use the appropriate error! (And know when there is no need to visualize error.)
- There are more ways of representing error than just bars. Consider alternative ways of visualizing error like violin plots, box plots, confidence bands...
- Make sure error bars / bands are clearly visible (color, line width, line pattern, caps)
- Always make the meaning of an error bar/band (1 SD, 95% CI, ...) explicit in a caption or annotation
- For more, check out Claus Wilke's guide: https://clauswilke.com/dataviz/visualizing-uncertainty

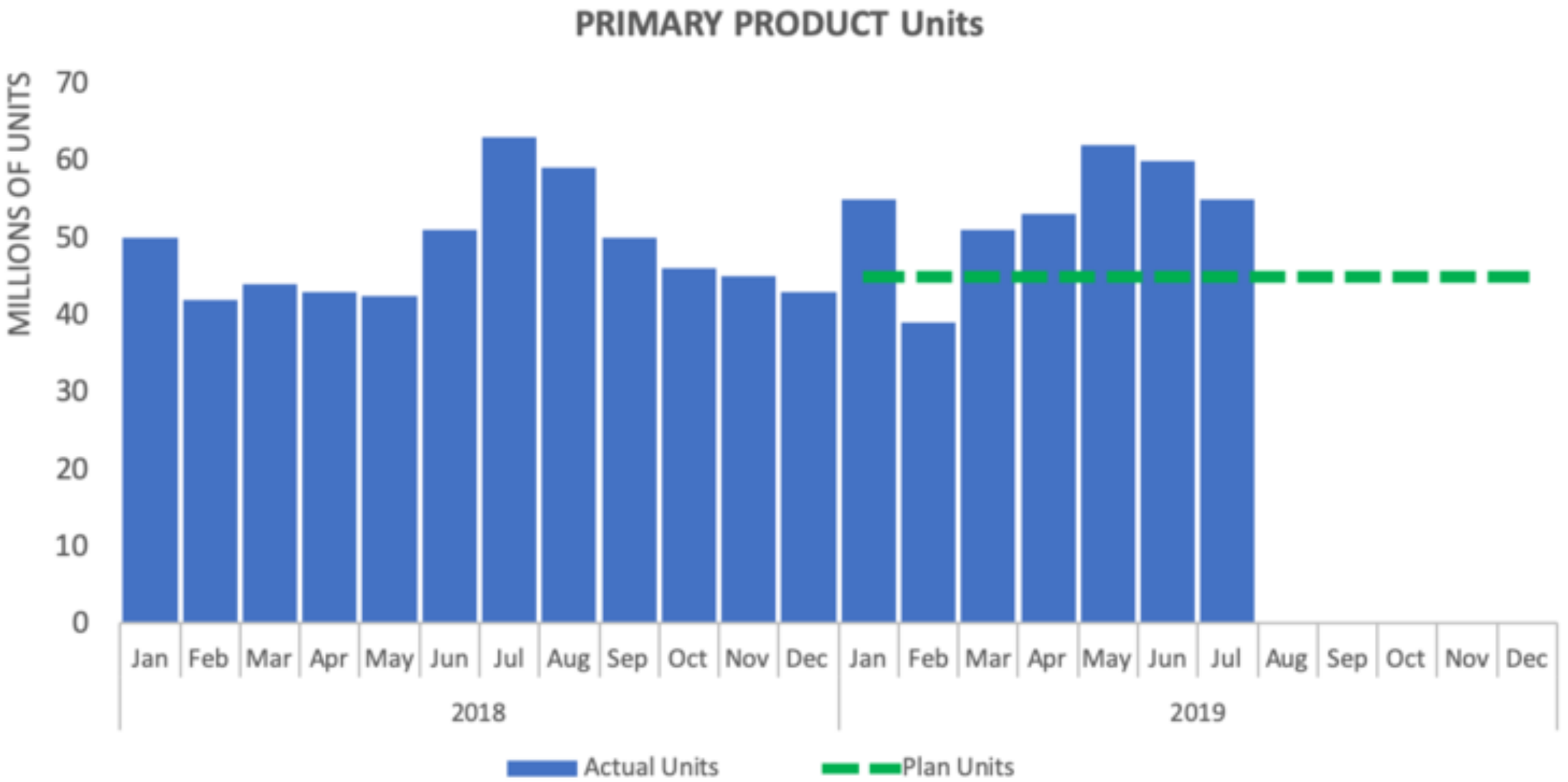
1. **Know your audience!**
   - For public vizzes, this usually requires bar, line, and scatter plots.
   - Basically never use undefined acronyms or symbols.
2. **Question the defaults.** This means fonts, colors, sizes, shapes, and more.
3. **Use descriptive titles:** *tell* your viewer what they should see, and don't be redundant.
4. **Be selective in data and design.** Consider your data-to-ink ratio and remove distractions.
5. **Use color sparingly, intentionally, and consistently.**
   - Keep branding in mind
   - Remember colorblindness
6. **Use reasonable axis tick** formats and frequencies, not just the default.
7. **Consider creative alternatives to legends.**
8. **Don't be afraid of annotations.** Use whatever software you prefer to add them.
9. **Make your text bigger** than you'd think. Nobody likes to squint!
10. **Use the blink test.**

Improve! This! Viz!

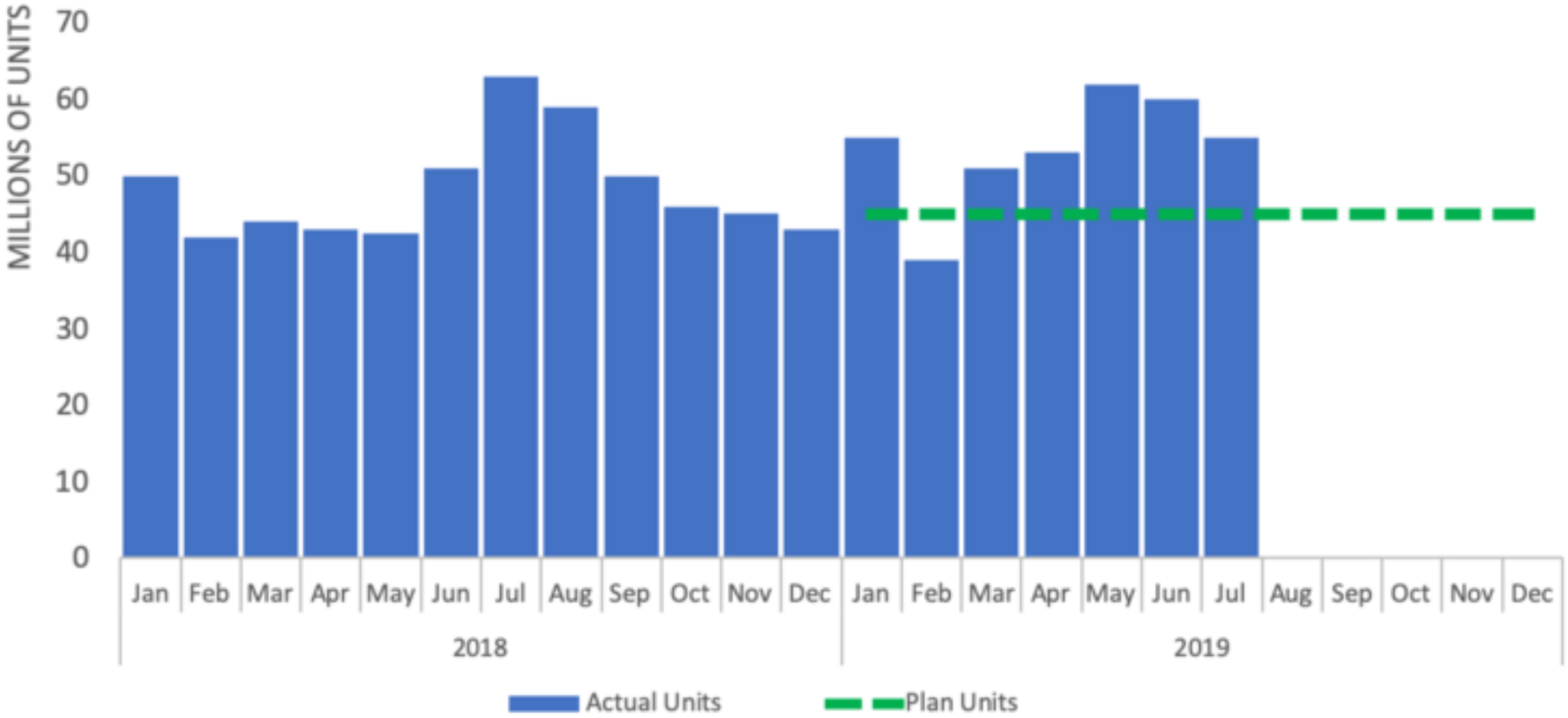

PRIMARY PRODUCT Units
70,000,000
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Actual Units
Plan Units


PRIMARY PRODUCT Units
70,000,000
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Actual Units
Plan Units

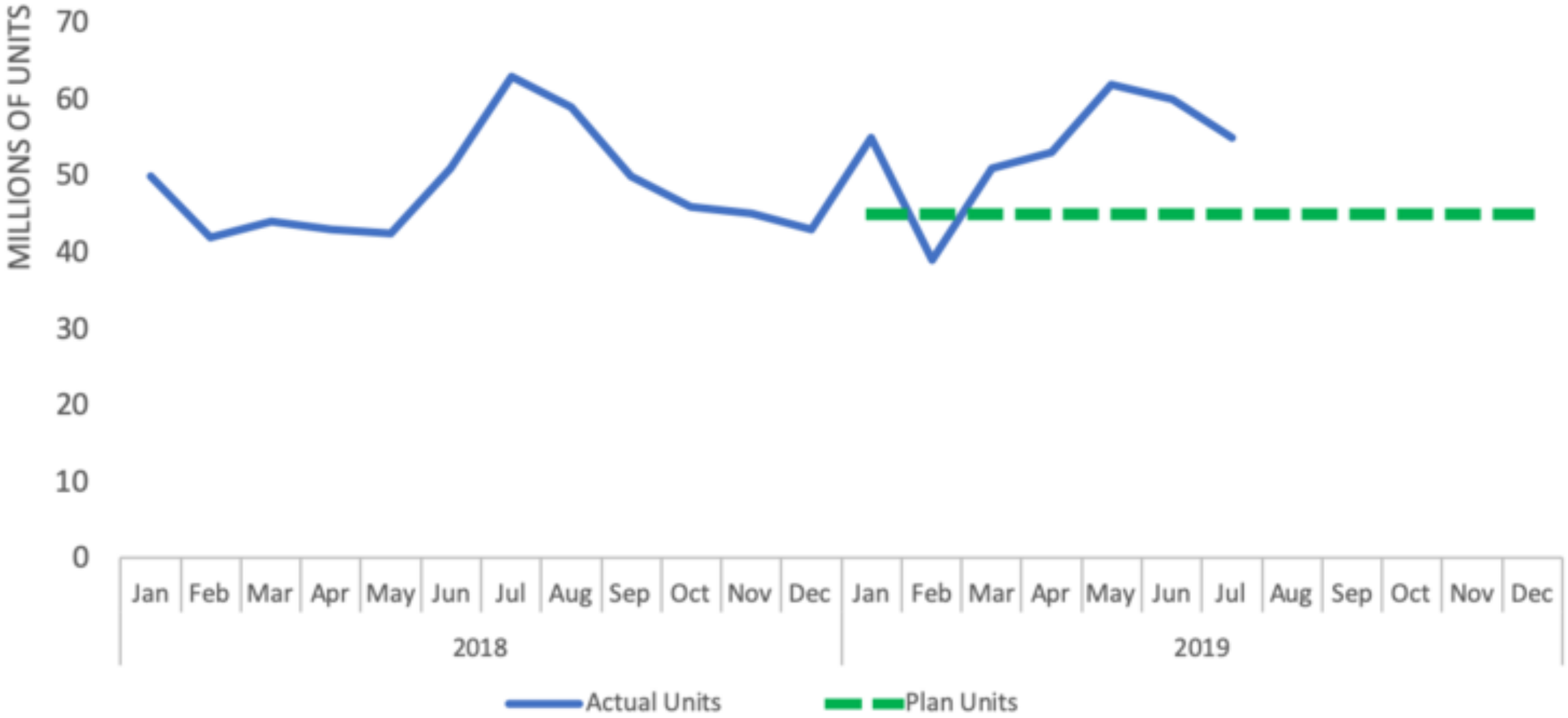

PRIMARY PRODUCT Units
MILLIONS OF UNITS
70
60
50
40
30
20
10
0
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Actual Units
Plan Units


PRIMARY PRODUCT Units
MILLIONS OF UNITS
70
60
50
40
30
20
10
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2018
2019
Actual Units
Plan Units

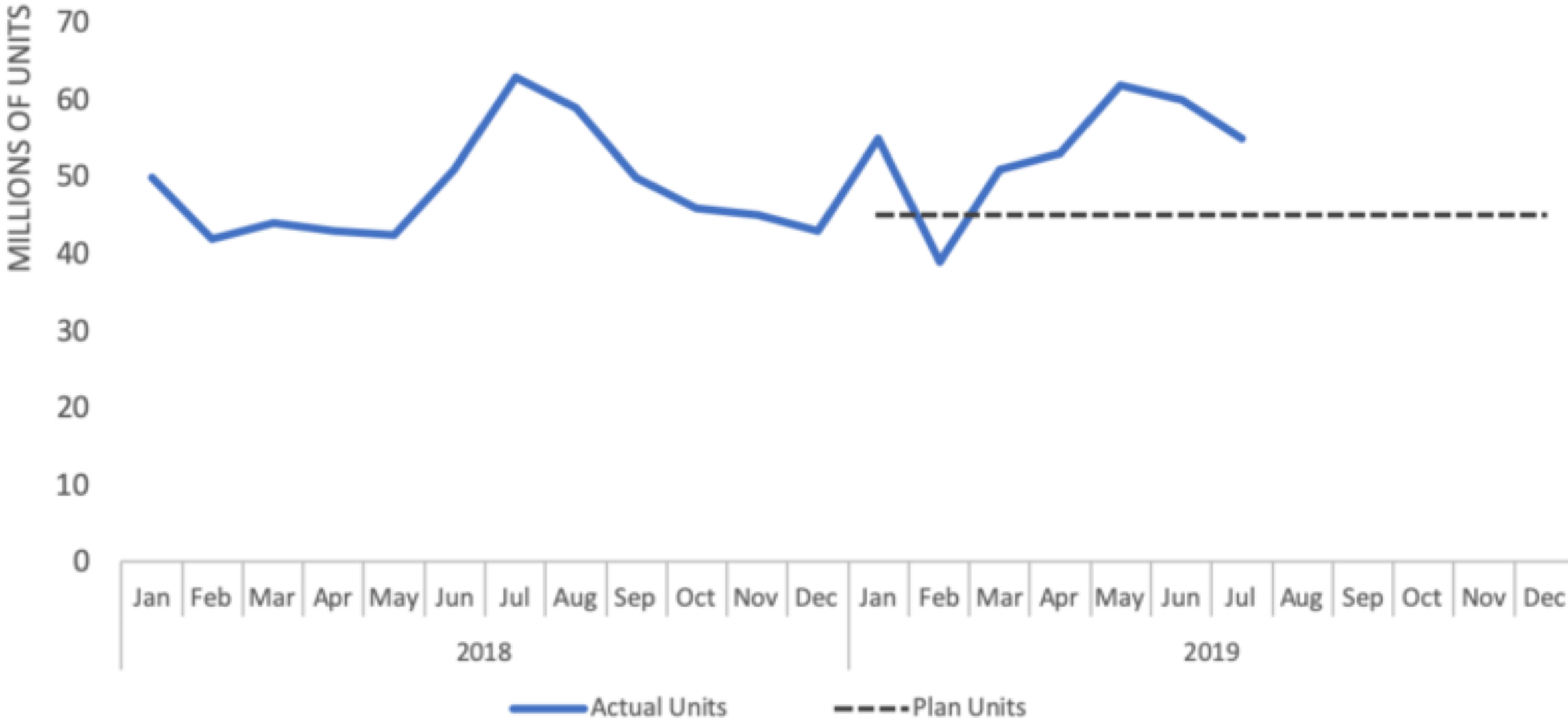
PRIMARY PRODUCT: unit sales



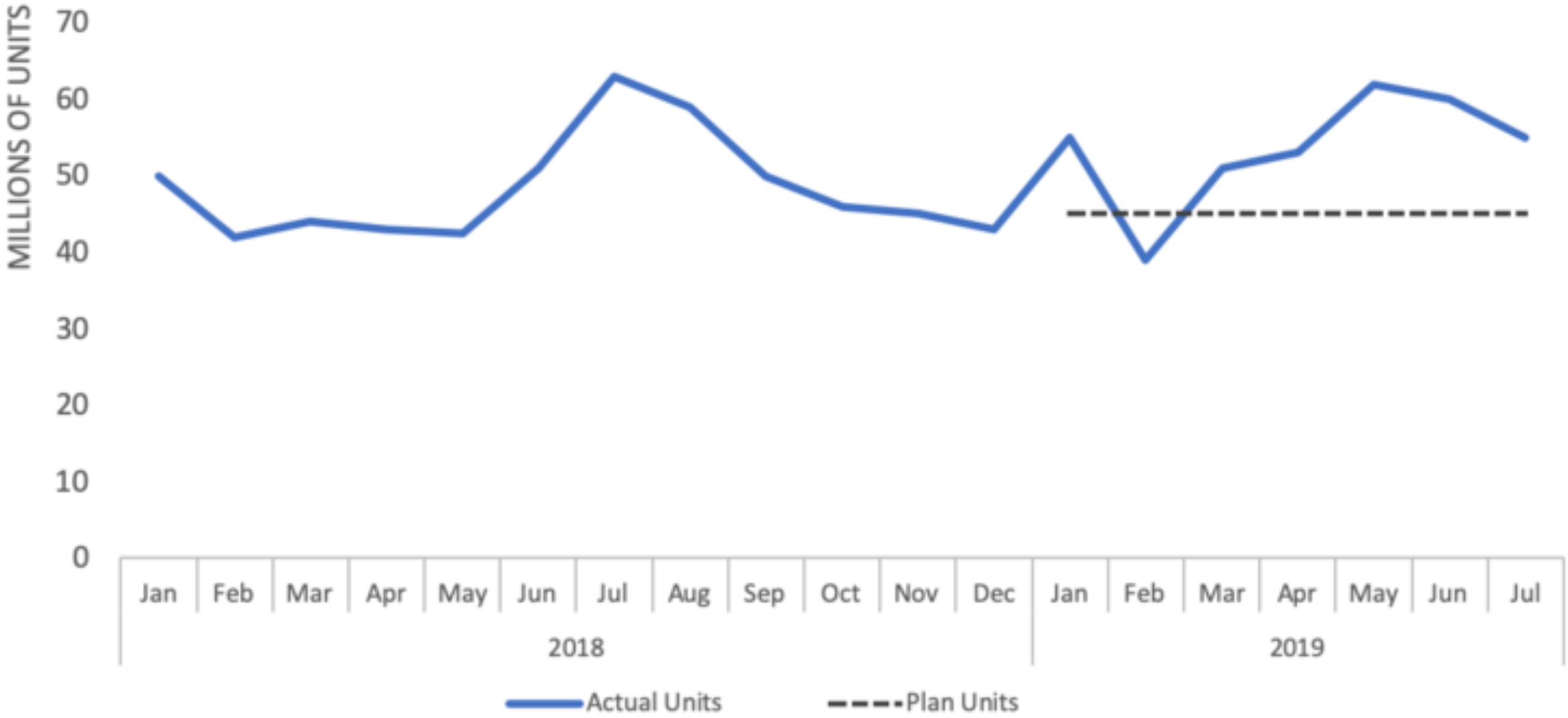
PRIMARY PRODUCT: unit sales



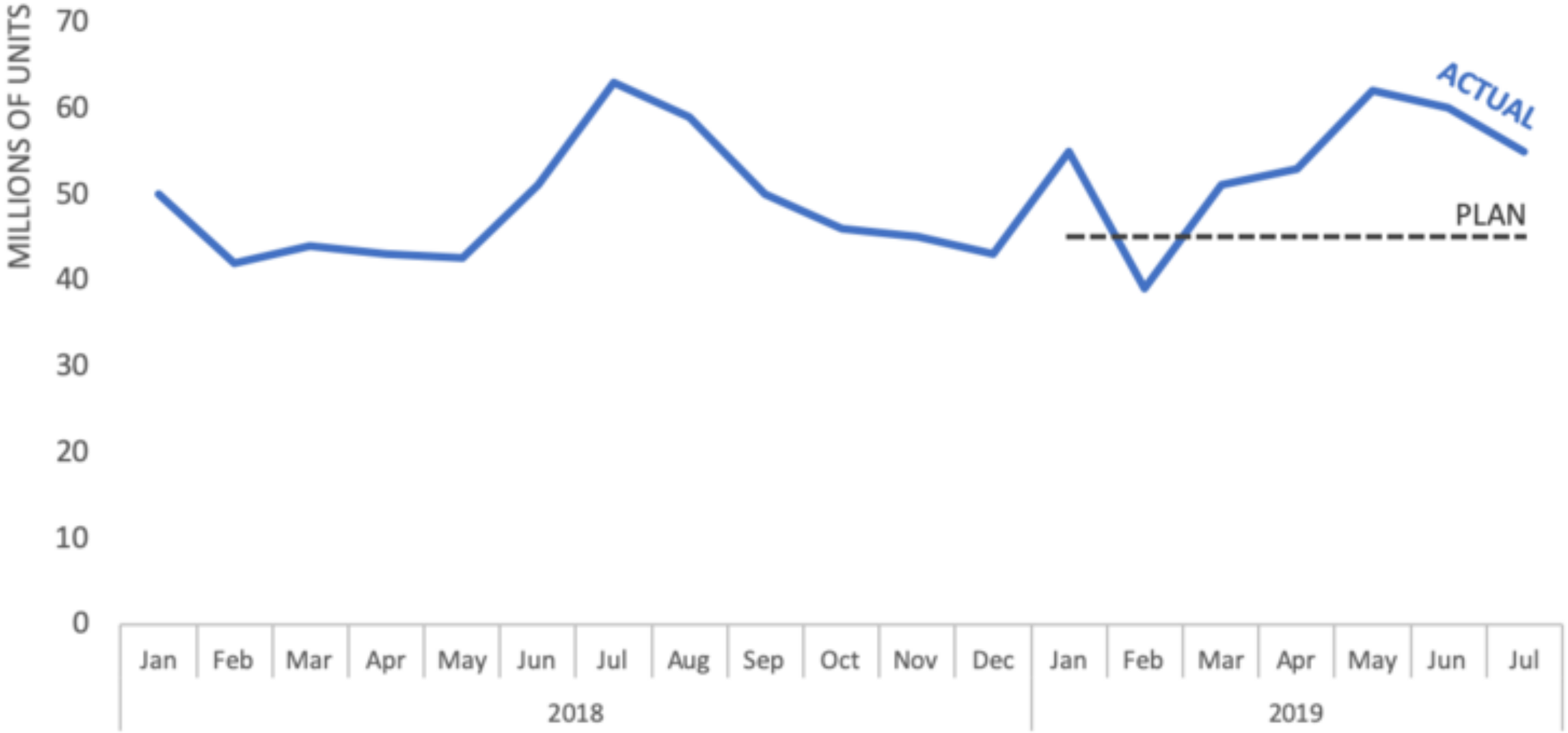
PRIMARY PRODUCT: unit sales



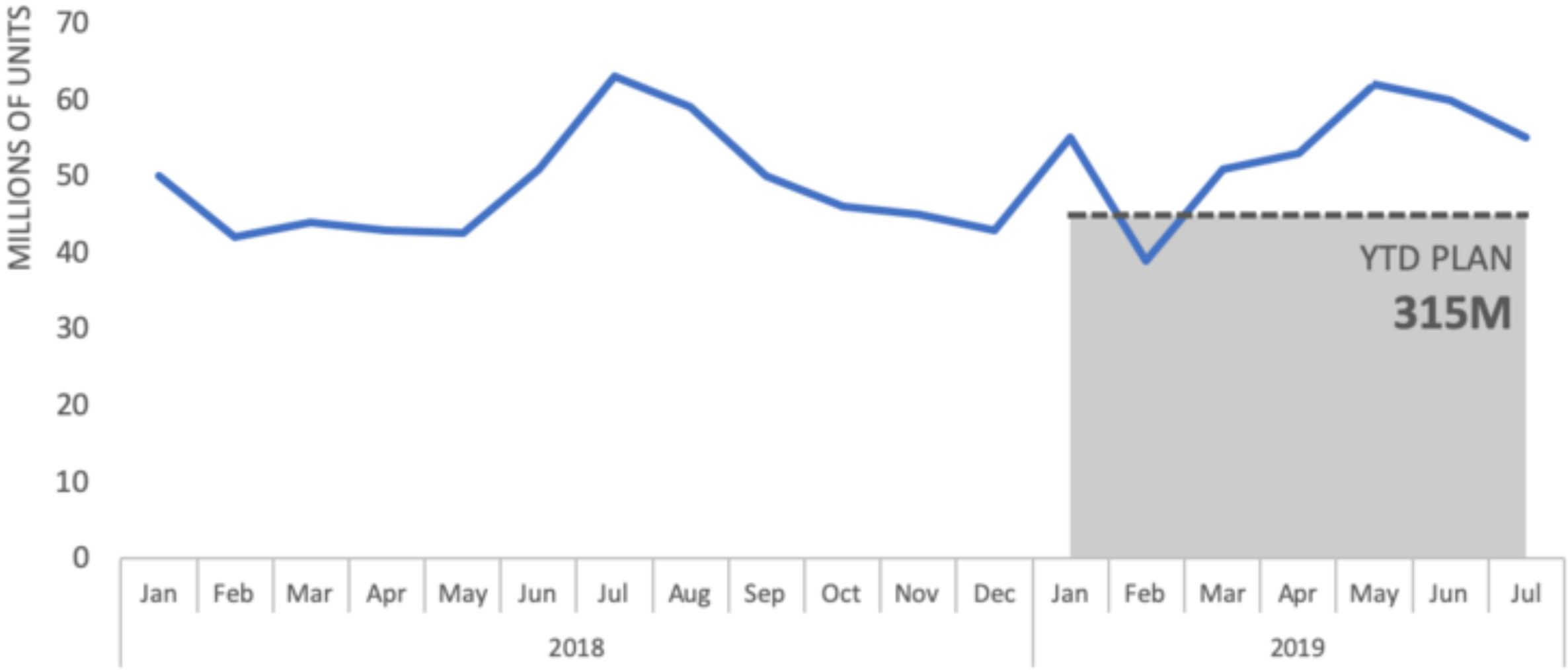
PRIMARY PRODUCT: unit sales



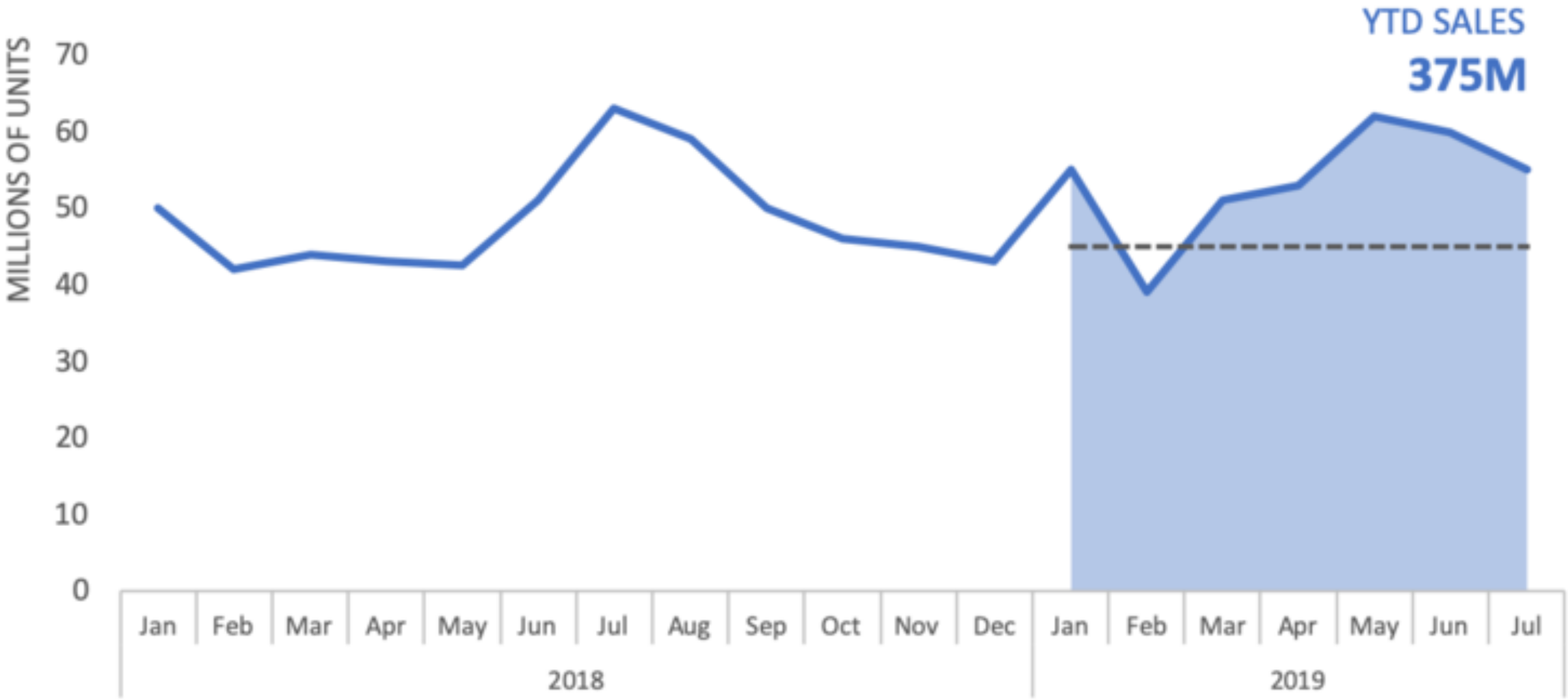
PRIMARY PRODUCT: unit sales



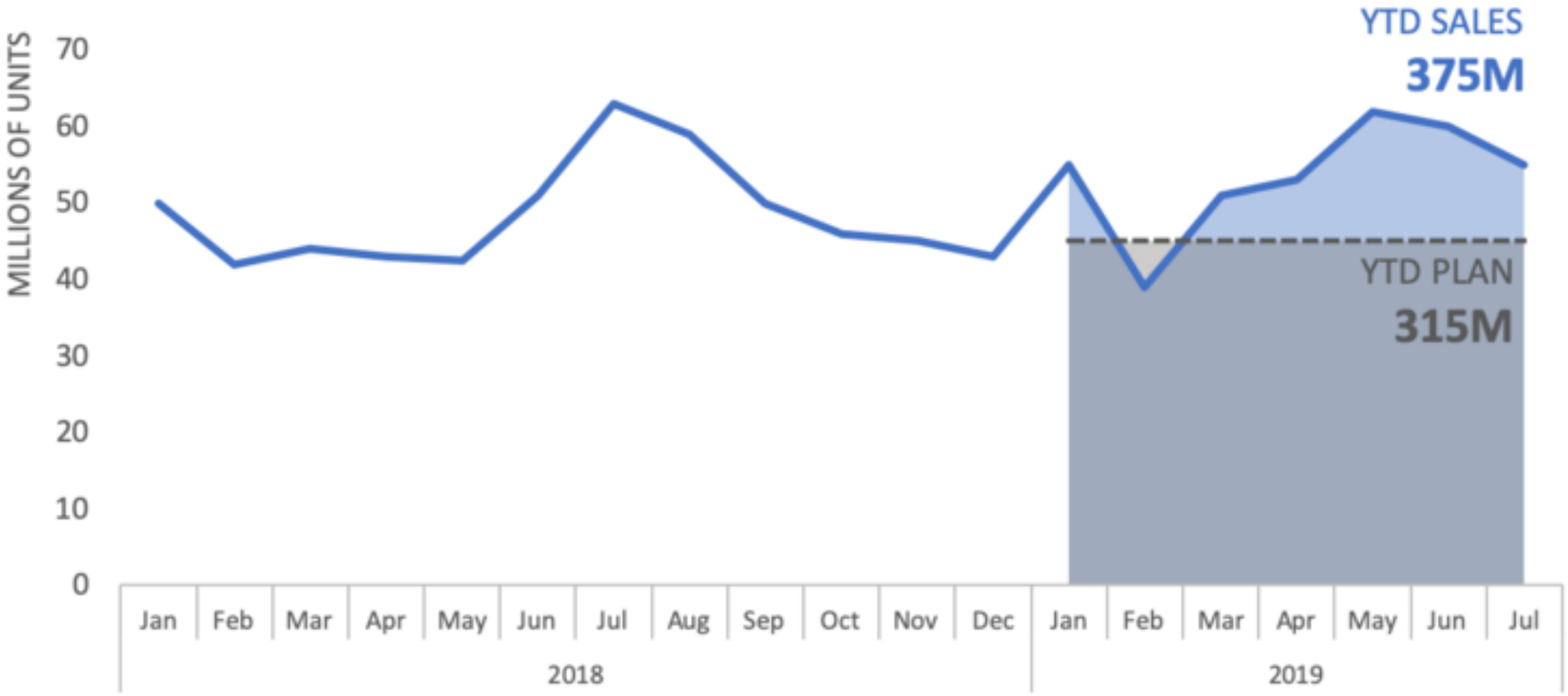
PRIMARY PRODUCT: unit sales



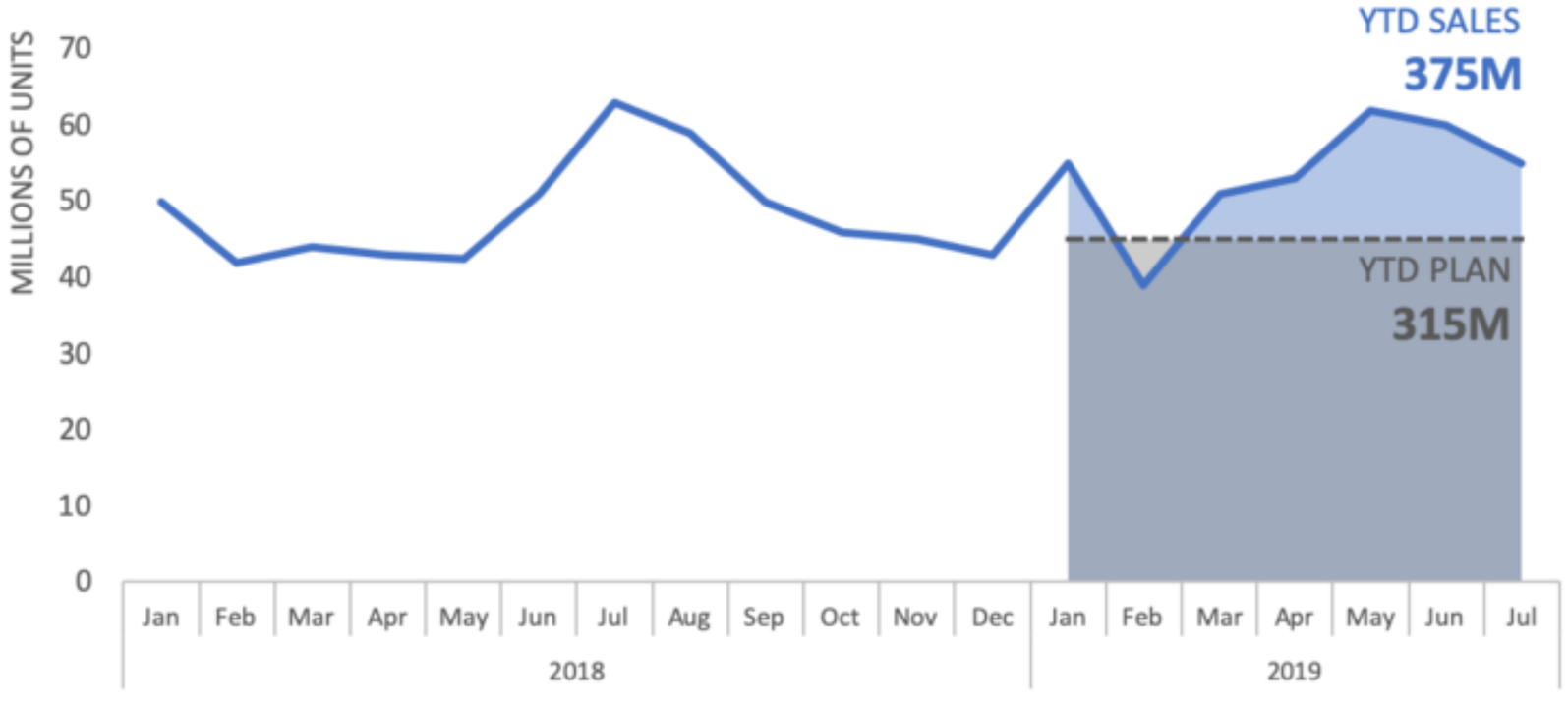
PRIMARY PRODUCT: unit sales

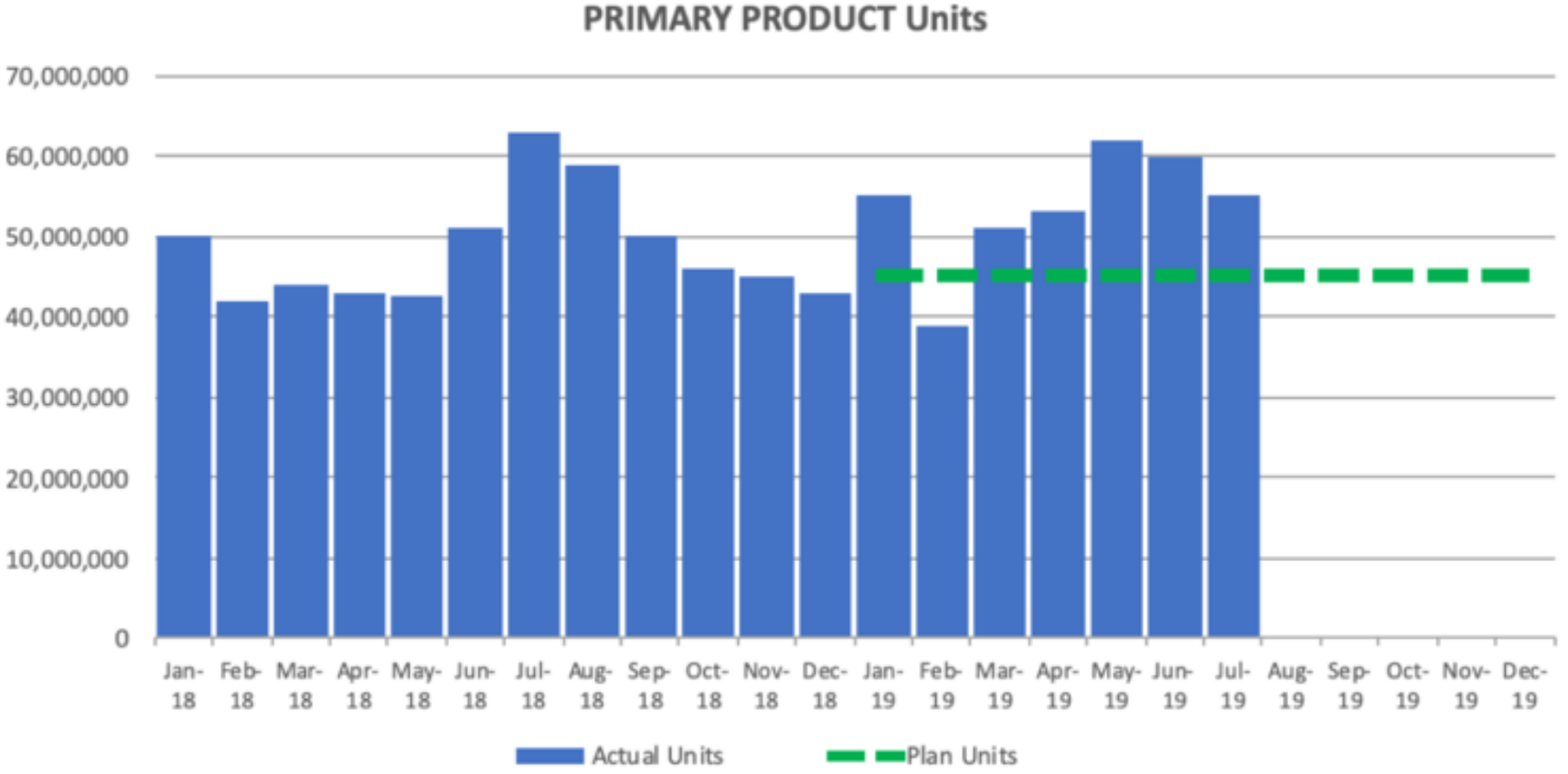


PRIMARY PRODUCT: **unit sales exceed plan by nearly 20% year to date**



PRIMARY PRODUCT: **unit sales exceed plan by nearly 20% year to date**





Group Exercise:

Write the headline

http://tiny.cc/INFO290_headlines

Individual Exercise:

Visualizing tabular data by hand

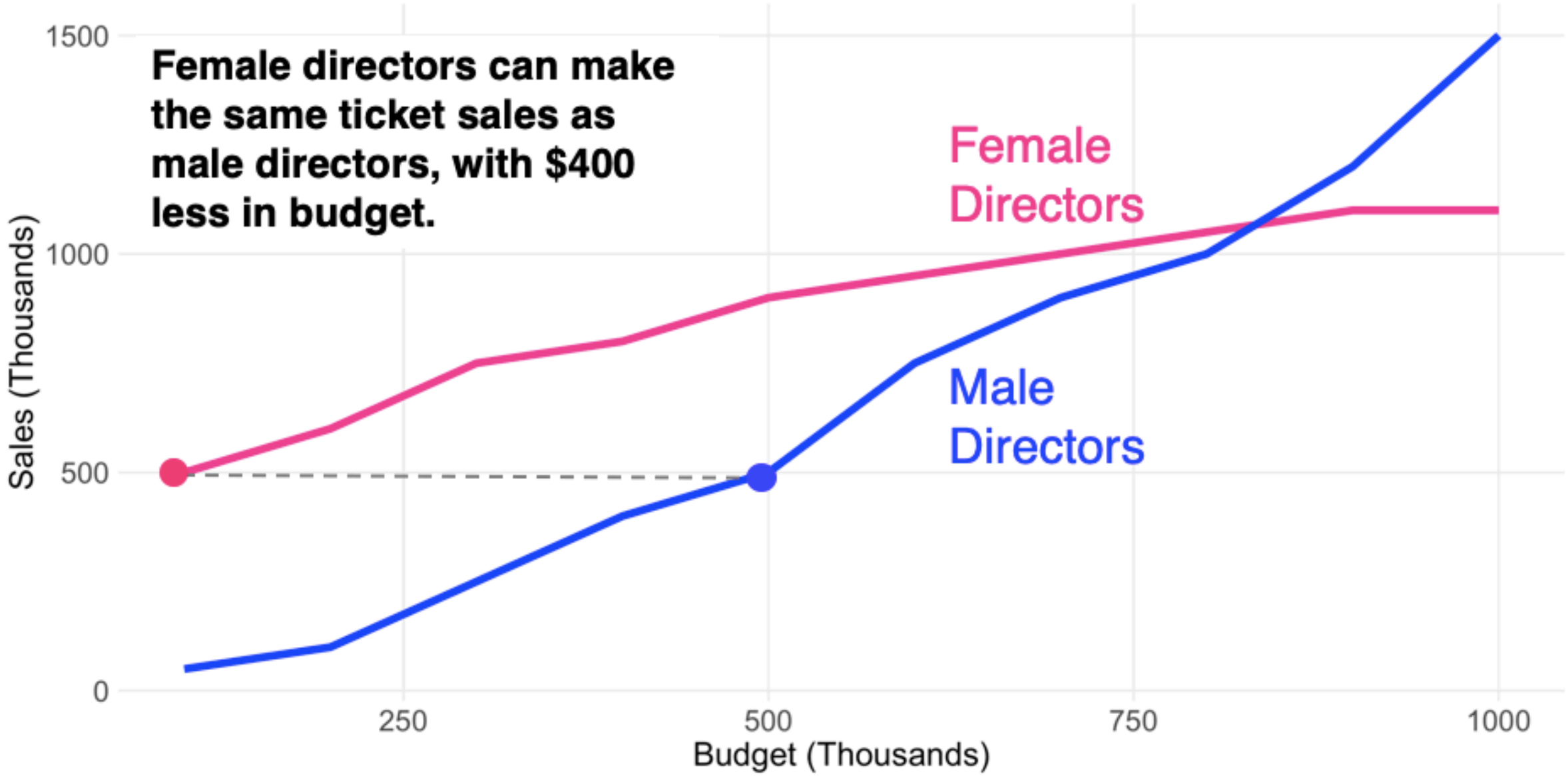
Individual Exercise:

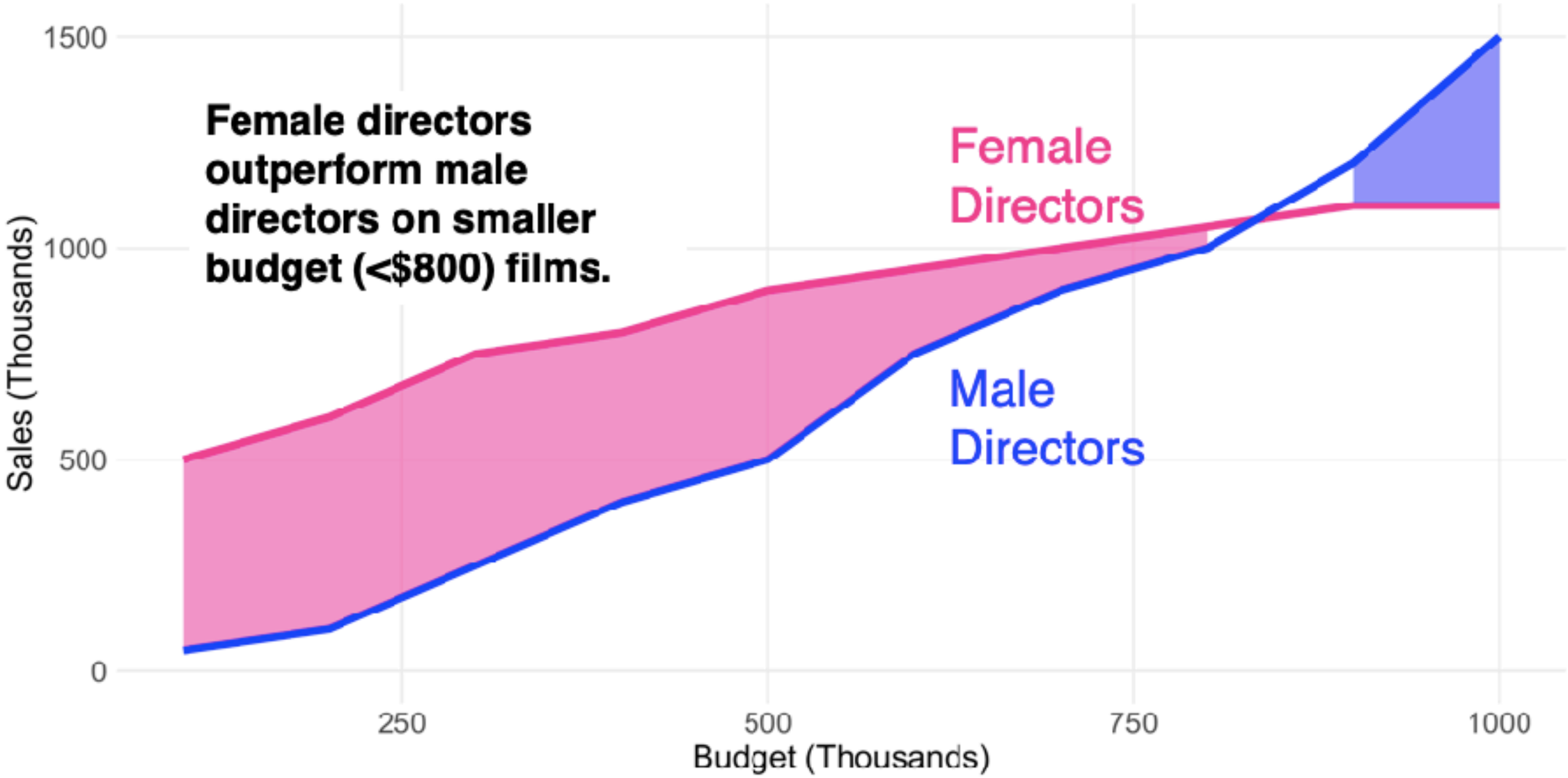
Visualizing tabular data by hand

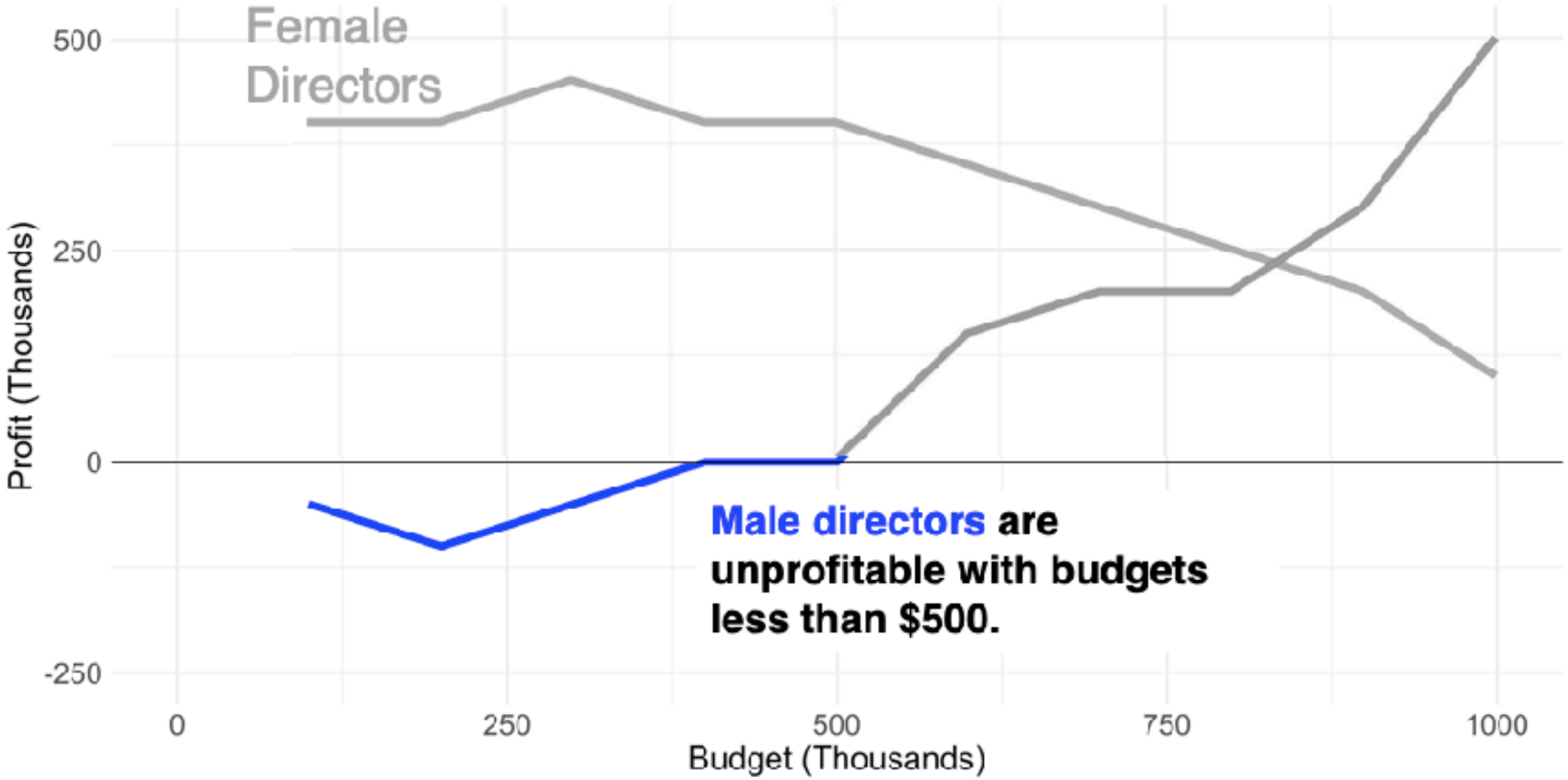
Take 7 minutes to draw, by hand, a visualization of the below data:

Budget (thousands)	Director Gender	Ticket Sales (thousands)
100	Female	500
200	Female	600
300	Female	750
400	Female	800
500	Female	900
600	Female	950
700	Female	1,000
800	Female	1,050
900	Female	1,100
1,000	Female	1,100

Budget (thousands)	Director Gender	Ticket Sales (thousands)
100	Male	50
200	Male	100
300	Male	250
400	Male	400
500	Male	500
600	Male	750
700	Male	900
800	Male	1,000
900	Male	1,200
1,000	Male	1,500


Female directors can make the same ticket sales as male directors, with $400 less in budget.
Female Directors
Male Directors
Sales (Thousands)
1500
1000
500
0
250
500
750
1000
Budget (Thousands)


Female directors outperform male directors on smaller budget (<$800) films.
Female Directors
Male Directors
Sales (Thousands)
1500
1000
500
0
250
500
750
1000
Budget (Thousands)


Female
Directors
500
250
0
-250
Profit (Thousands)
0
250
500
750
1000
Budget (Thousands)
Male directors are
unprofitable with budgets
less than $500.

Going interactive and beyond


+ableau
SOFTWARE


Leaflet


plotly

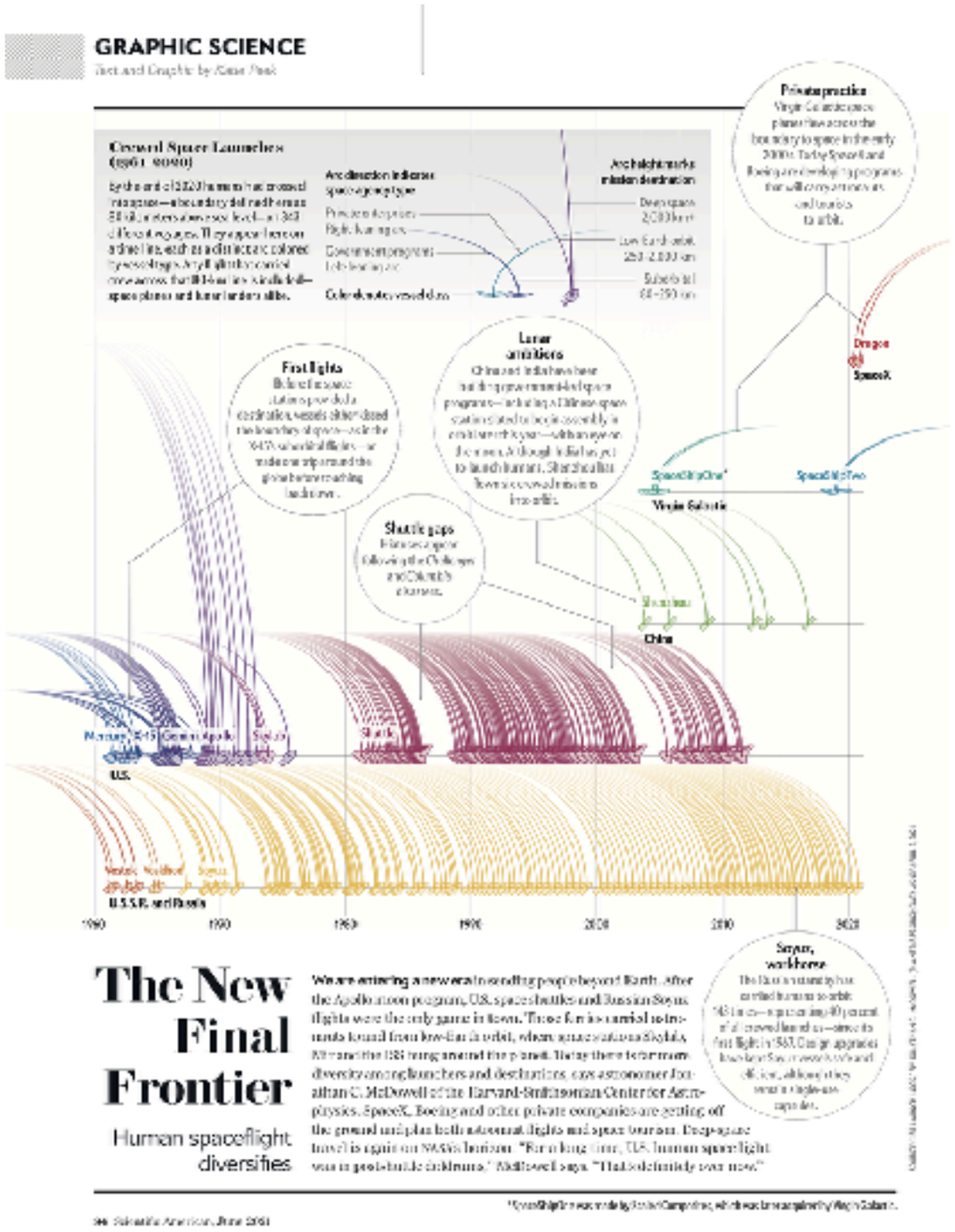

bokeh

Examples

- Tableau: Massachusetts Trial Court
- Leaflet: Police Violence Happens Here
- Plotly: Tracking COVID-19 in MA Prisons
- Bokeh: COVID-19 in Portugal
- D3: NASA Spacecraft

Where to find viz inspiration

- NYT's *The Upshot*
- *The Pudding*
- Tableau's public gallery
- Graphic science in *Scientific American*



Helpful resources

- Paul Tol's colorblind-friendly categorical color palettes
- Data Viz Project: catalog of different viz types
- Gina Reynolds' ggplot flipbook
- Kira Tebbe's Effective Data Visualization workshop
- Free online book: Fundamentals of Data Visualization
- *Storytelling with Data* books

Find these resources:

https://github.com/laurenmarietta/effective-data-viz-2025

lauren@ischool.berkeley.edu