

# Effective Data Visualization Workshop

---

Kira Tebbe - February 25, 2020

# Who is this person I'm listening to???

Data science + sociology = seeing patterns through the noise



- 1. Why is data visualization important?**
- 2. What is wrong with these plots?**
- 3. What are the important parts of a visualization?**
- 4. What are my best practices?**
- 5. How would you visualize this data?**

# Why is data viz important?

Data = information

Data visualization = communication

Good data visualization = effective communication

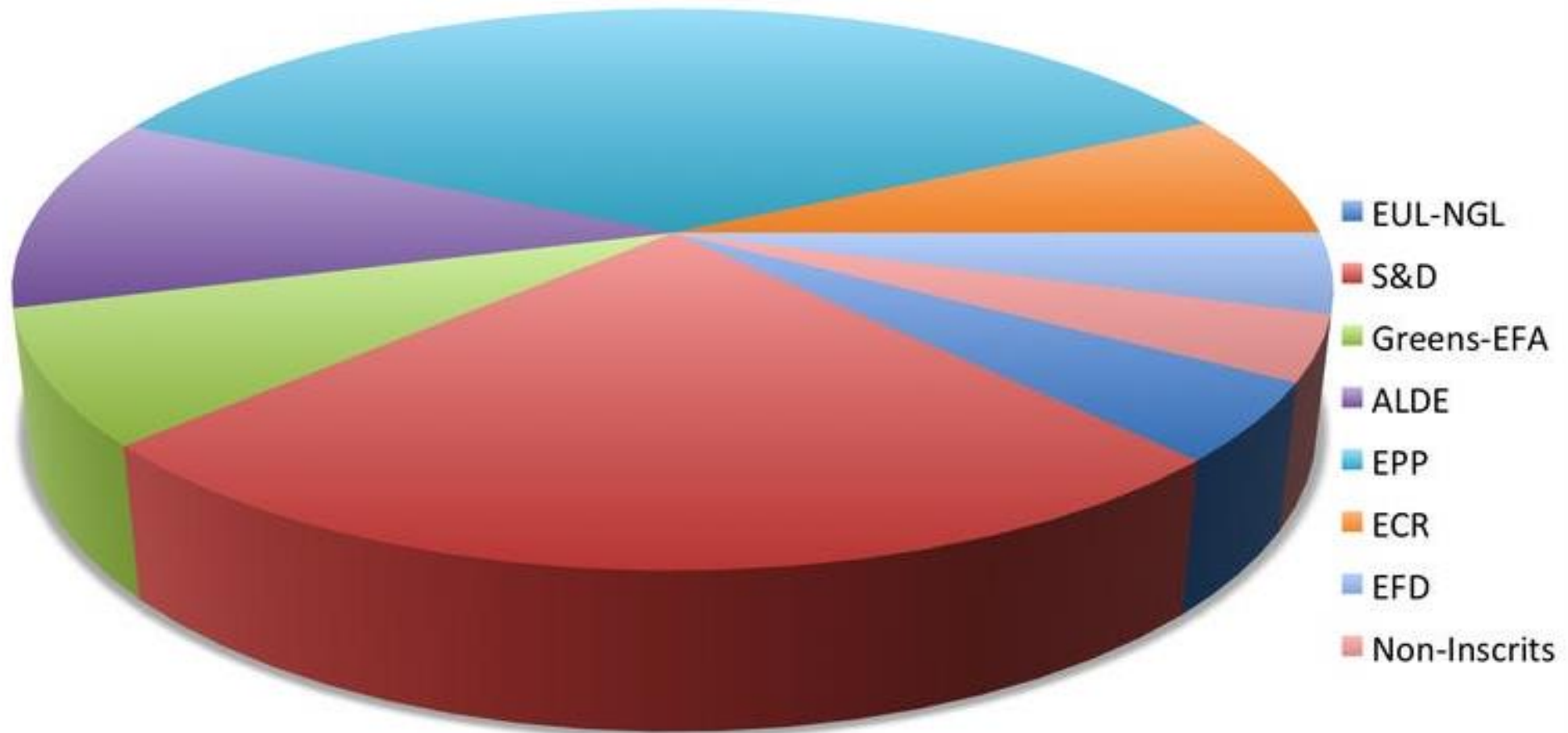
# What do mediocre visualizations give us?

- Misleading results
- Incorrect interpretations
- Confusion
- Lack of attention
- Worse impression of work quality

What's wrong  
with these plots?

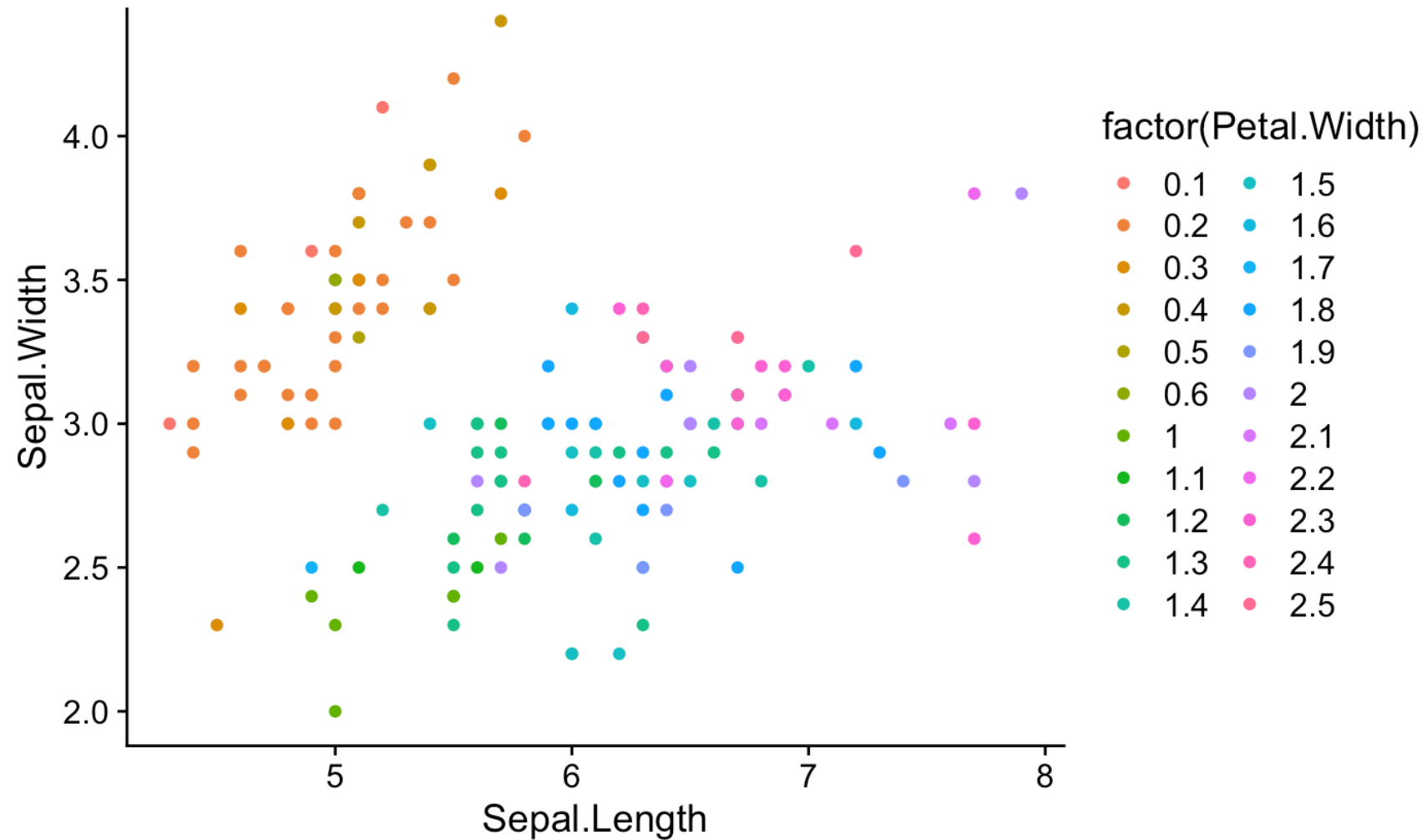
# What could be improved?

## European Parliament Party Breakdown

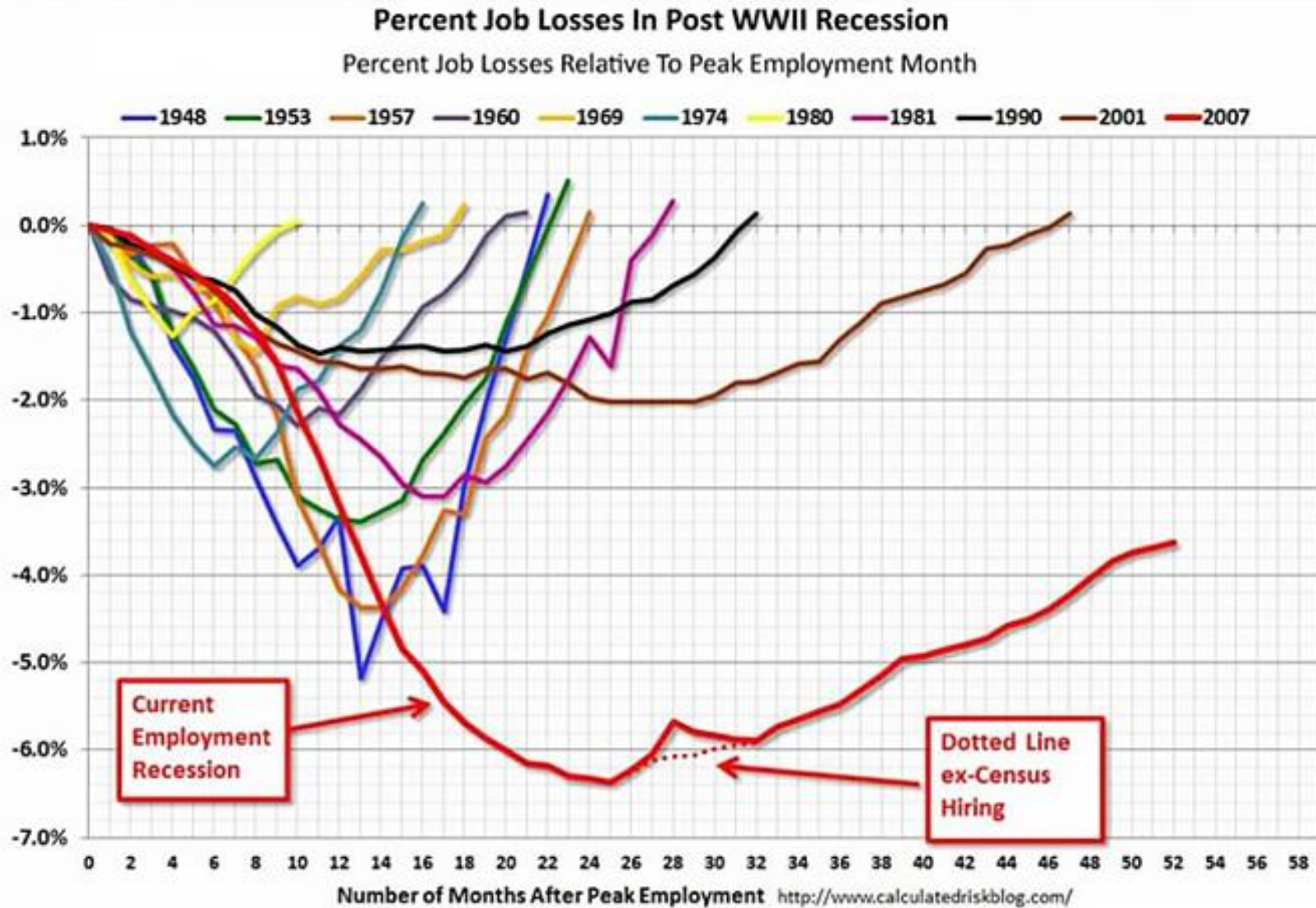




# What could be improved?

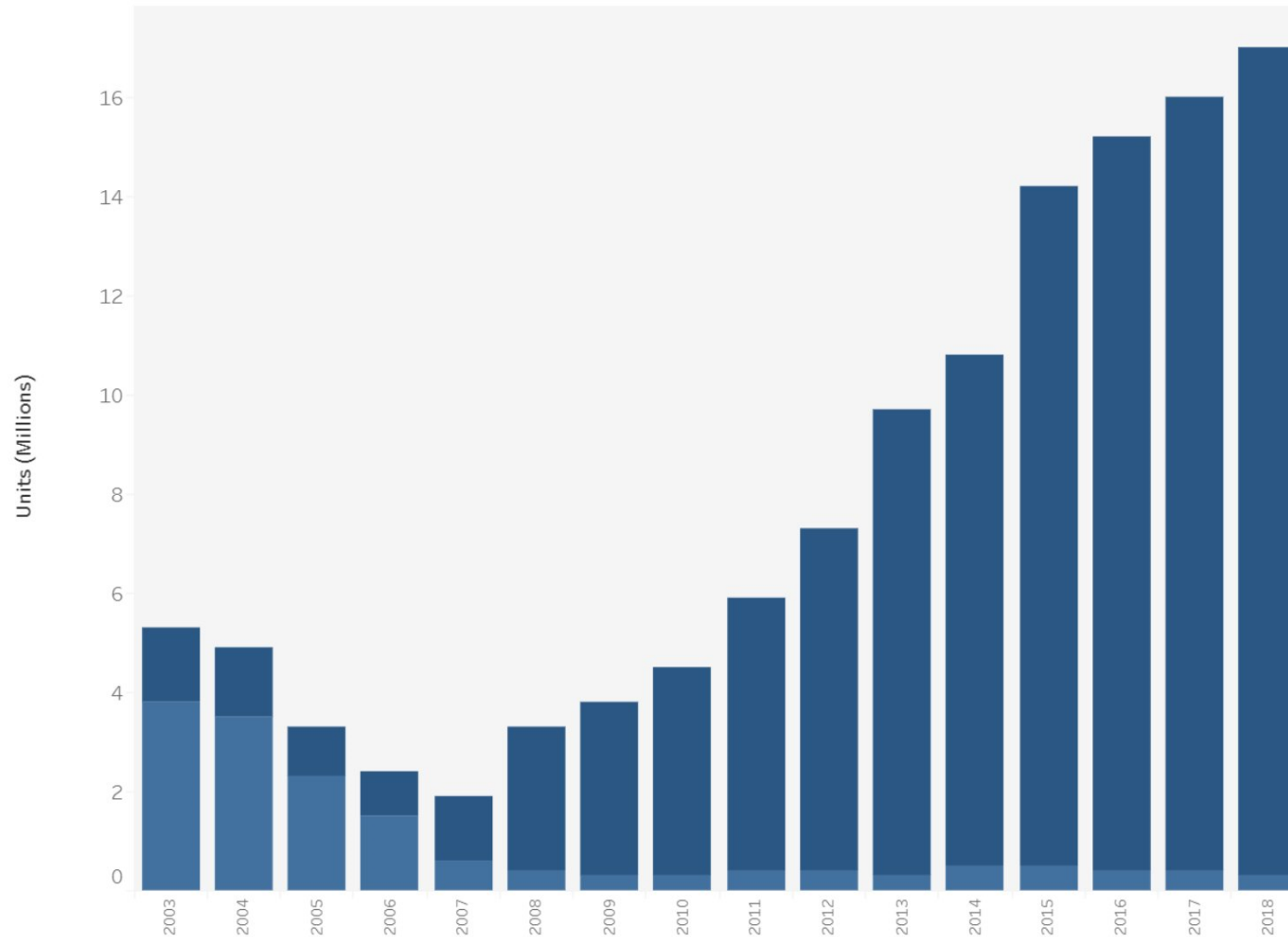


# What could be improved?



# What could be improved?

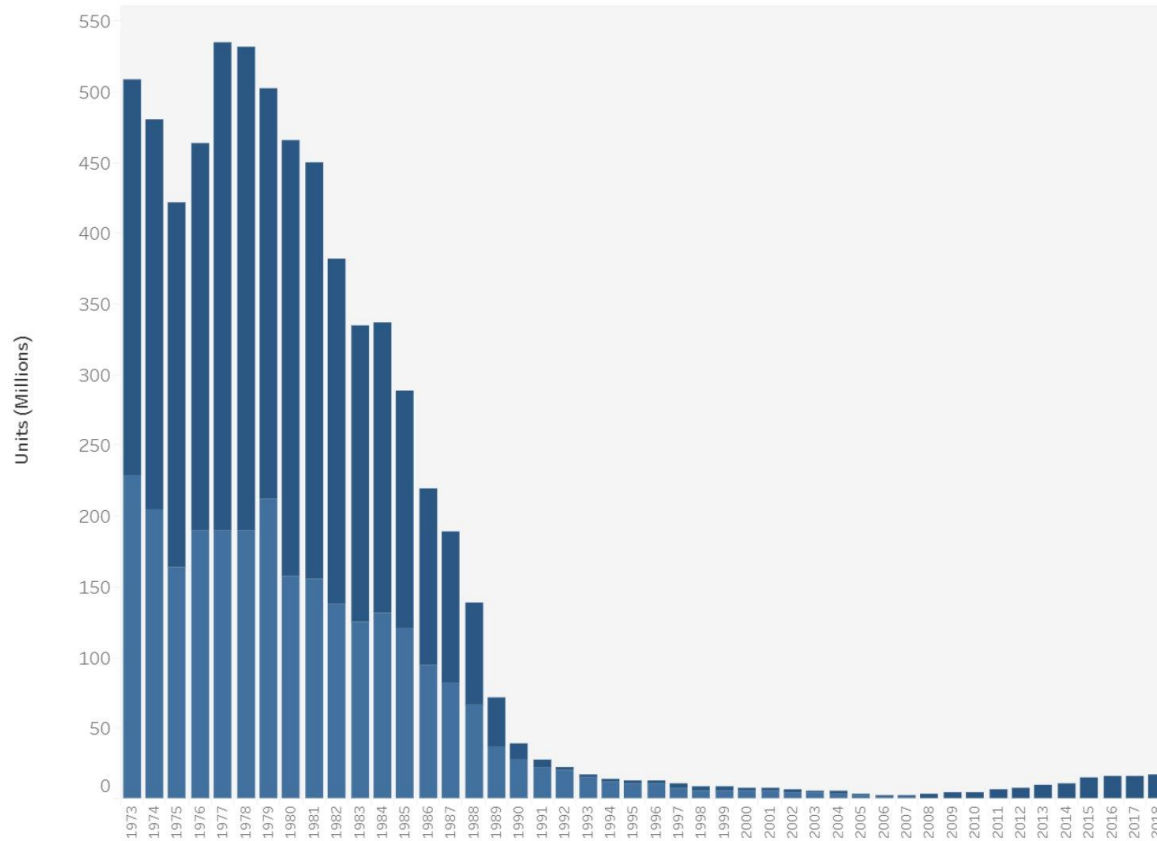
2003 to 2018, Format(s): LP/EP & Vinyl Single  
Source: RIAA



# What could be improved?

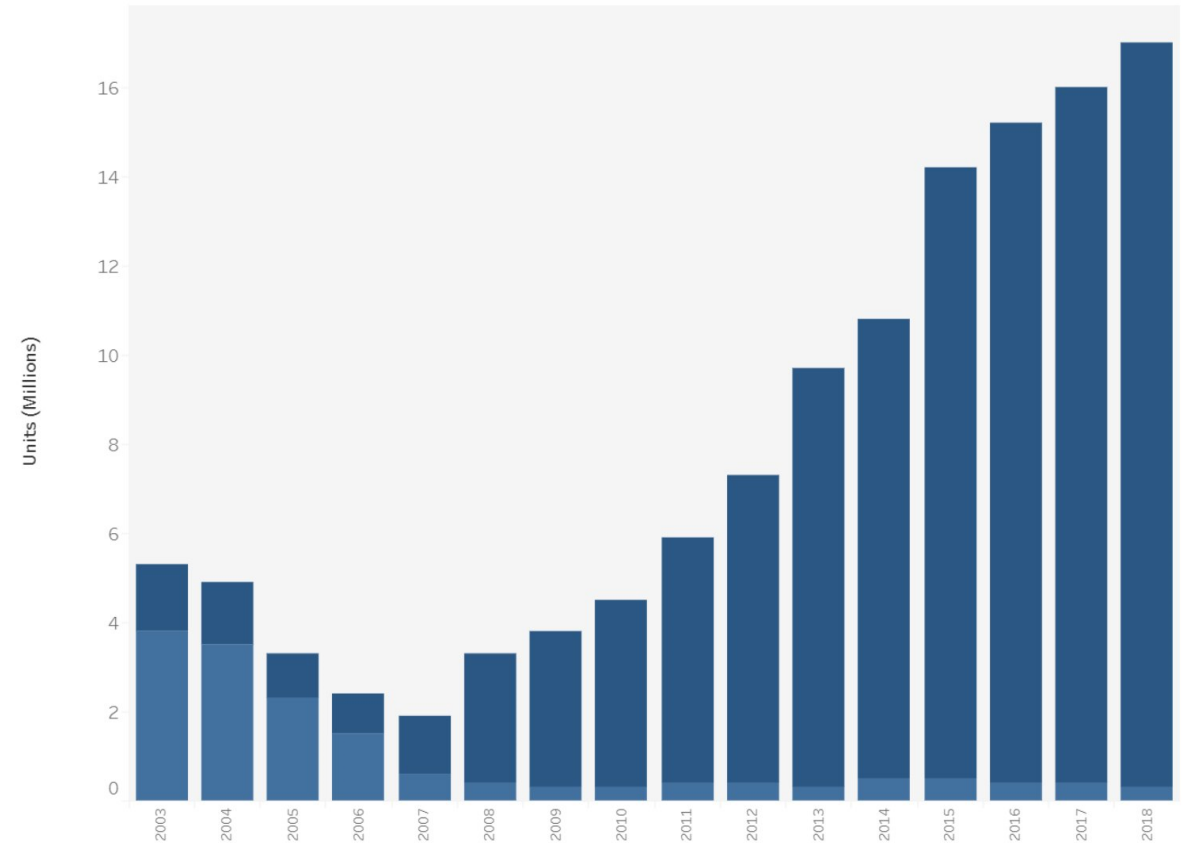
1973 to 2018, Format(s): LP/EP & Vinyl Single

Source: RIAA

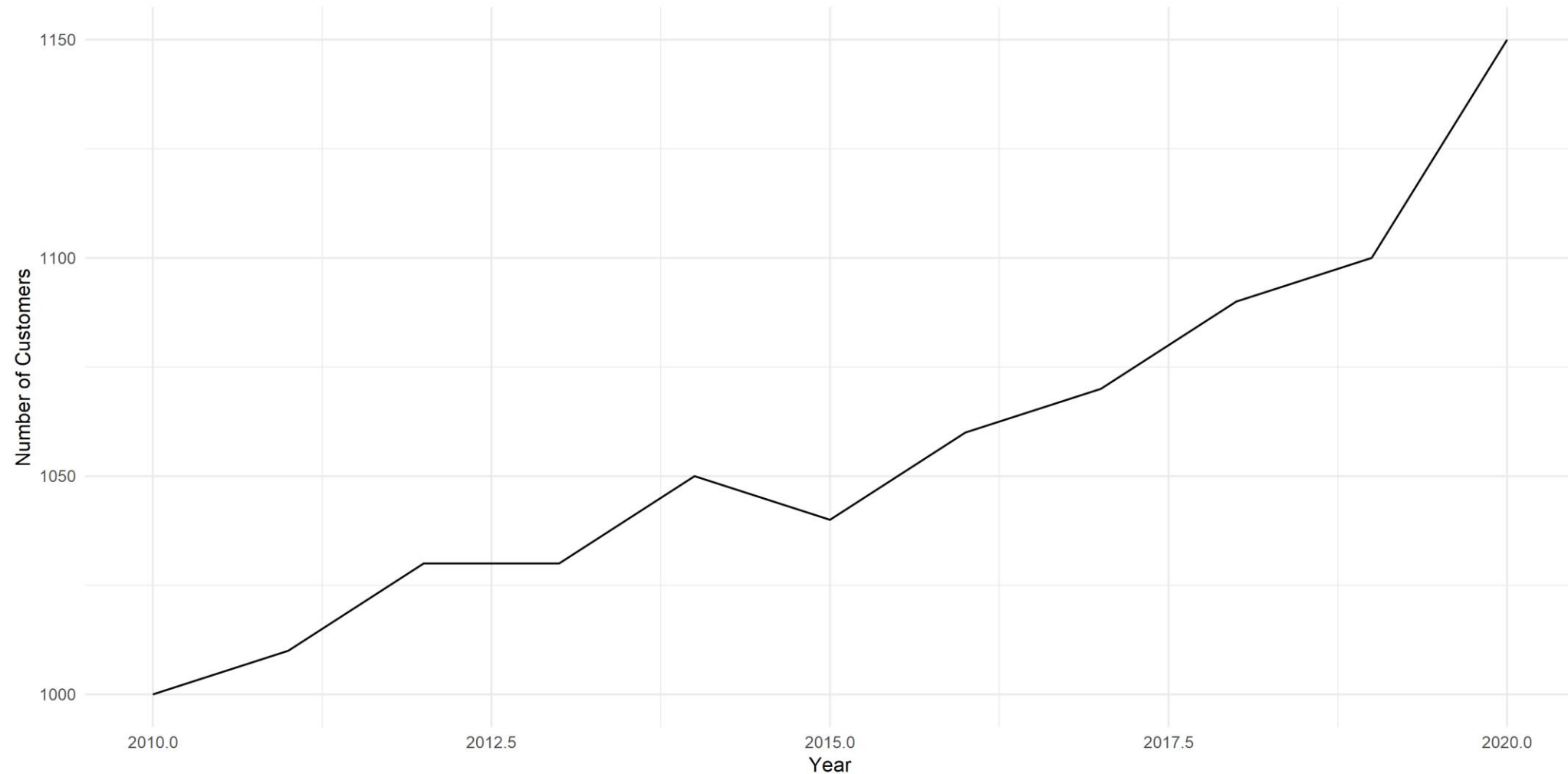


2003 to 2018, Format(s): LP/EP & Vinyl Single

Source: RIAA



# What could be improved?



# Parts of a visualization

Background color

Title and subtitle  
(size, font, color,  
location)

Major & minor  
gridline frequency  
and color

Plot type & aesthetics  
(color, size, shape, opacity)

Plant petal sizes are grouped by species  
And Virginia are the biggest!

Legend title (size,  
font, color, location,  
existence)

Species

● setosa  
▲ versicolor  
■ virginica

X and Y  
axis labels  
(size, font,  
color, tilt)



X and Y axis tick marks  
(frequency, label, size, color)

Added text on the  
plot/animation

# Seven best practices



# For context: two types of data visualization

1. Exploratory → meant for you, don't yet know what the data says
2. **Explanatory** → meant for audience, you know what you want to say
  - What we're focusing on!

# Seven best practices

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

# Seven best practices

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

# Seven best practices

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

# SEVEN BEST PRACTICES

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

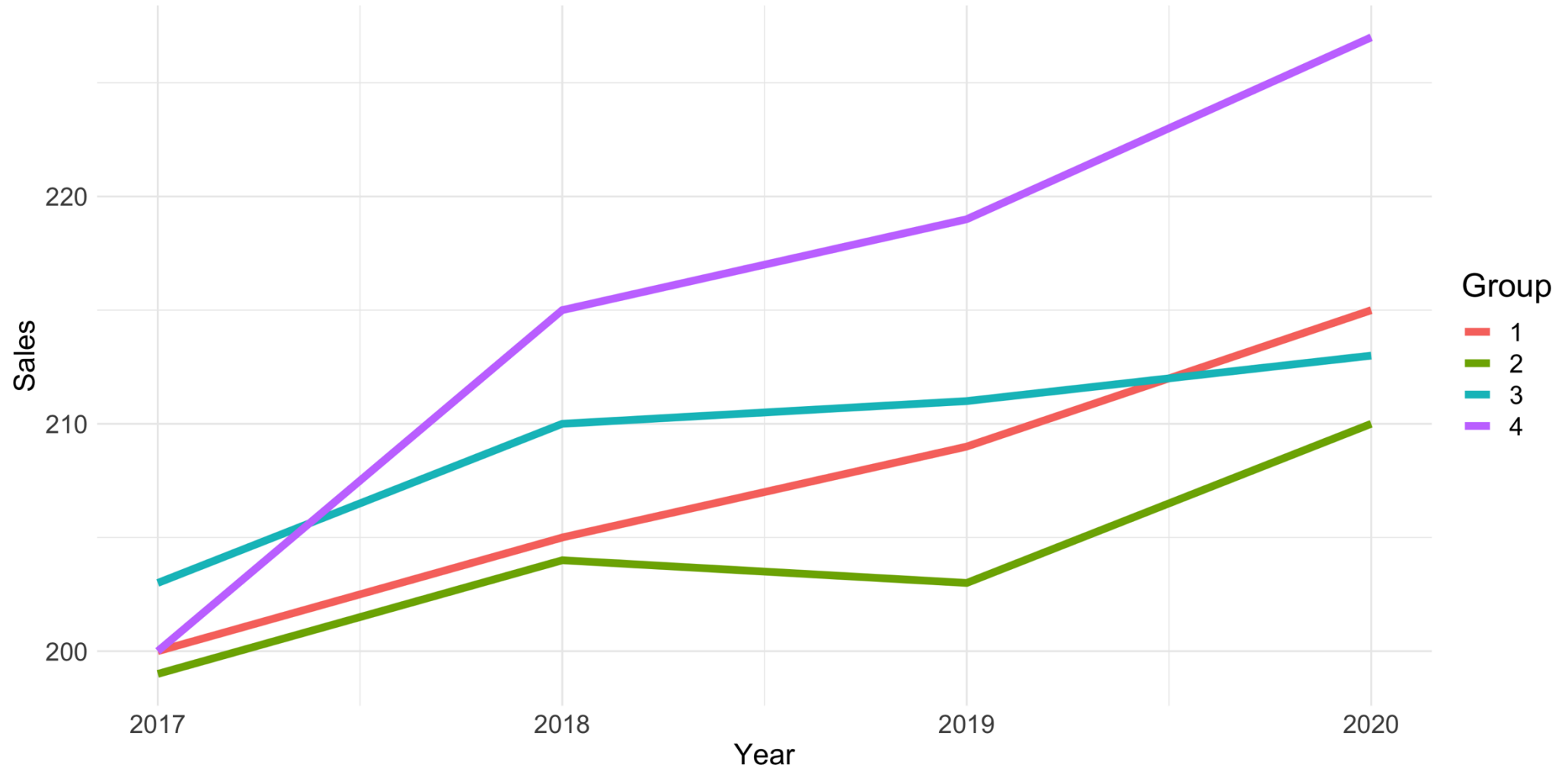
# Seven best practices

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

# Seven best practices

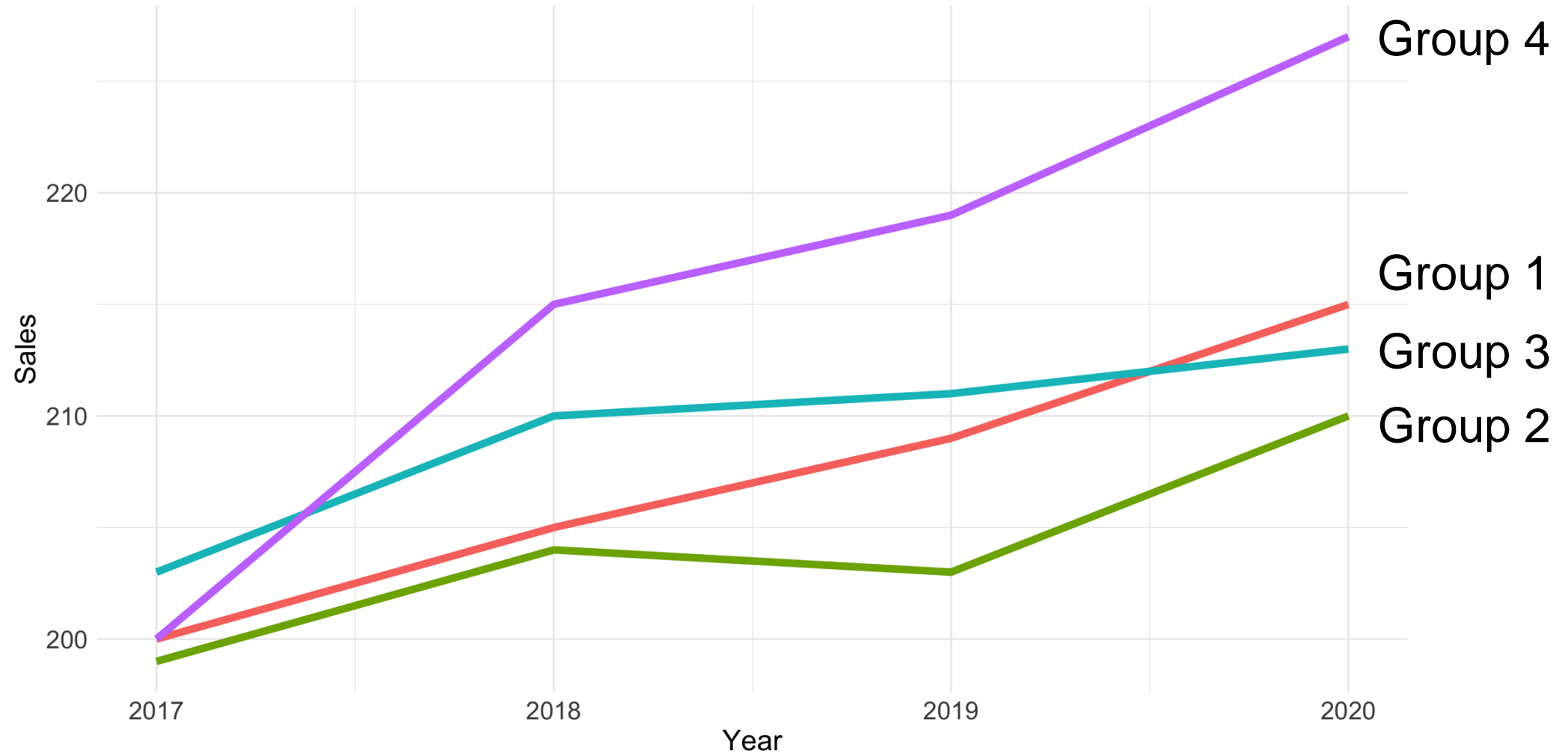
1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. **Legends (usually) suck**
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

# Why do legends suck?

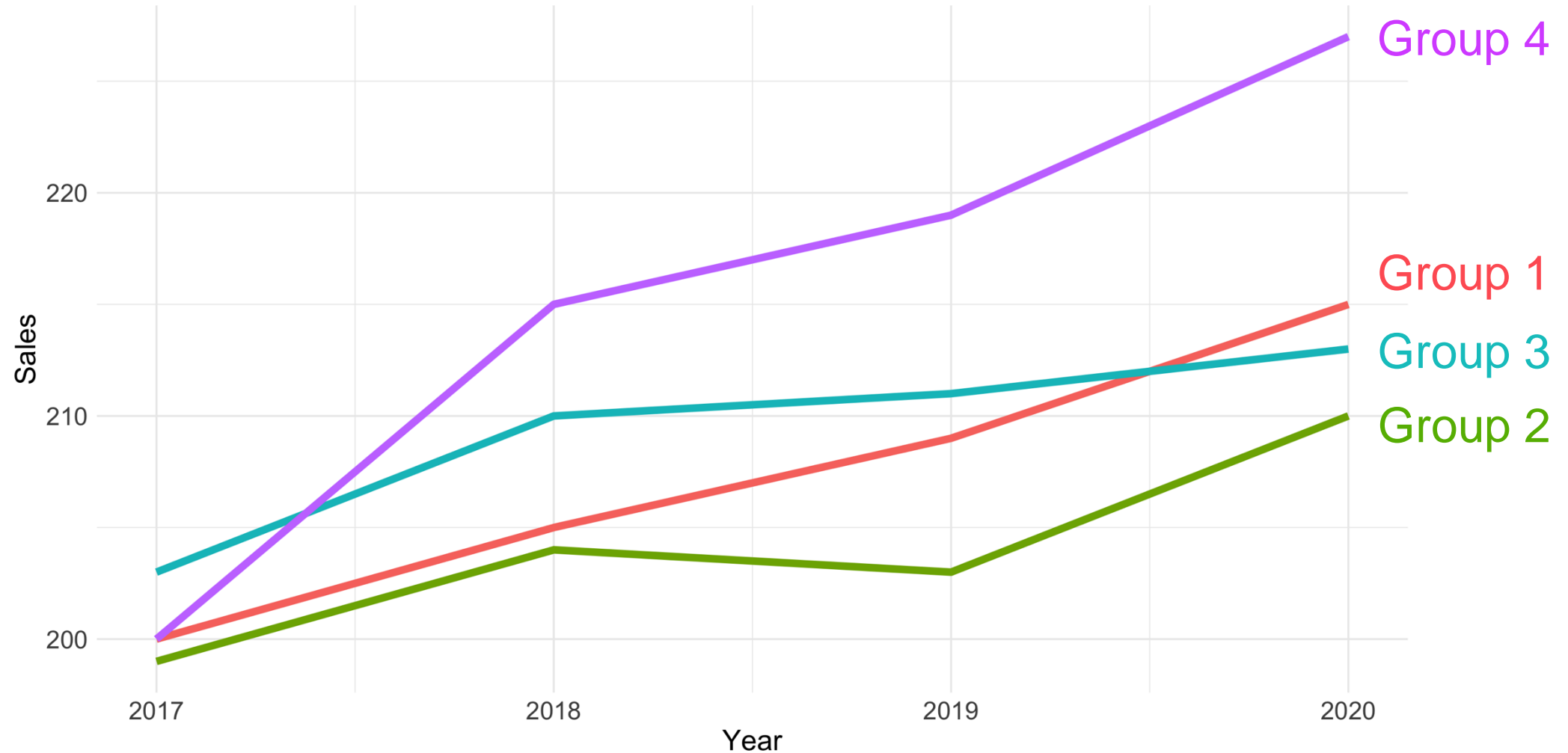




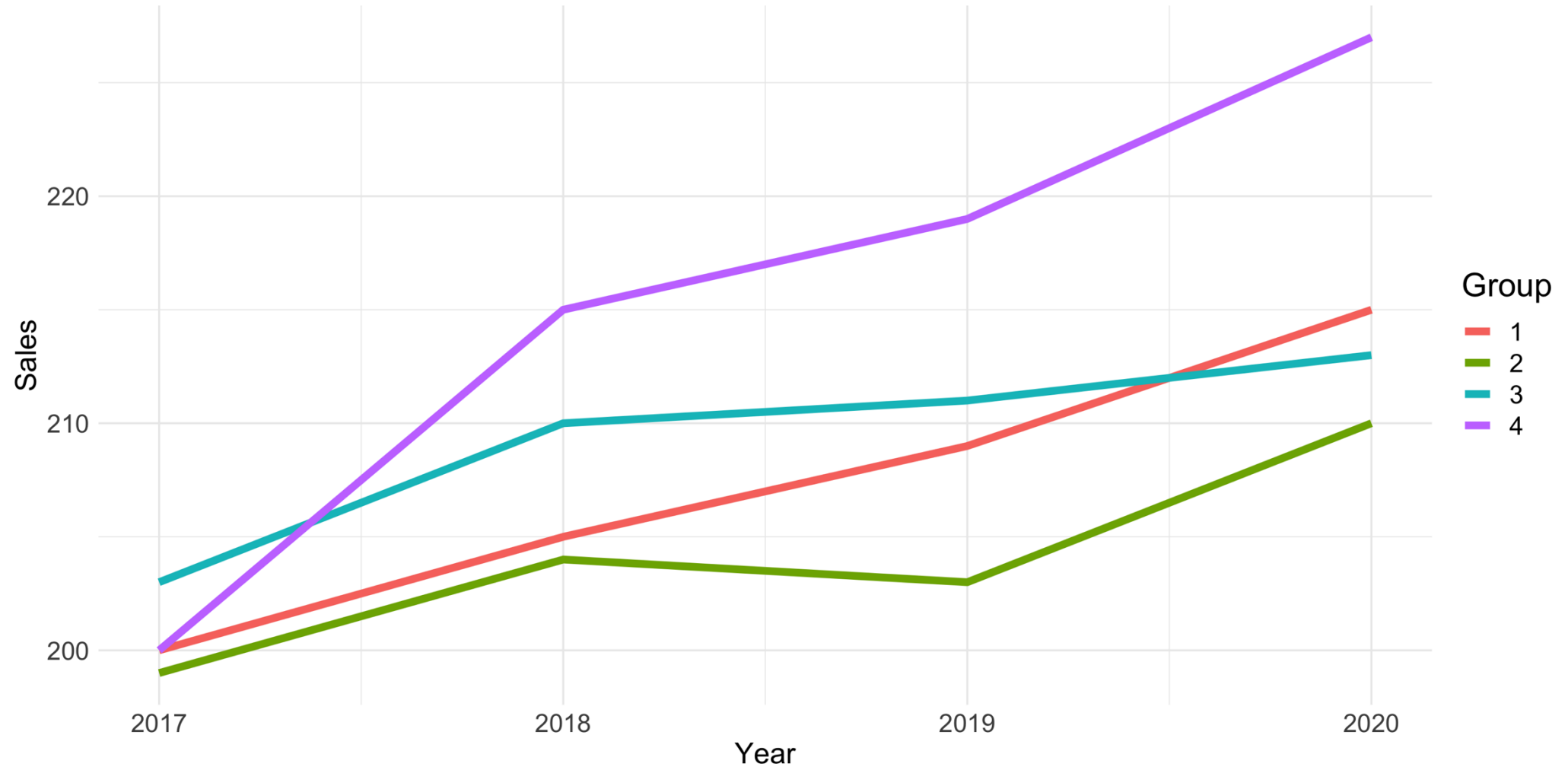
# Label your data directly



# Label your data directly



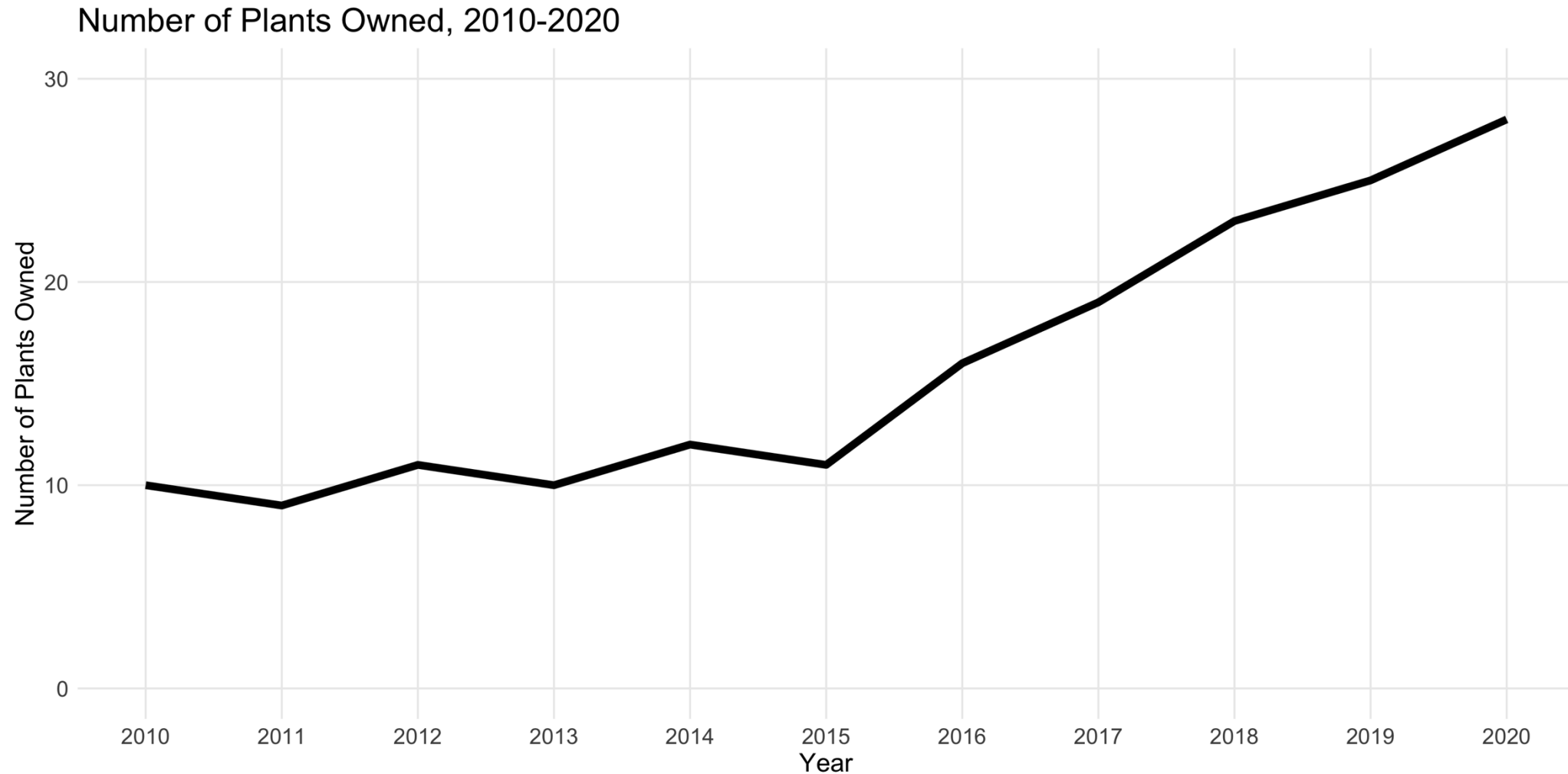
# Label your data directly



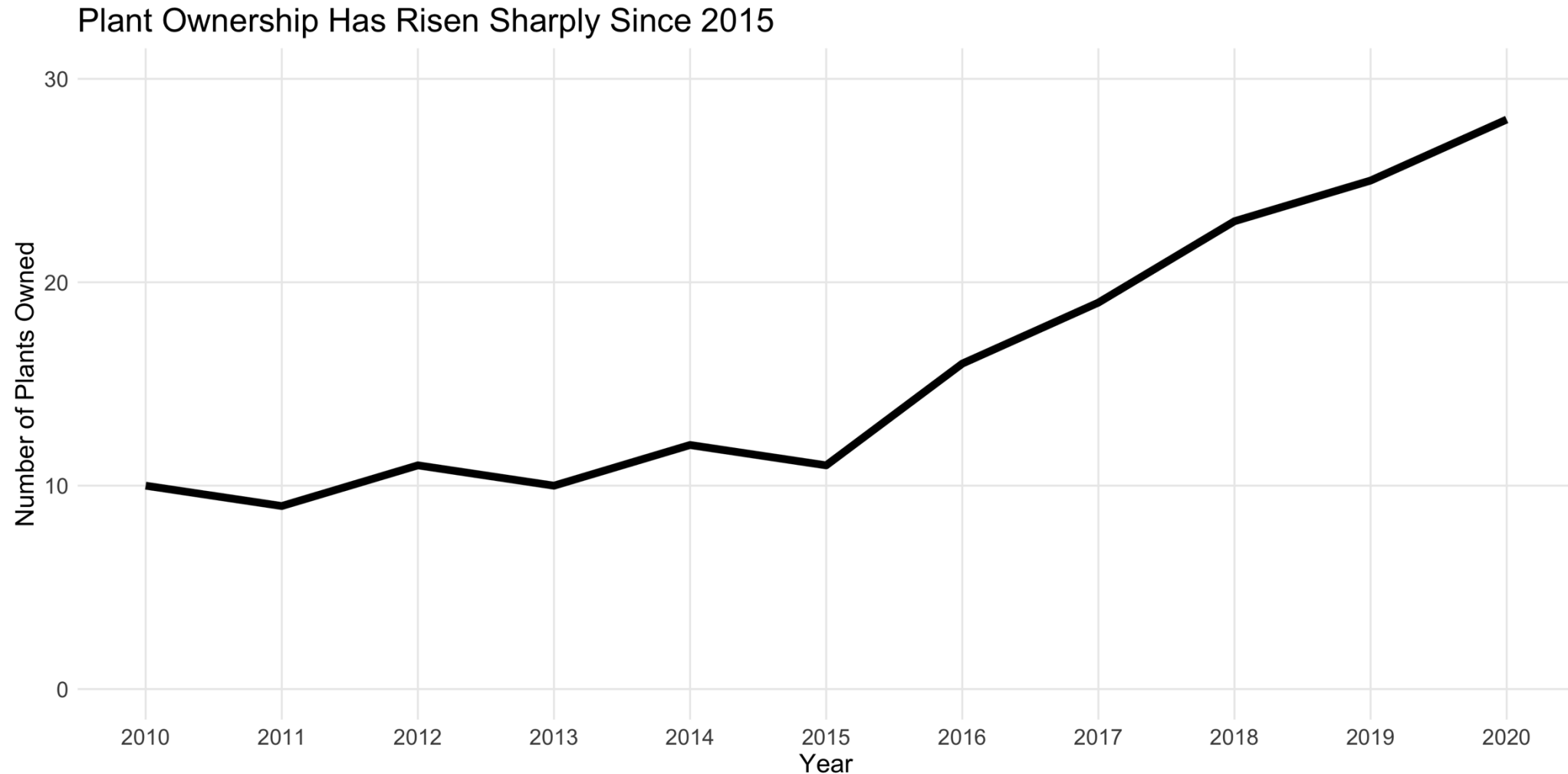
# Seven best practices

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

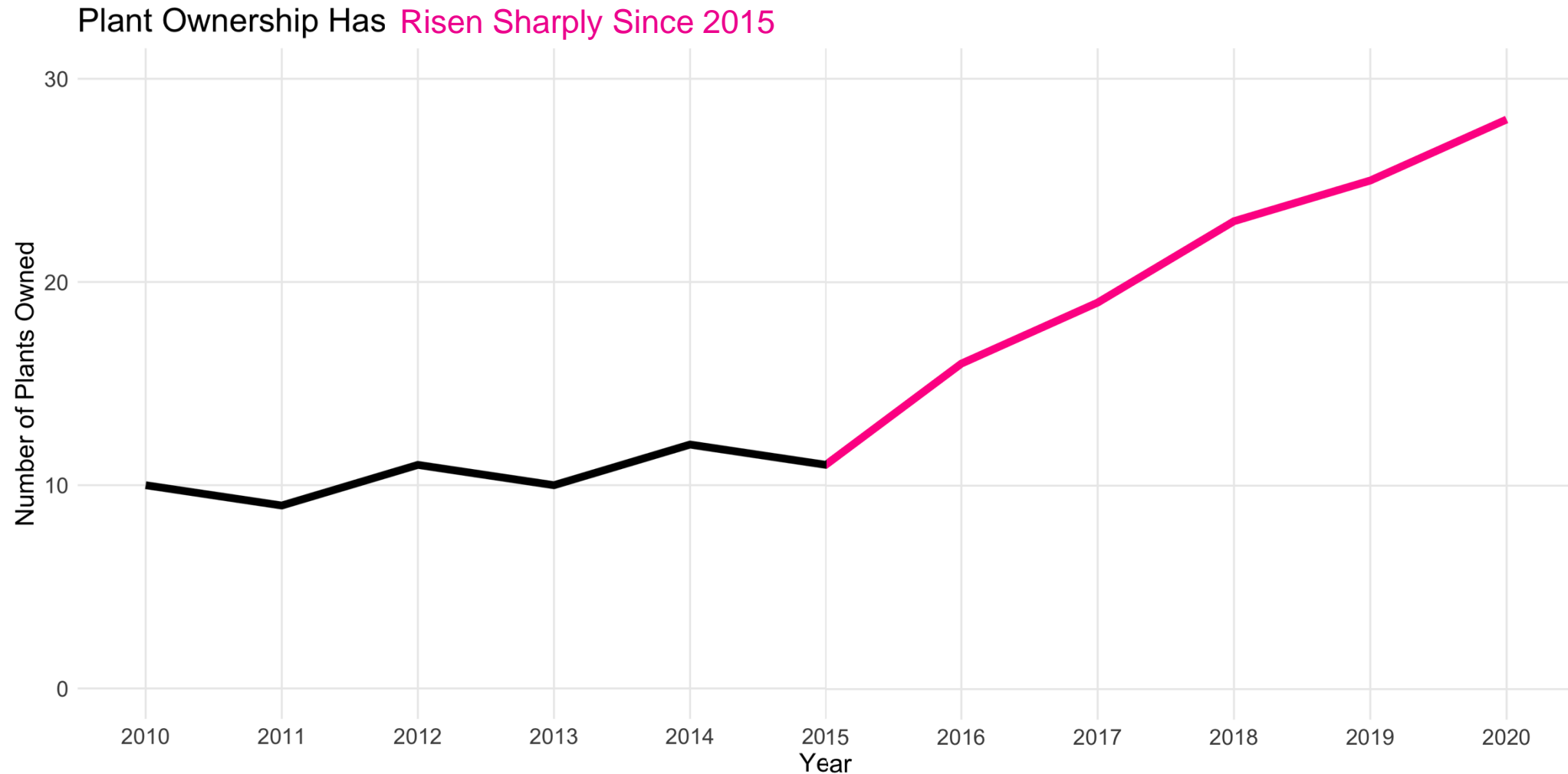
# Tell your viewer what they should see



# Tell your viewer what they should see



# Tell your viewer what they should see

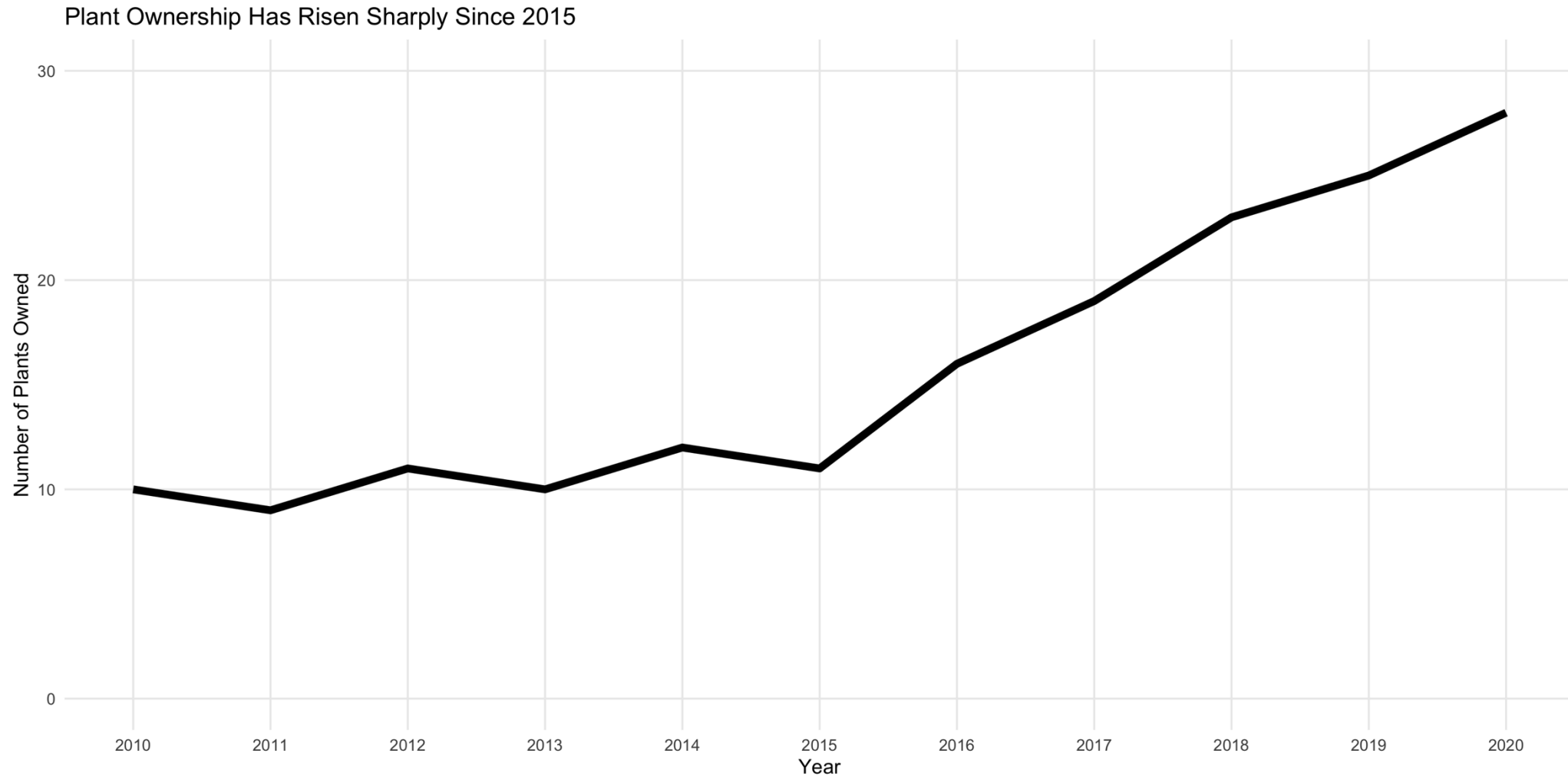


# Seven best practices

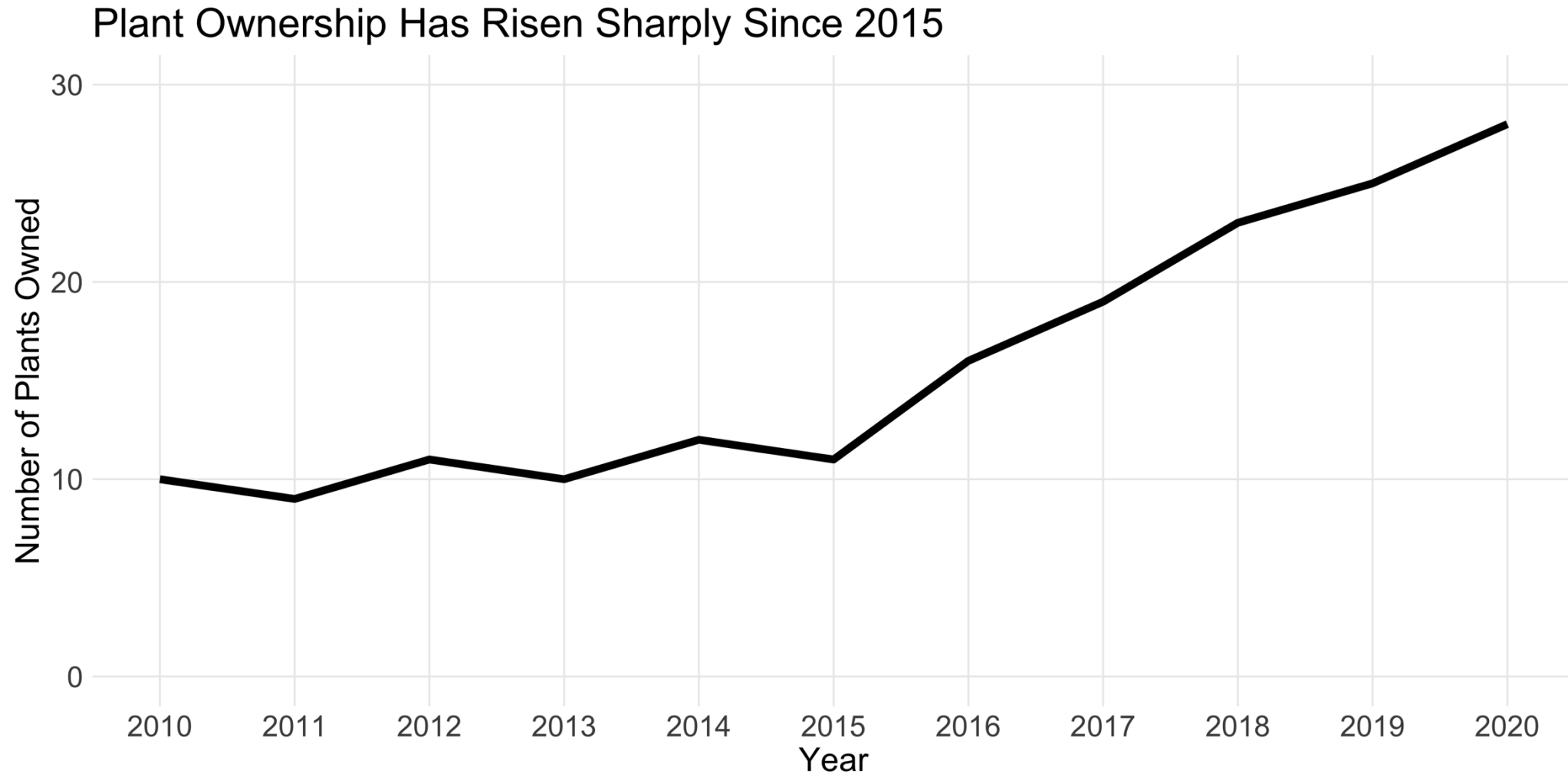
1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. **Make your text bigger than you'd think → nobody likes to squint**
7. Consider your data-to-ink ratio → remove distractions



# Nobody likes to squint!



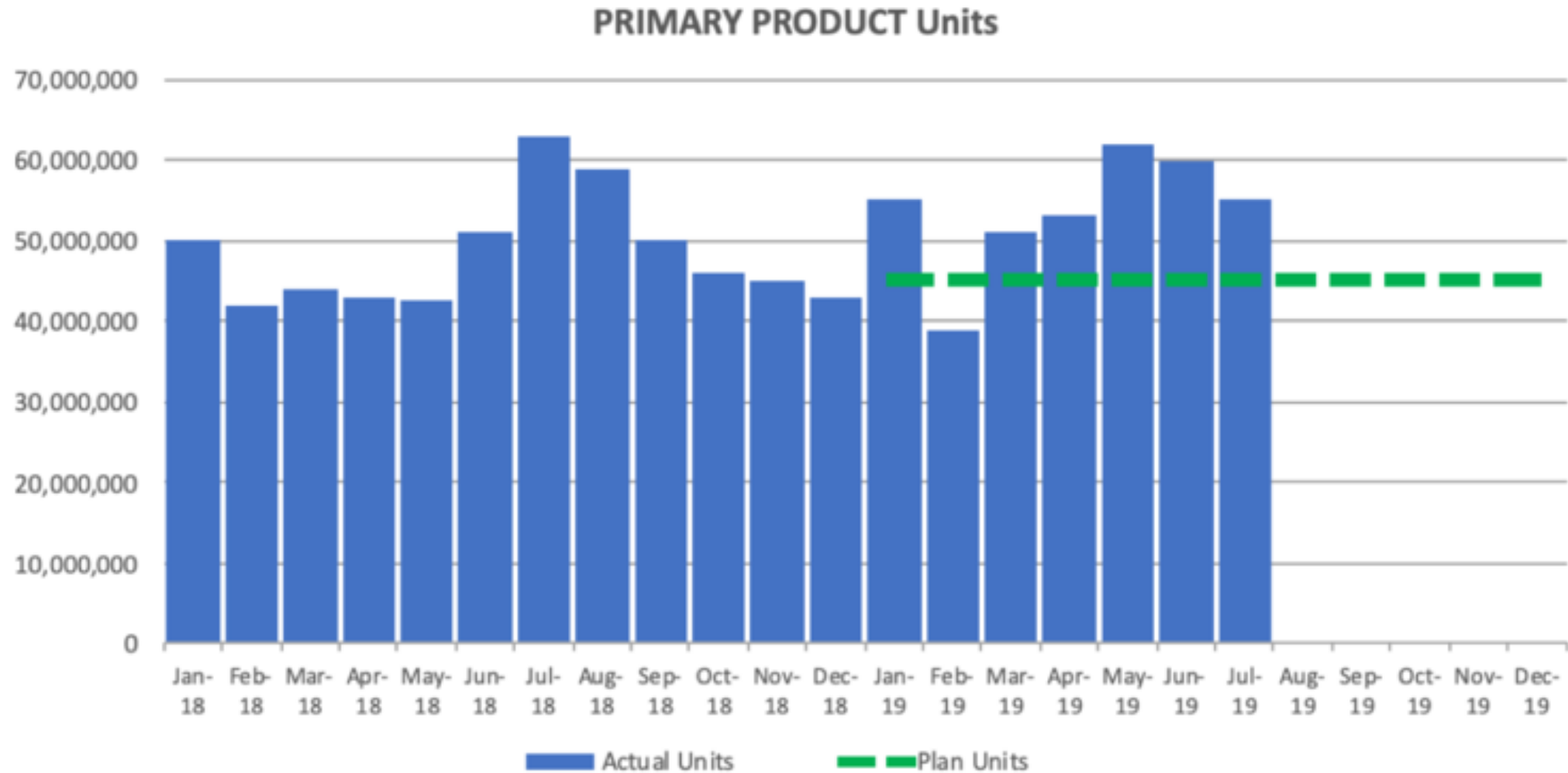
# Nobody likes to squint!



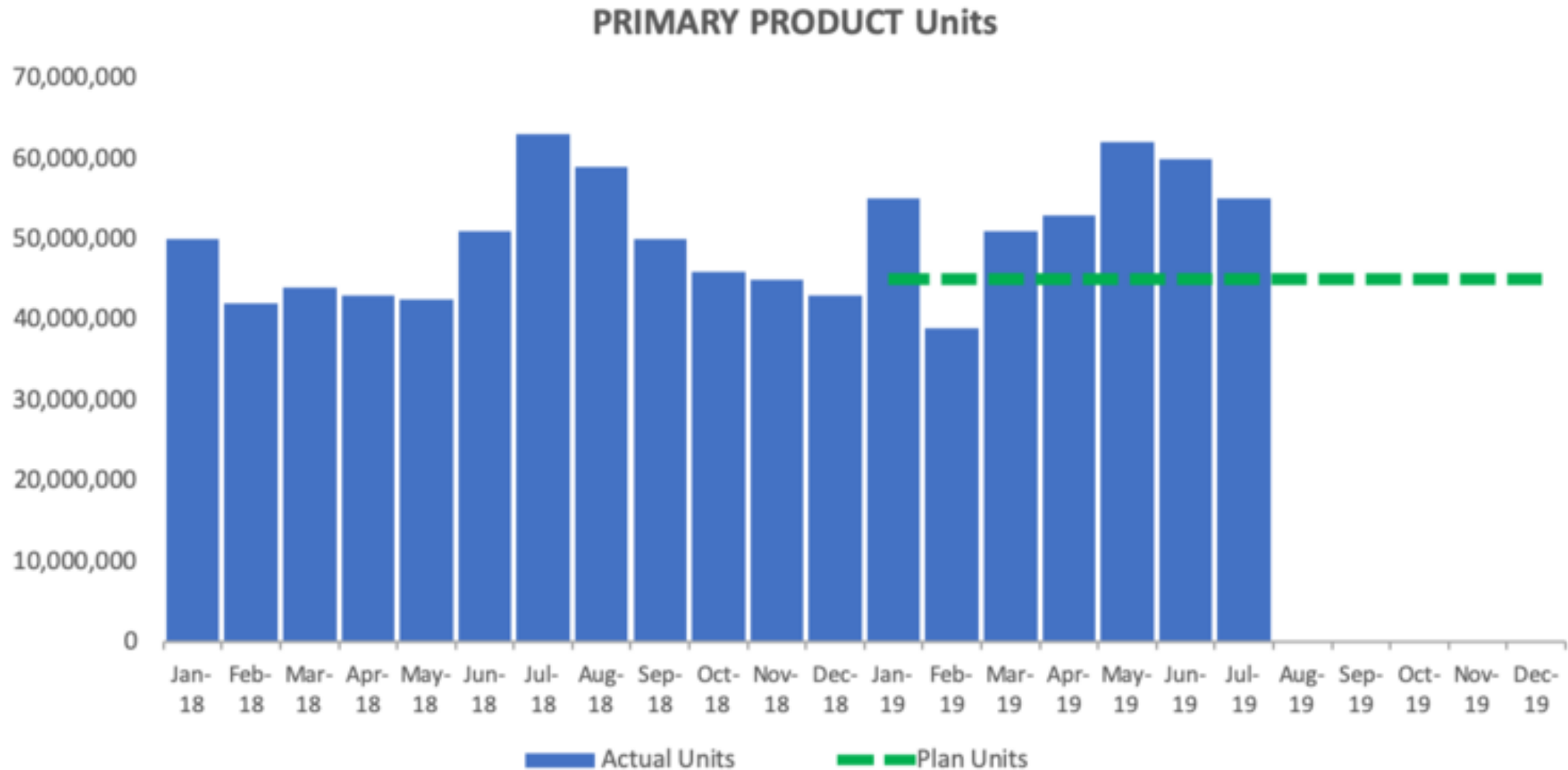
# Seven best practices

1. Believe in bar, line, and scatter plots → your audience already recognizes them
2. Use reasonable axis tick lengths → the default may not make sense
3. Use color sparingly → and when you do use it, be intentional and internally consistent
  - Color gradients for density
  - If referring to the same thing in multiple plots, use the same color each time
  - Make your gridlines gray and don't use minor gridlines
4. Legends (usually) suck
5. Use descriptive titles → tell your viewer what they should see
  - Traditional scientific titles are not explanatory
  - Don't make your audience struggle – make your point as clear as possible
6. Make your text bigger than you'd think → nobody likes to squint
7. Consider your data-to-ink ratio → remove distractions

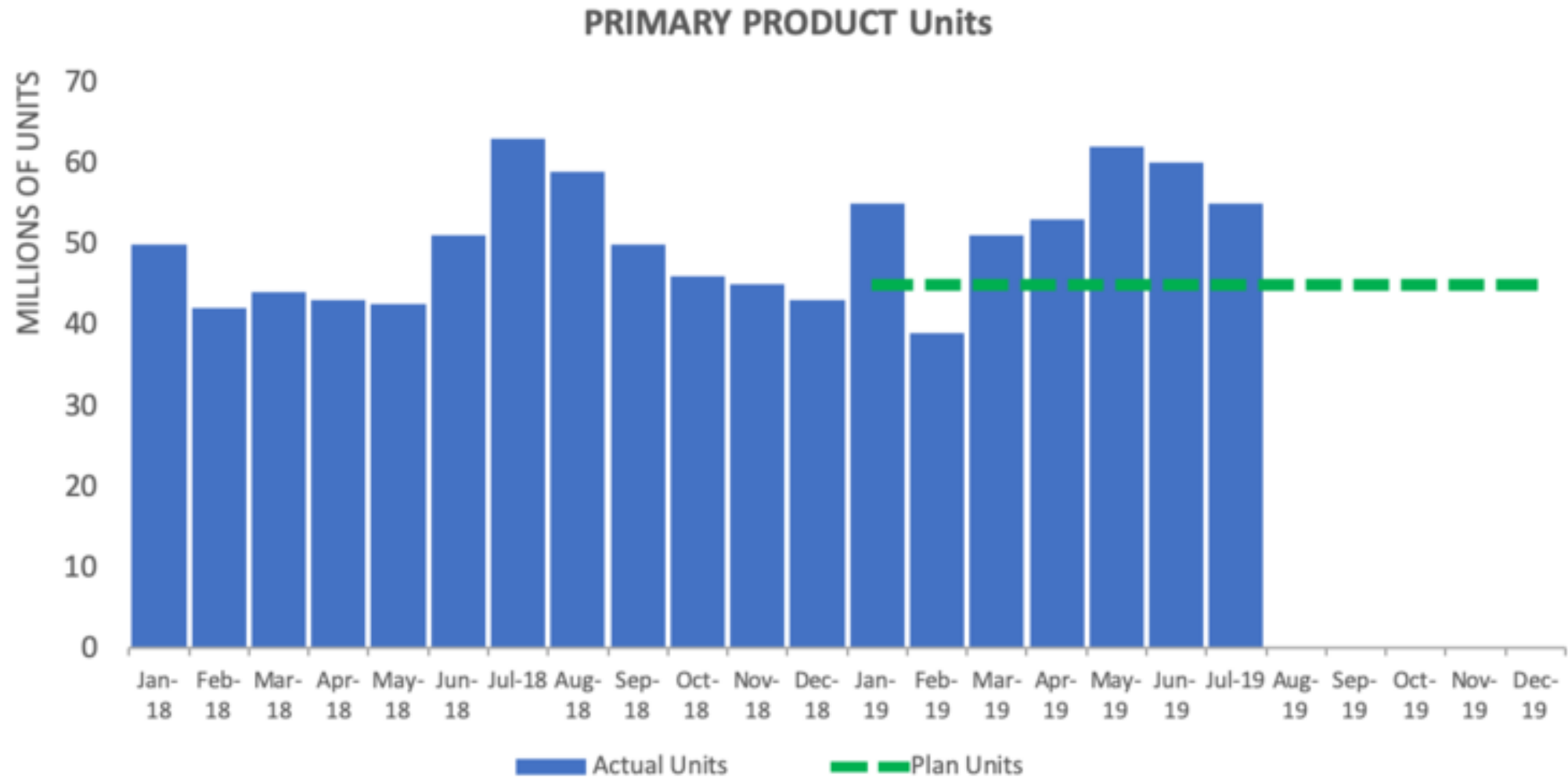
# Things to remove



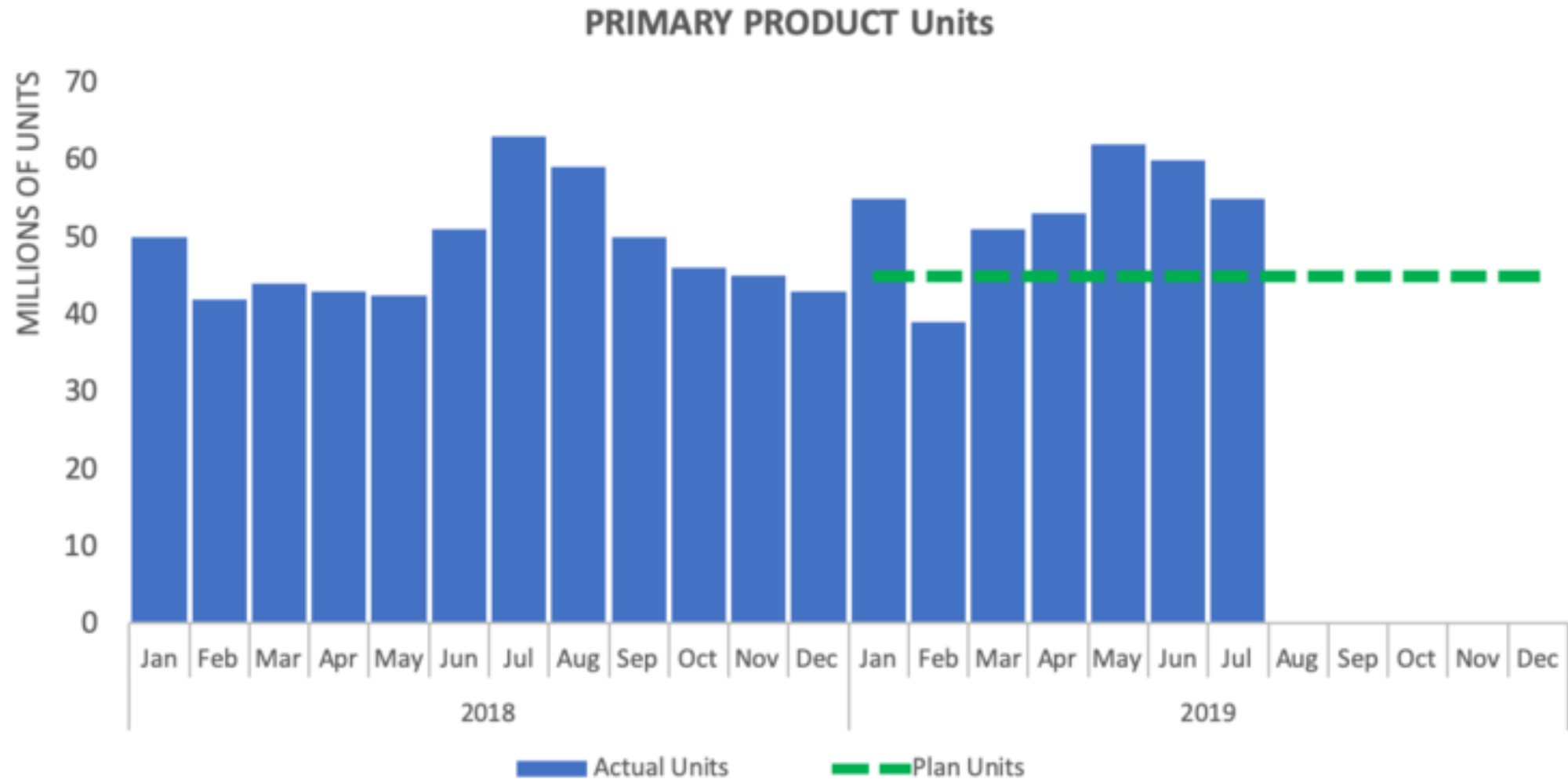
# Things to remove



# Things to remove

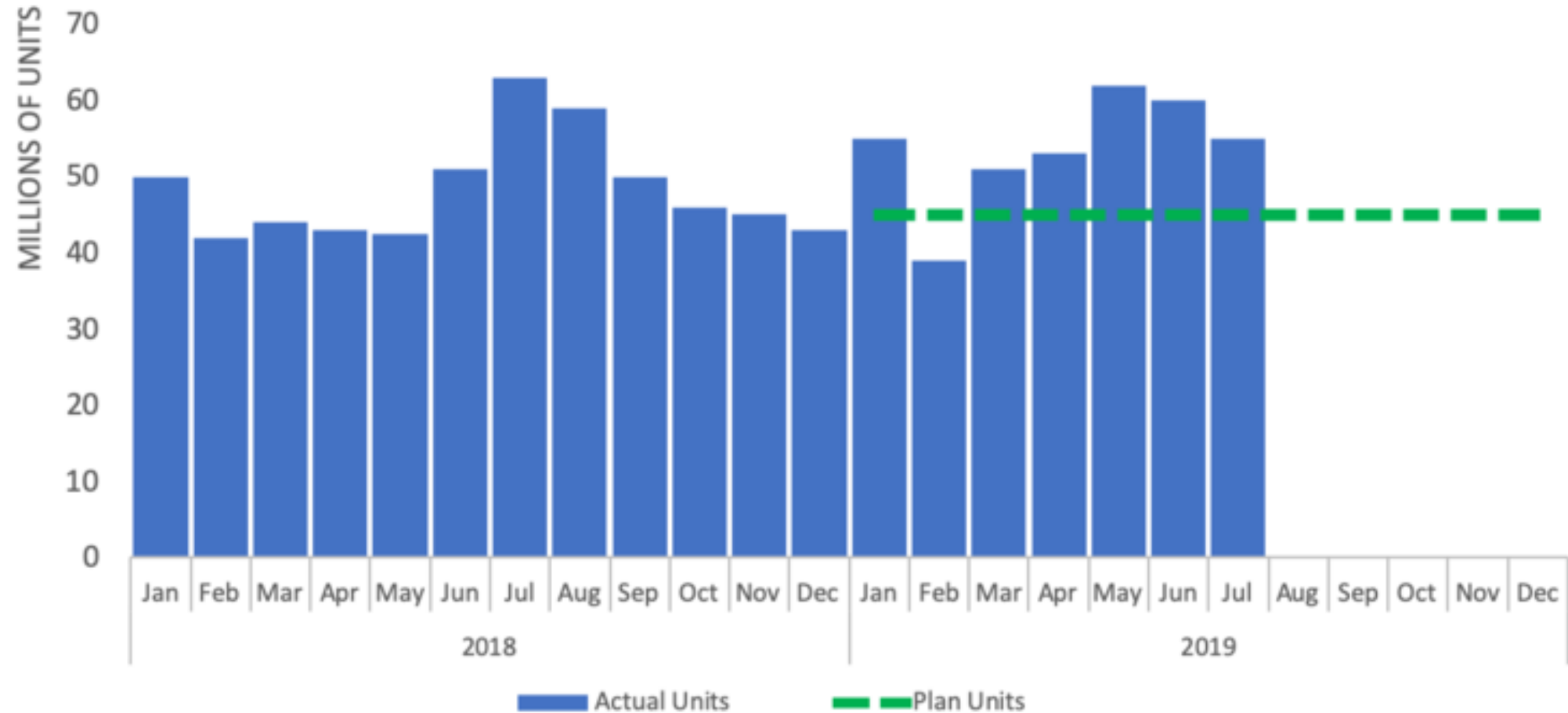


# Things to remove



# Things to remove

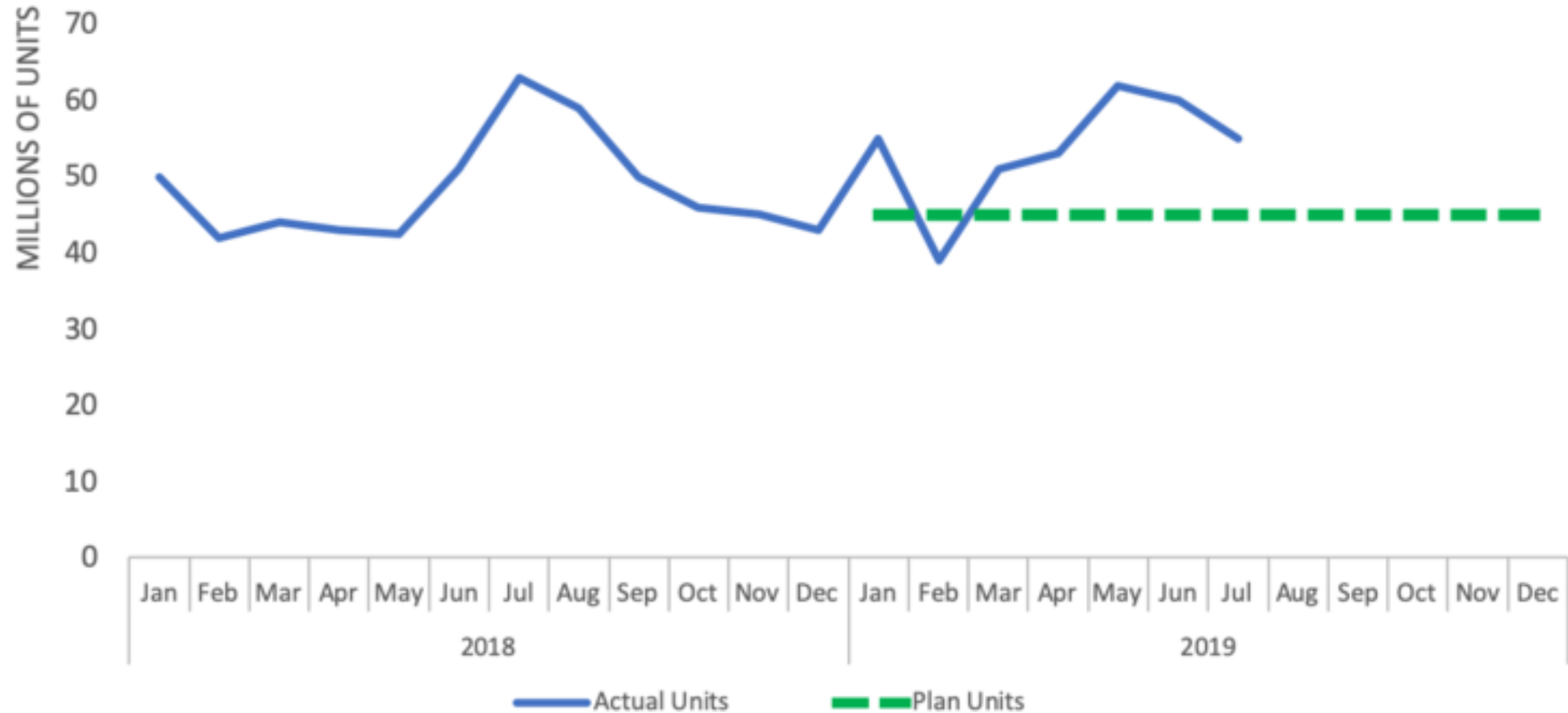
## PRIMARY PRODUCT: unit sales





# Things to remove

## PRIMARY PRODUCT: unit sales



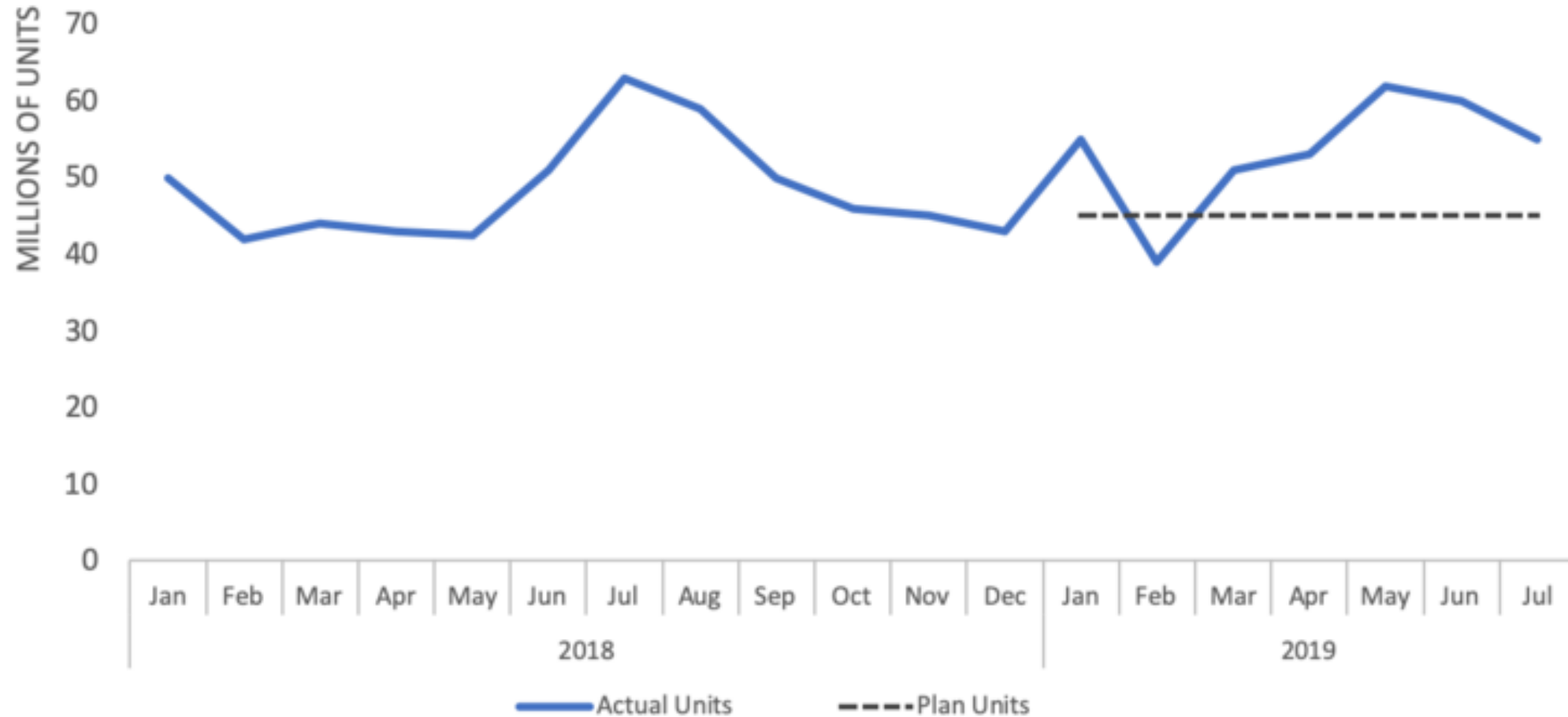
# Things to remove

## PRIMARY PRODUCT: unit sales



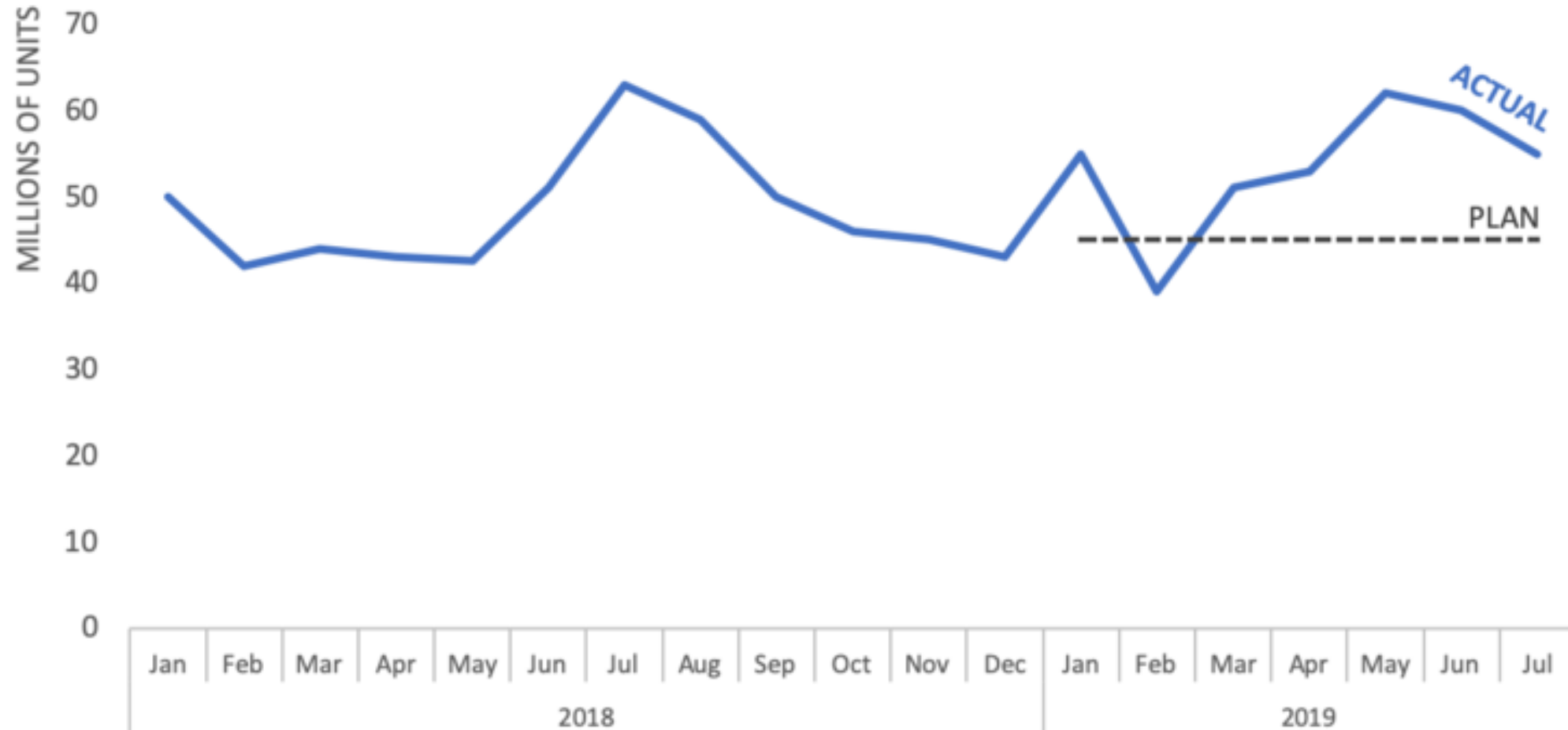
# Things to remove

## PRIMARY PRODUCT: unit sales



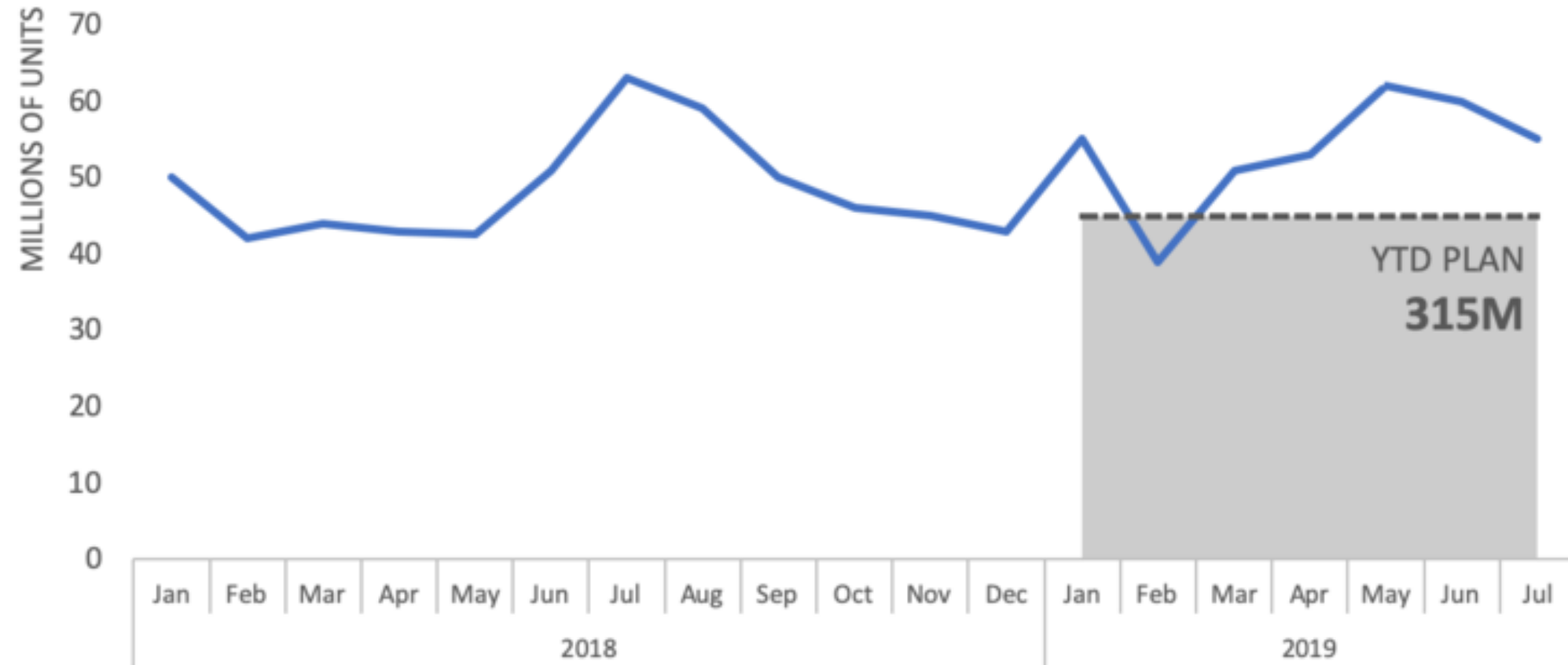
# Things to remove

PRIMARY PRODUCT: unit sales



# Things to remove

## PRIMARY PRODUCT: unit sales



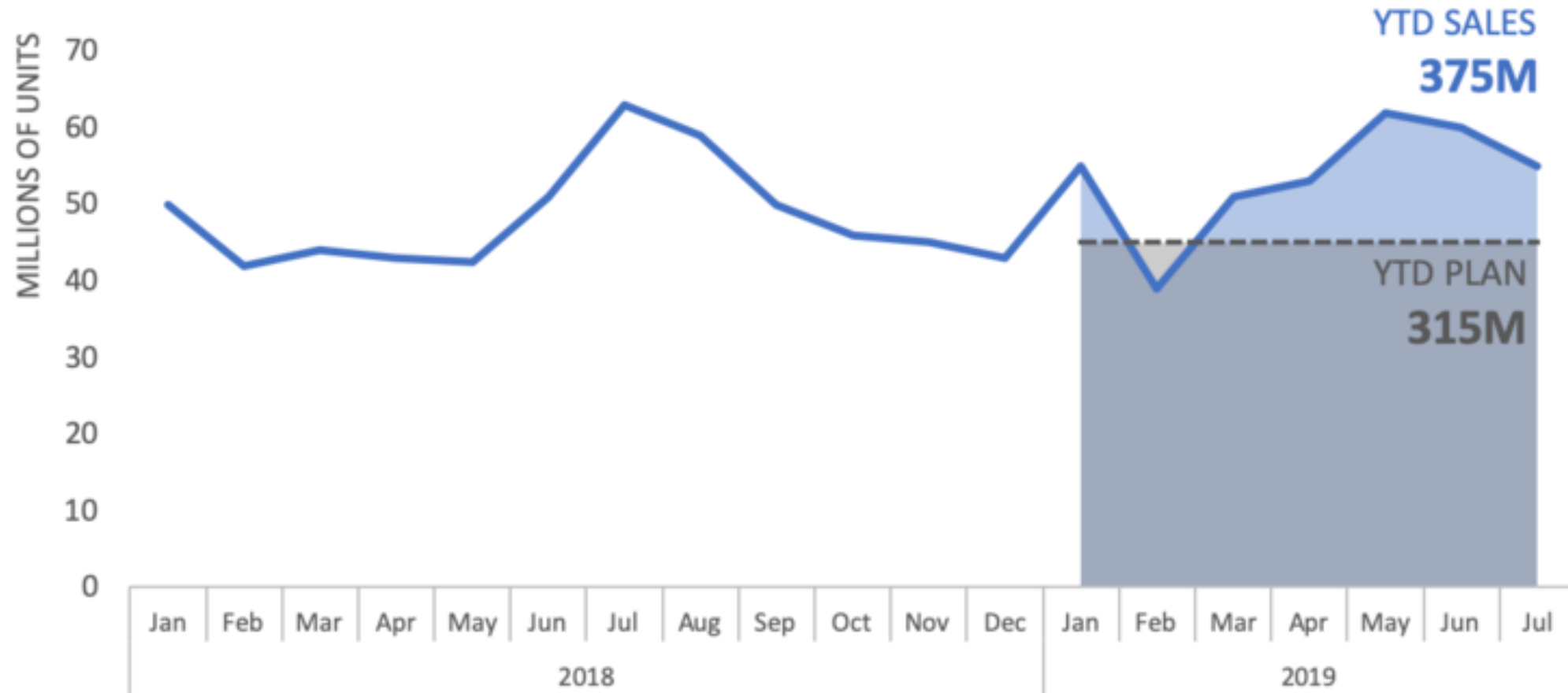
# Things to remove

PRIMARY PRODUCT: unit sales

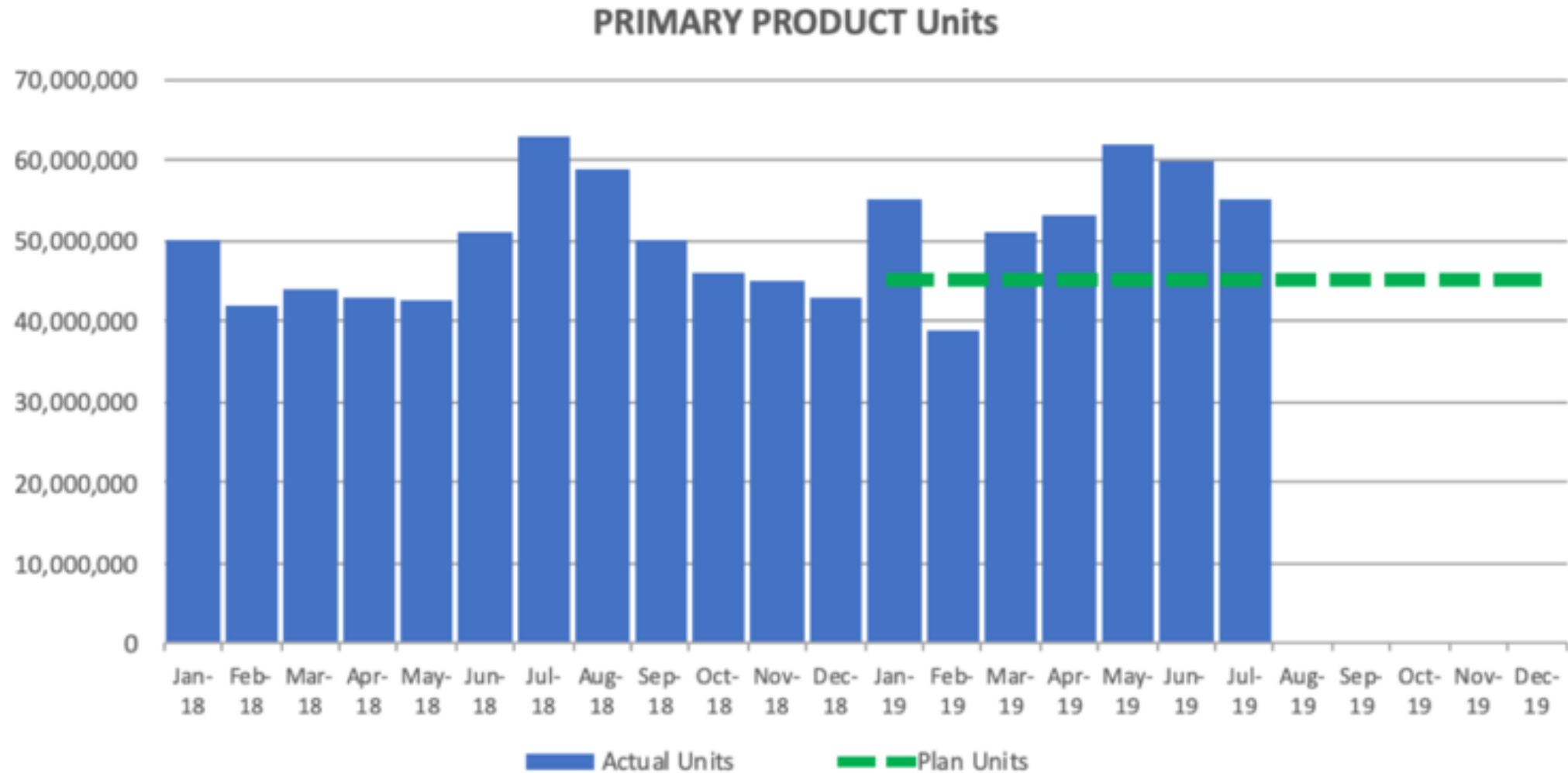


# Things to remove

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



# Things to remove



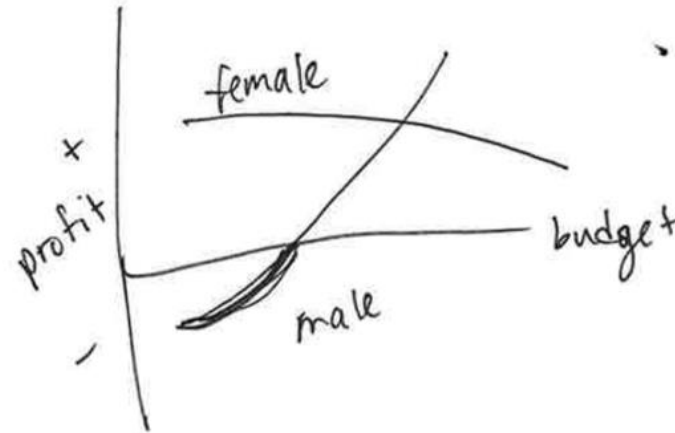
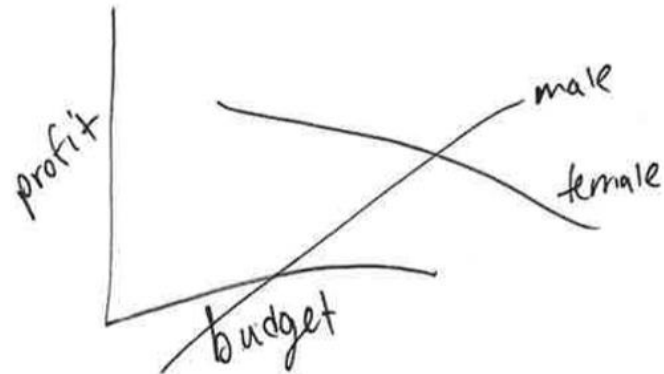
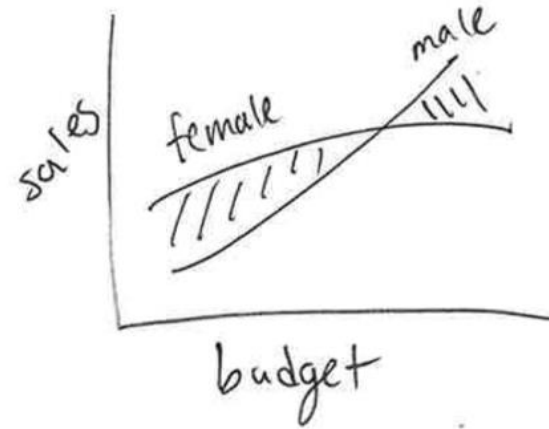
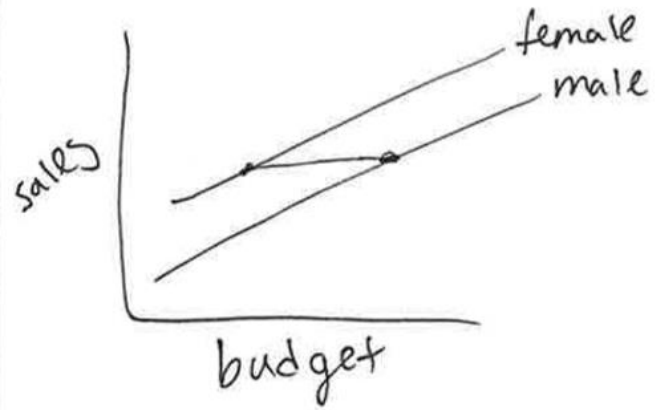


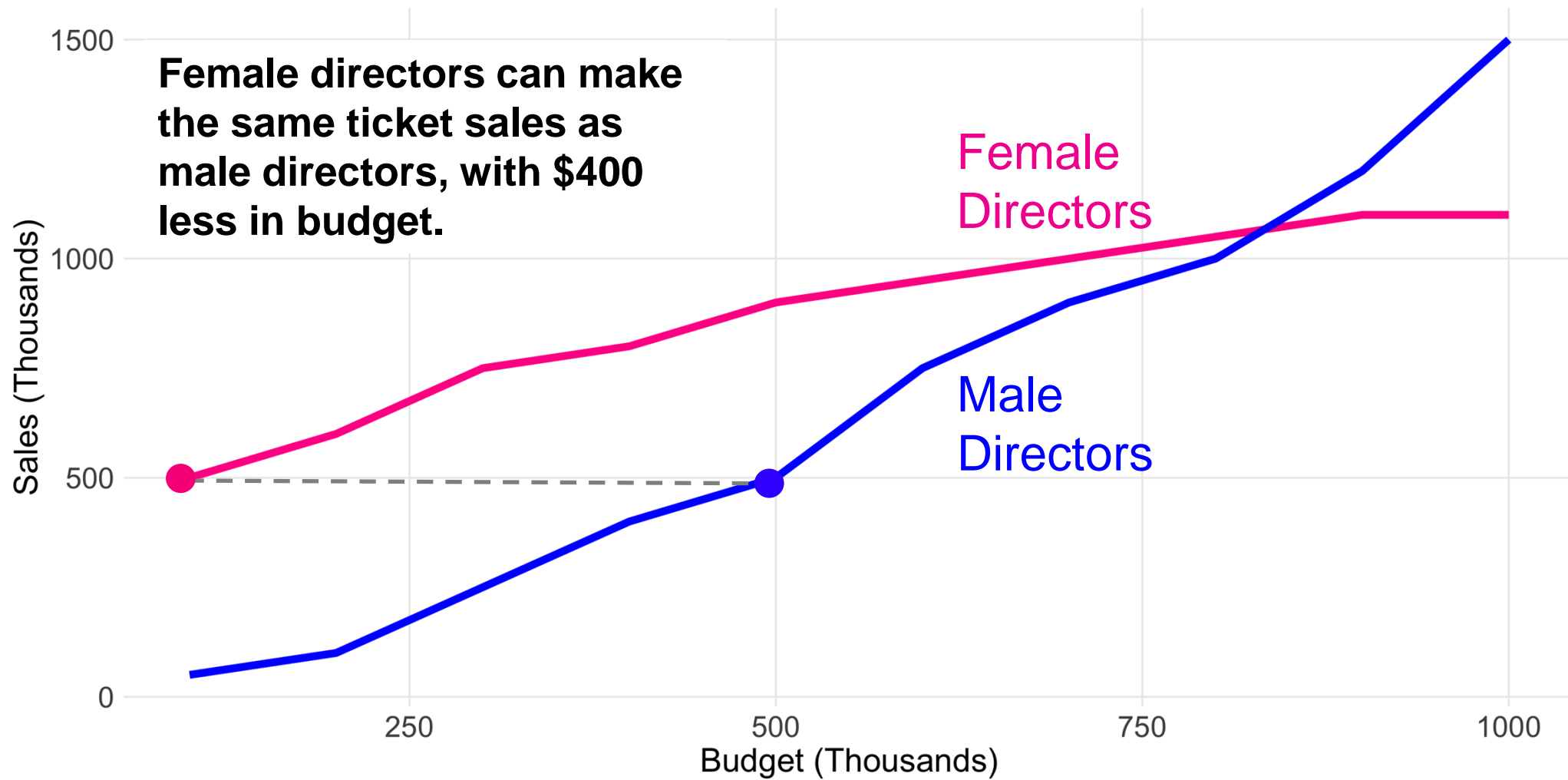
# Drawing from tabular data

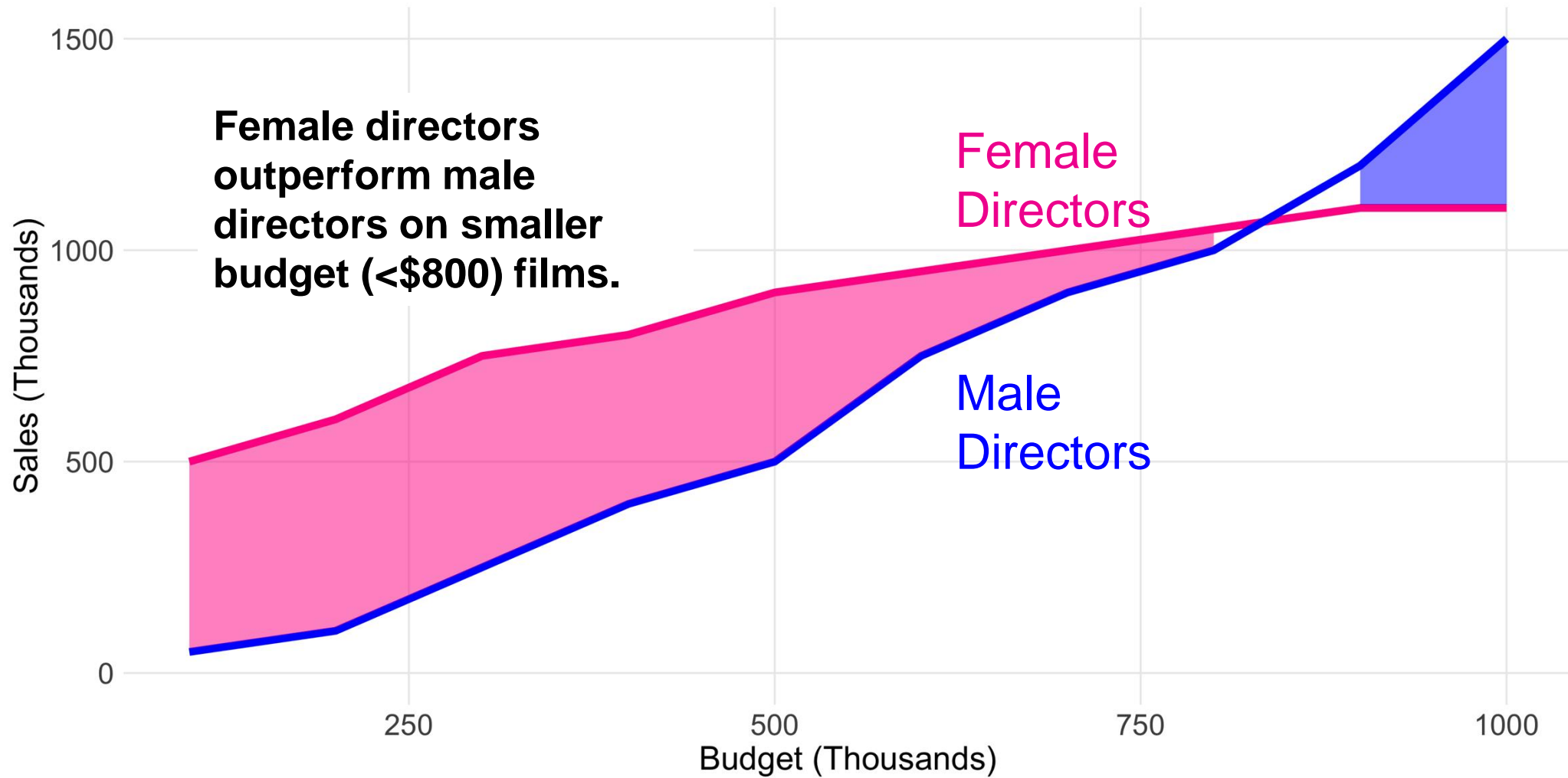
Budget (thousands)	Director Gender	Ticket Sales (thousands)
100	Female	500
200	Female	600
300	Female	750
400	Female	800
500	Female	900
600	Female	950
700	Female	1,000
800	Female	1,050
900	Female	1,100
1,000	Female	1,100

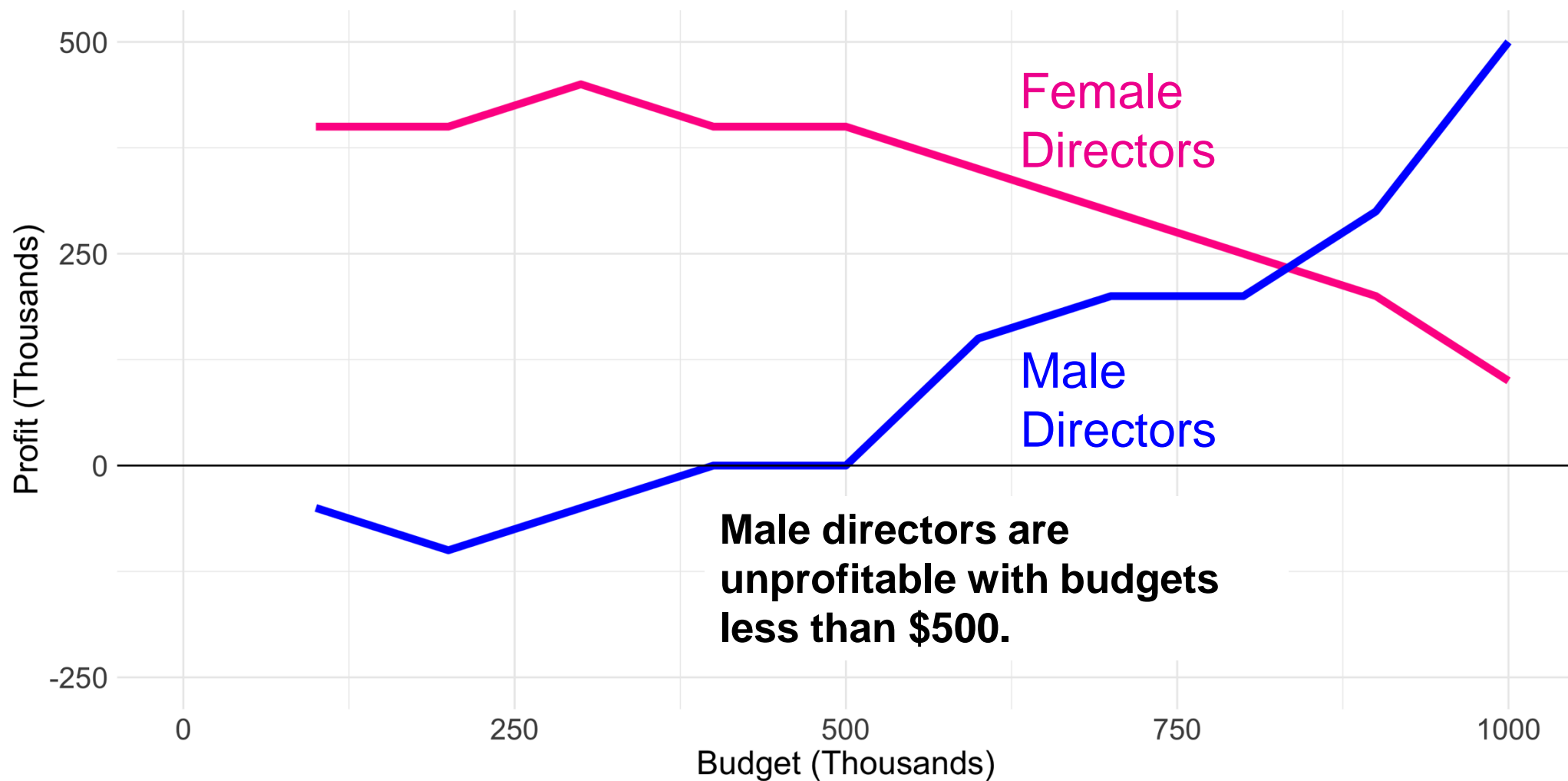
Budget (thousands)	Director Gender	Ticket Sales (thousands)
100	Male	50
200	Male	100
300	Male	250
400	Male	400
500	Male	500
600	Male	750
700	Male	900
800	Male	1,000
900	Male	1,200
1,000	Male	1,500

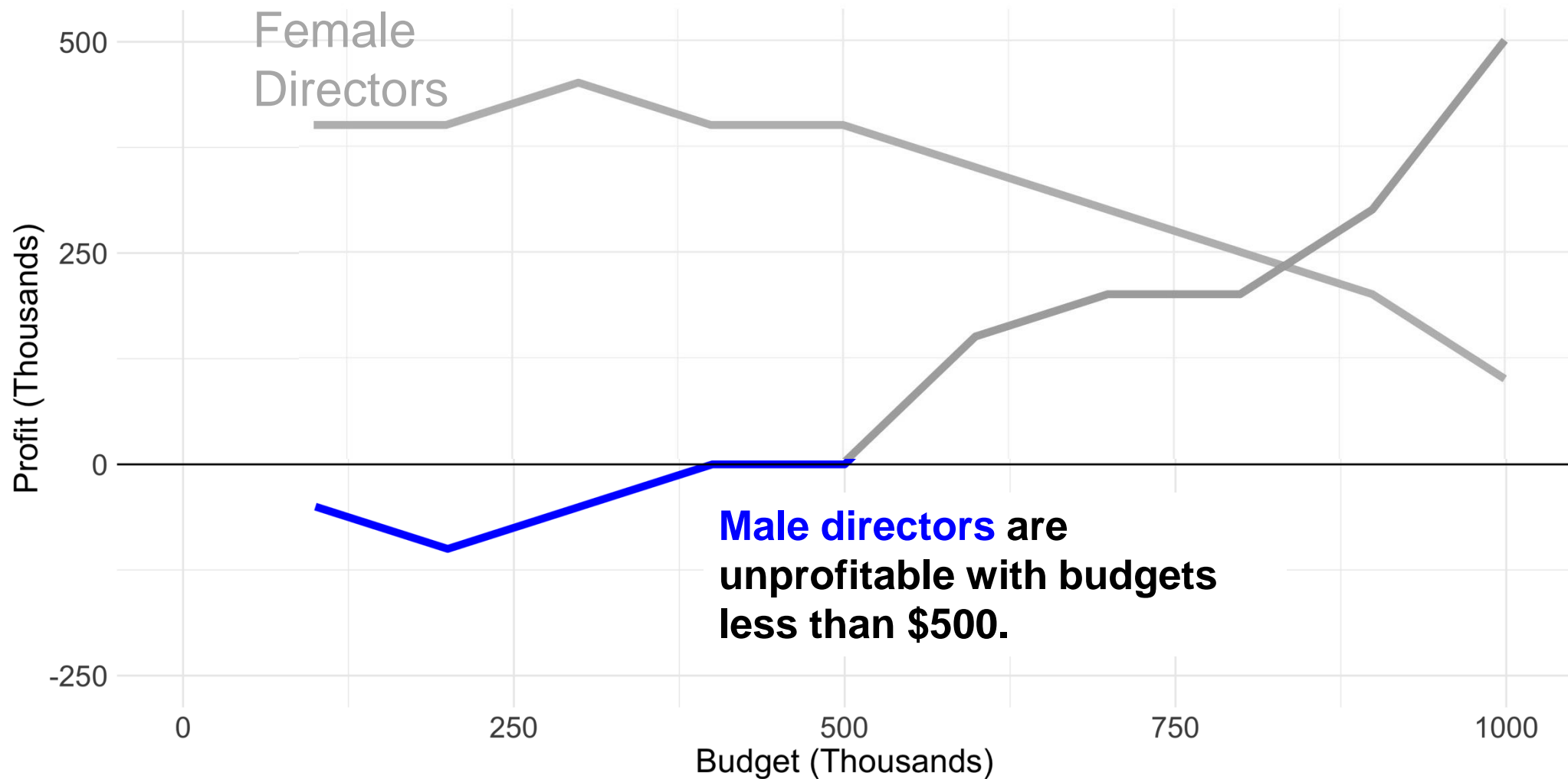
# Initial sketches of plots











**Male directors are unprofitable with budgets less than \$500.**

# Thank you!

<https://github.com/ktebbe/DataViz>



# Additional resources

- *Storytelling with Data* books - <http://www.storytellingwithdata.com/books>
- Free e-book on visualization- <https://serialmentor.com/dataviz/>
- Intro to *ggplot* in R - [https://www.williamrchase.com/slides/ggplot\\_intro.html](https://www.williamrchase.com/slides/ggplot_intro.html) and [https://evamaerey.github.io/ggplot\\_flipbook/ggplot\\_flipbook\\_xaringan.html#1](https://evamaerey.github.io/ggplot_flipbook/ggplot_flipbook_xaringan.html#1)
- *ggplot* examples <https://wilkelab.org/practicalgg>
- Slide deck on data communication - <https://www.amelia.mn/TeachingDataCommunication.pdf>