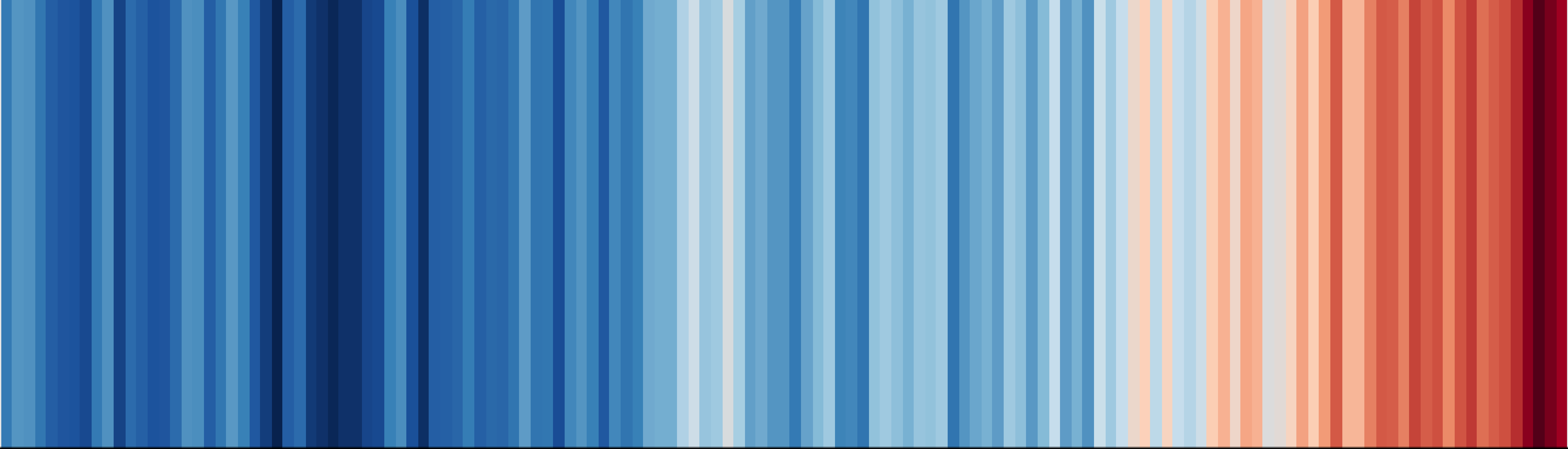


Week 5: *Visualizing Information*

EMSE 4575: Exploratory Data Analysis

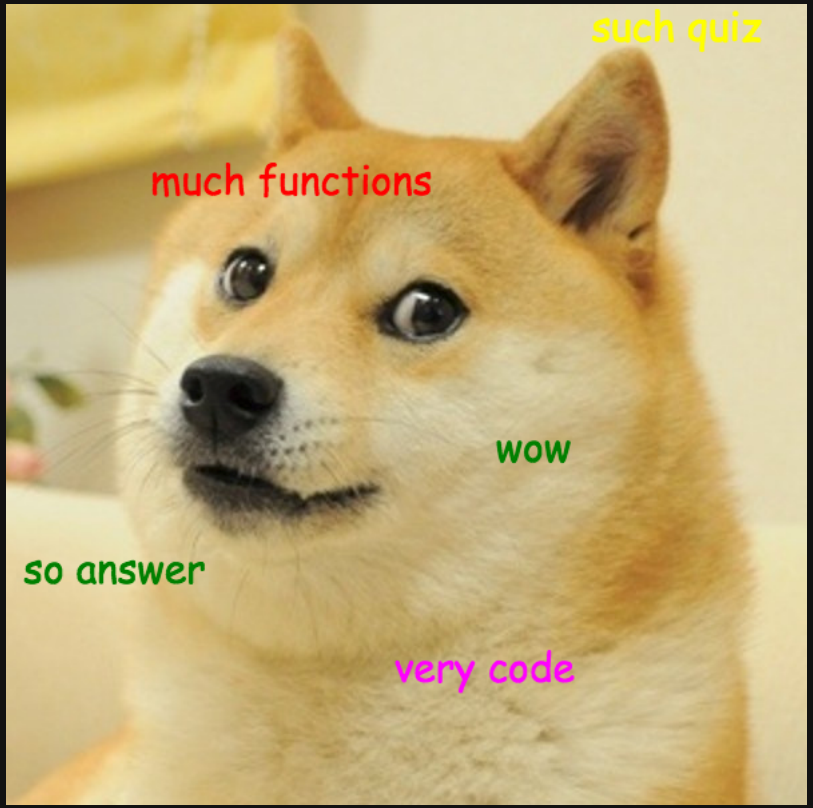
John Paul Helveston

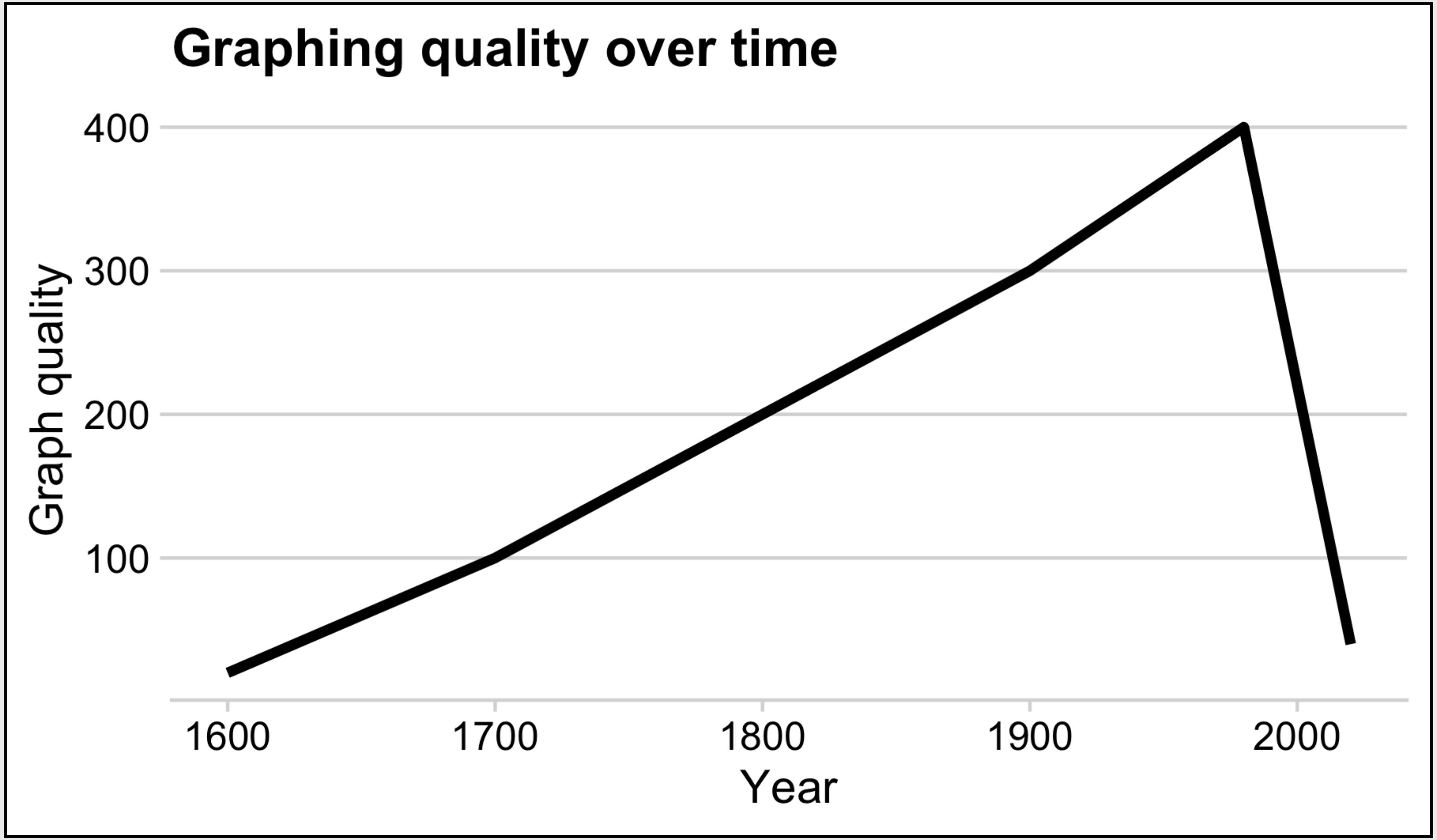
February 10, 2021

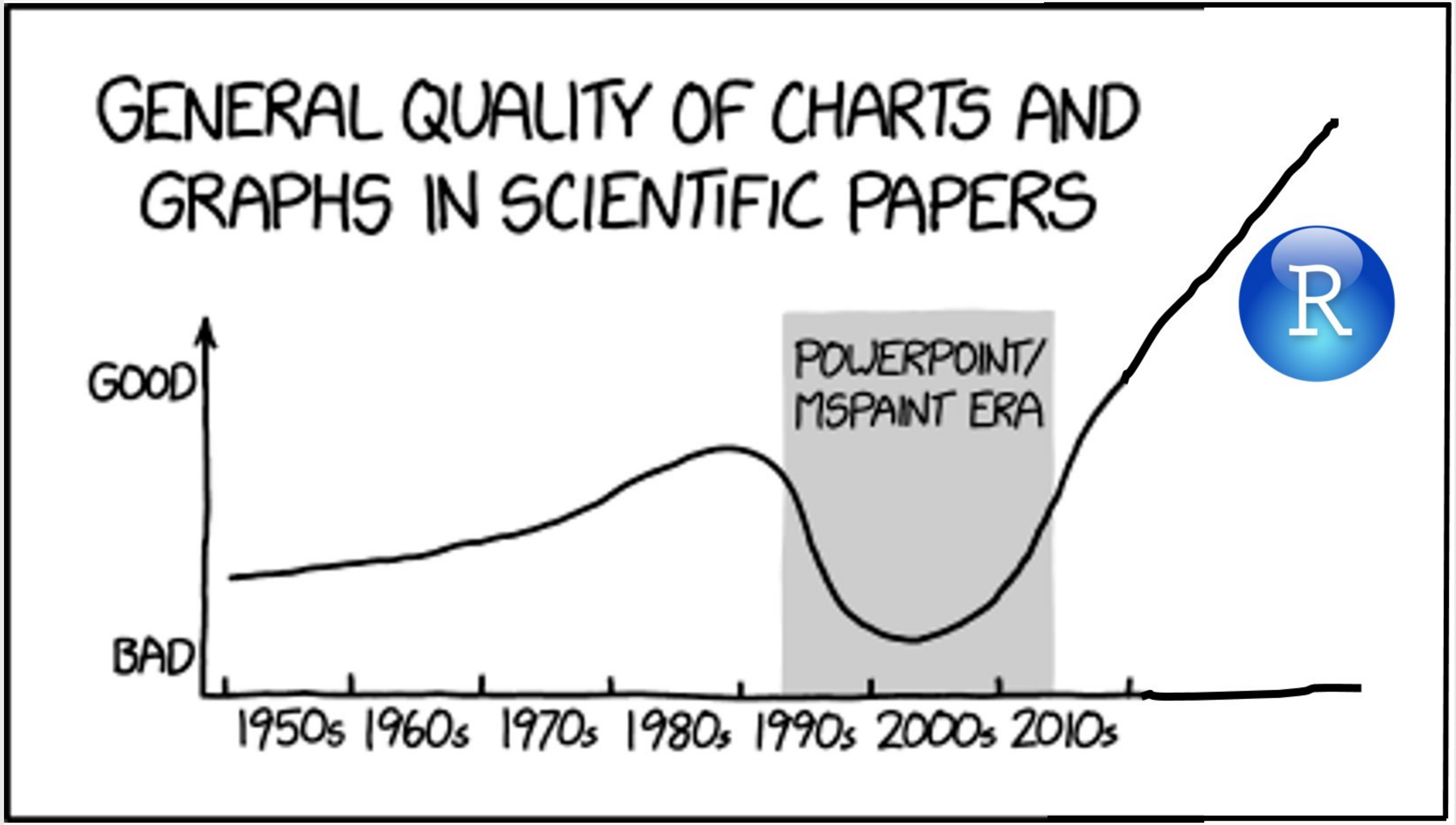
Quiz 1

- Go to `#classroom` channel in Slack for link
- Open up RStudio before you start - you'll probably want to use it.

05:00




Graphing quality over time
400
300
200
100
Graph quality
1600
1700
1800
1900
2000
Year


GENERAL QUALITY OF CHARTS AND GRAPHS IN SCIENTIFIC PAPERS
GOOD
BAD
POWERPOINT/ MSPAINT ERA
R
1950s 1960s 1970s 1980s 1990s 2000s 2010s

From here

"Having word processing software doesn't make us great writers."

— Stephen Few

We don't write paragraphs like this

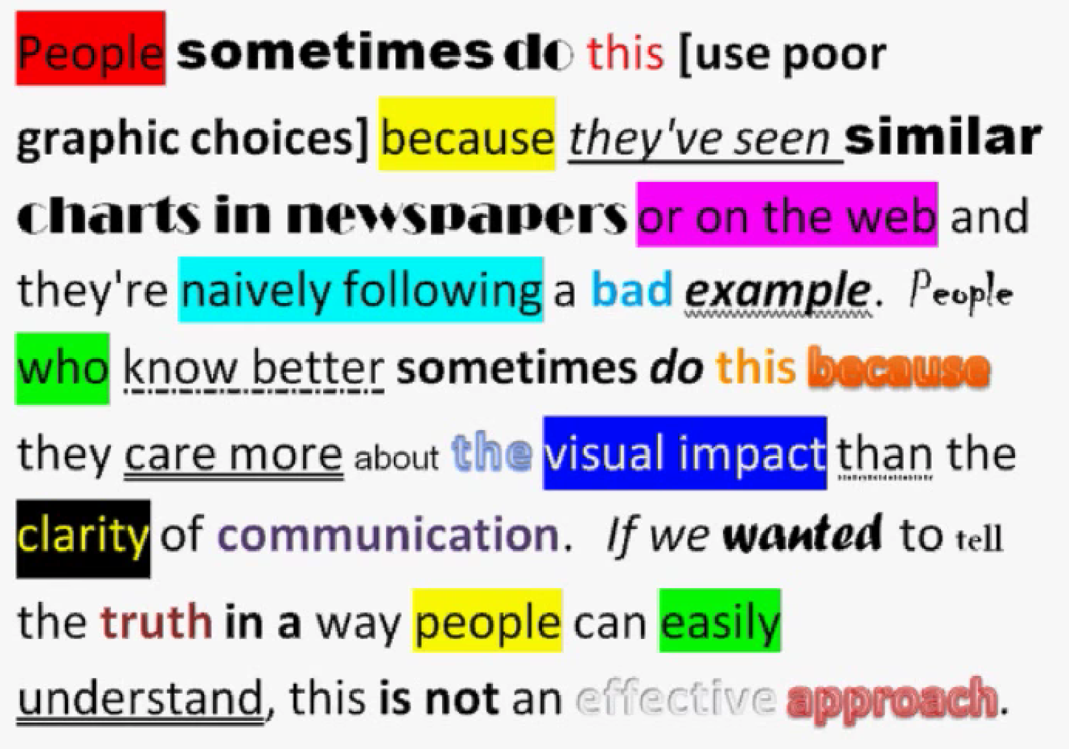


Image from Few (2012, pg. 227)

So don't make graphs like this

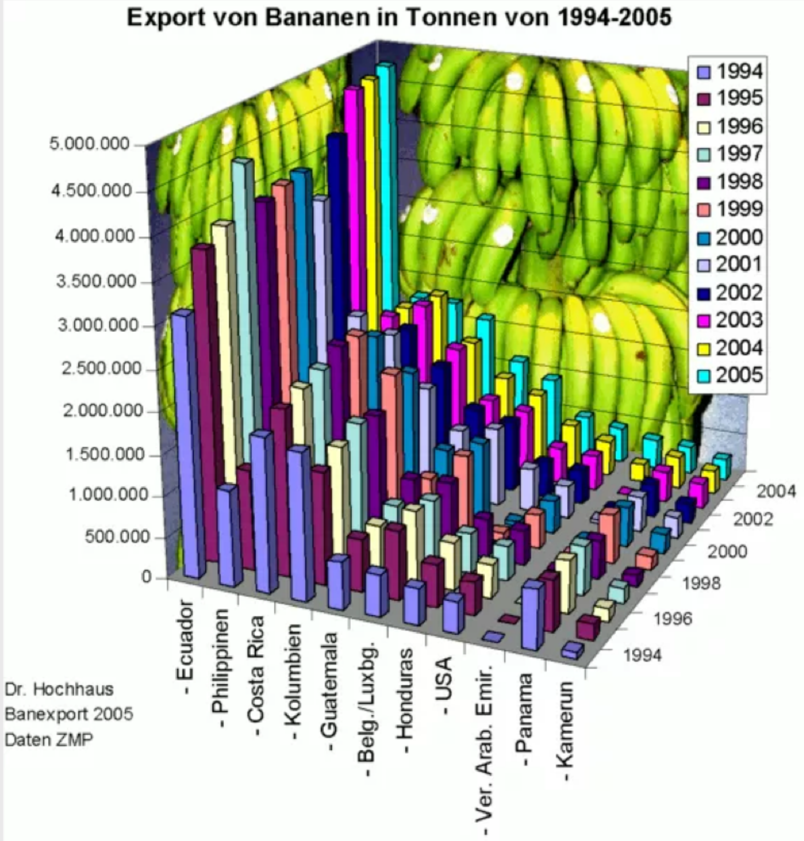


Image from excelcharts.com

Week 5: *Visualizing Information*

1. The Human Visual-Memory System

2. The Psychology of Data Viz

BREAK

3. 10 Data Viz Best Practices

4. Making a (good) ggplot

Week 5: *Visualizing Information*

1. The Human Visual-Memory System
2. The Psychology of Data Viz

BREAK

3. 10 Data Viz Best Practices
4. Making a (good) ggplot

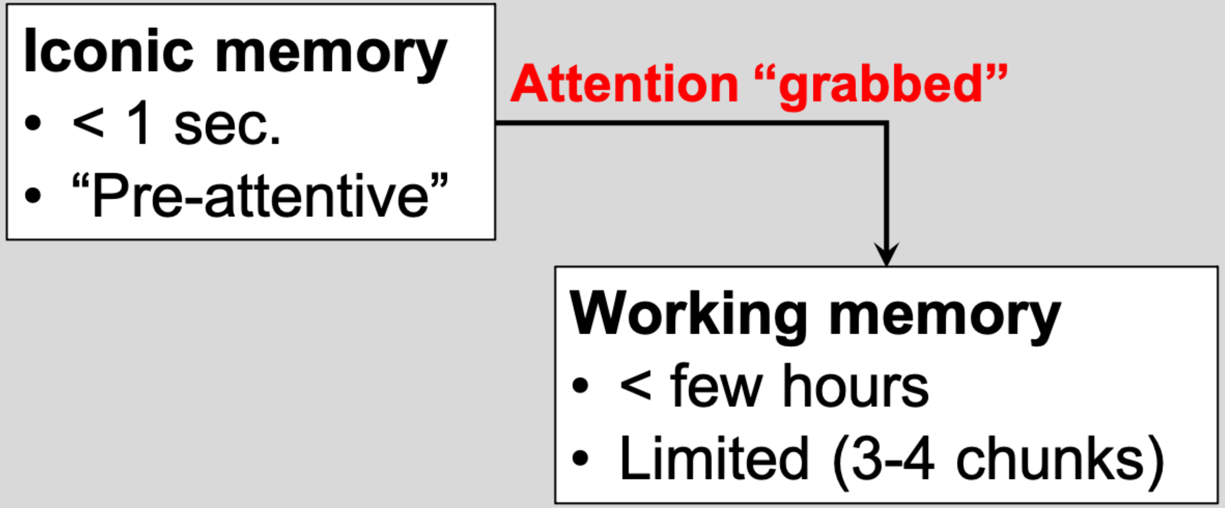
Good visualizations optimize for
the human visual-memory system

A (very) simplified model of the visual-memory system

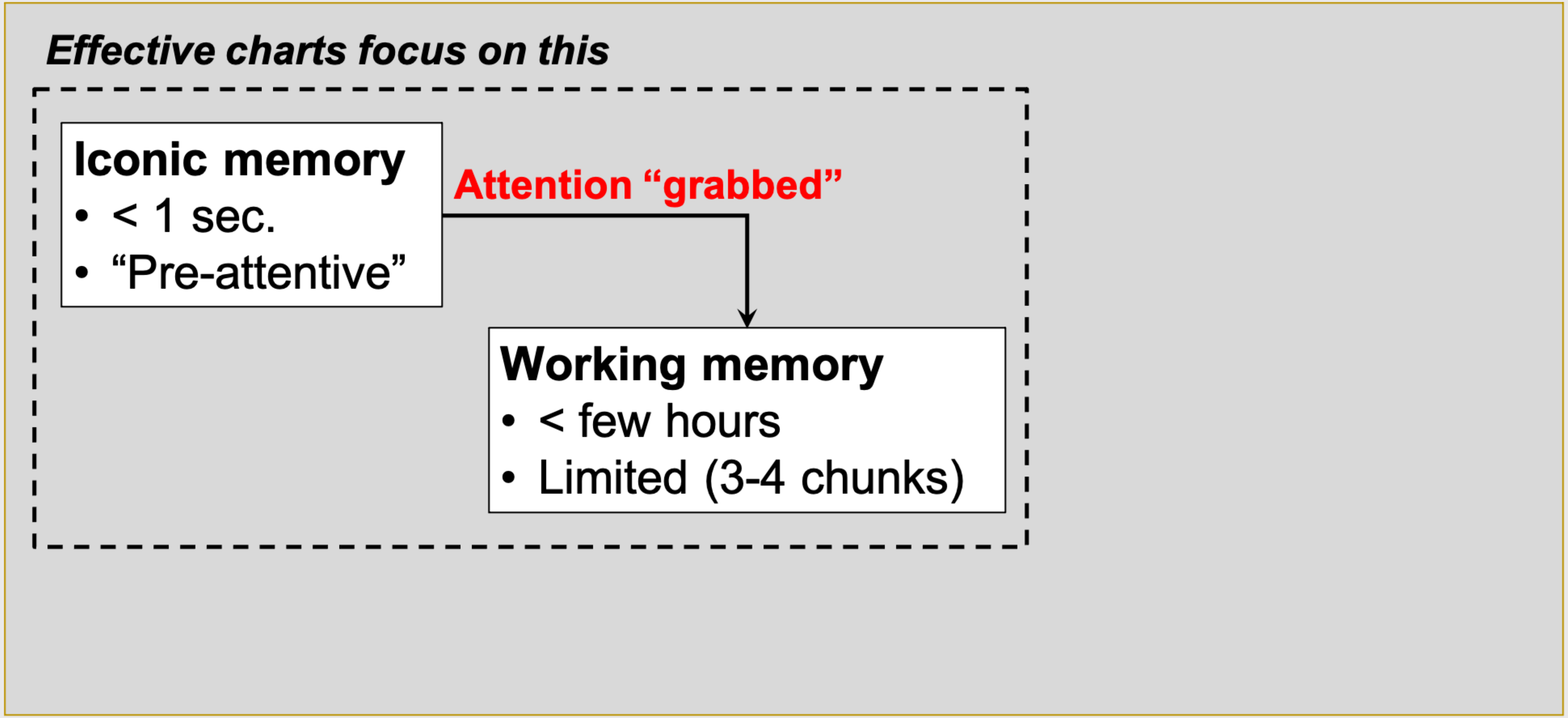
Iconic memory

- < 1 sec.
- “Pre-attentive”

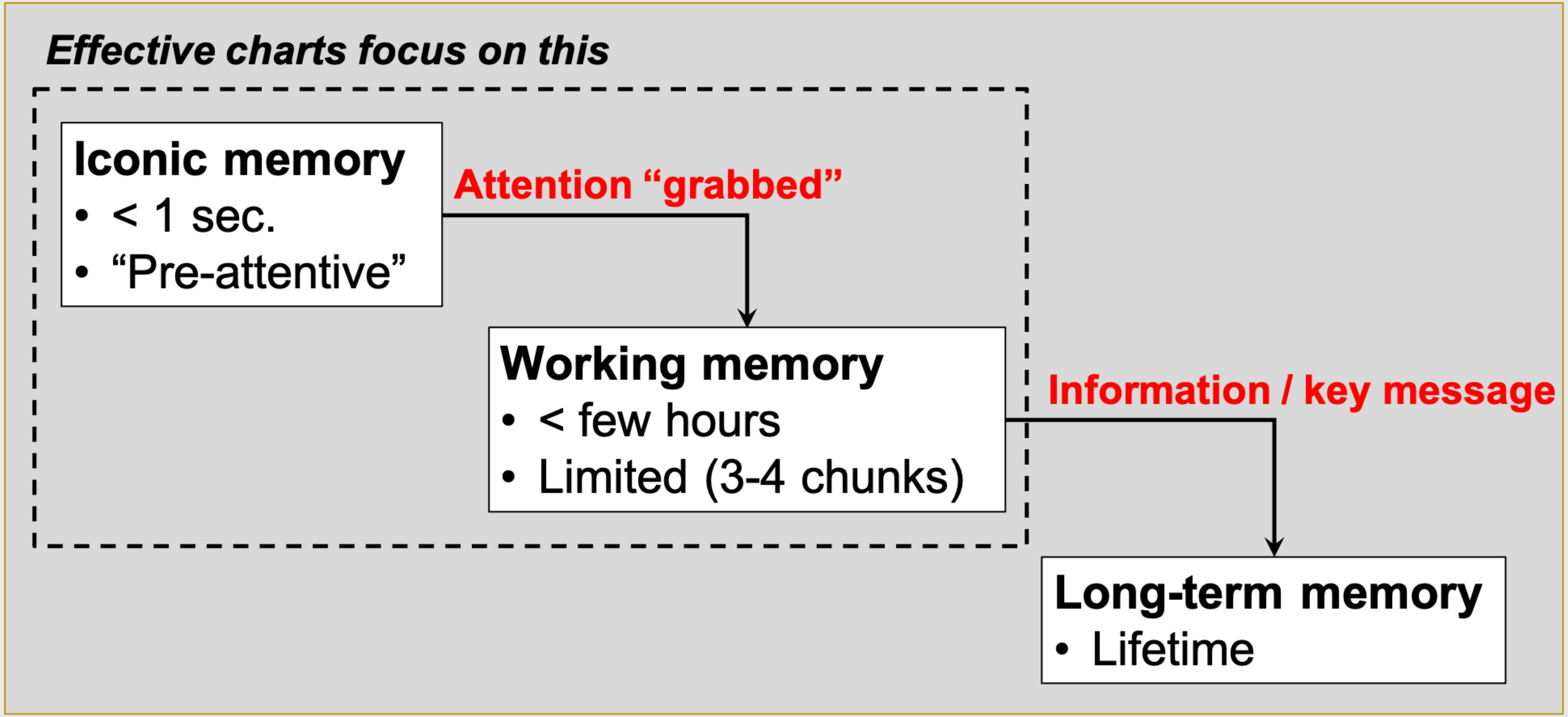
A (very) simplified model of the visual-memory system



A (very) simplified model of the visual-memory system



A (very) simplified model of the visual-memory system



Two objectives of effective charts:

1. Grab & direct attention (iconic memory)
2. Reduce processing demands (working memory)

The power of pre-attentive processing

Count all the "5"'s

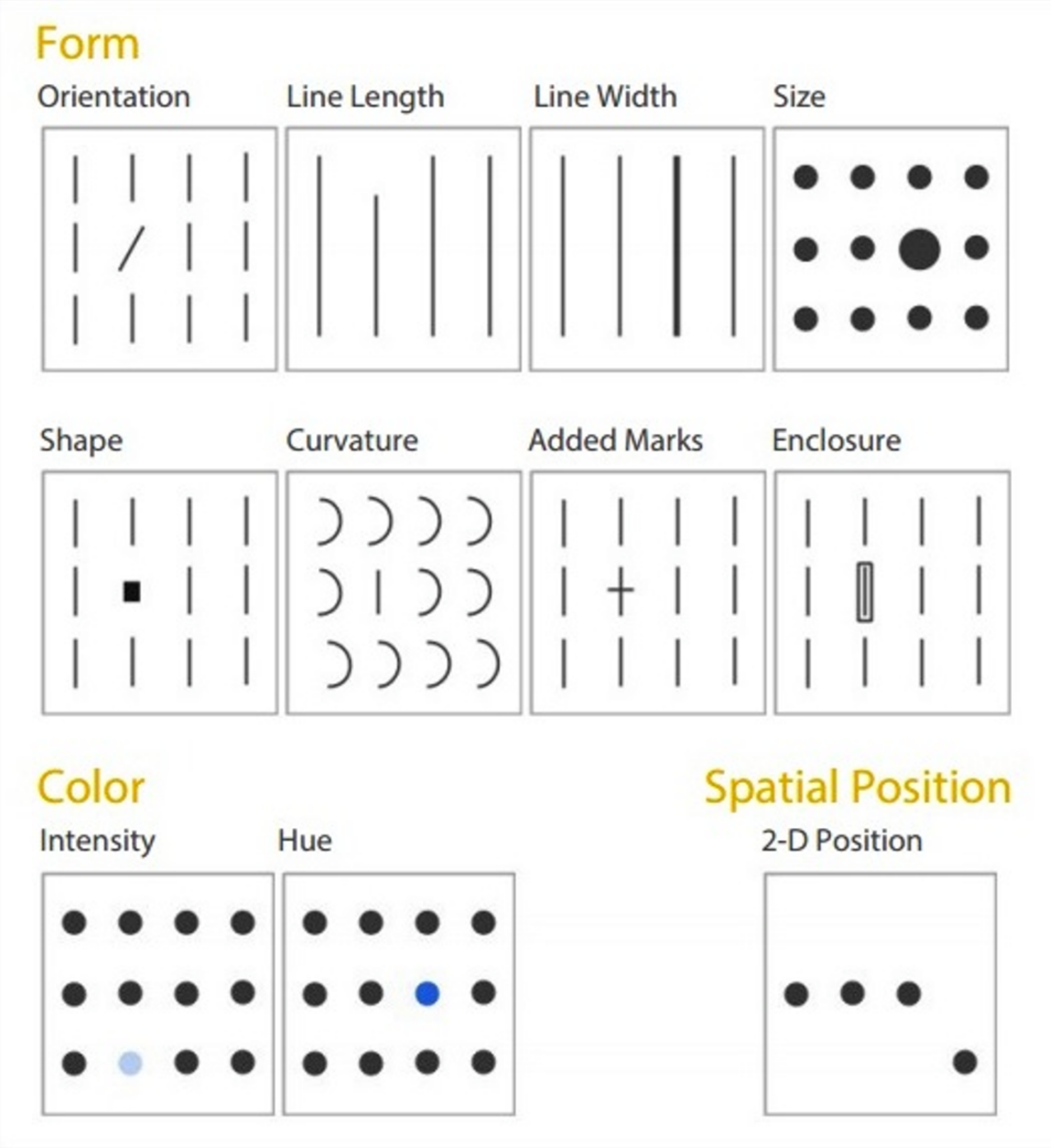
821134907856412043612
304589640981709812734
123450986124790812734
029860192837401489363
123479827961203459816
234009816256908127634
123459087162342015237
123894789237498230192

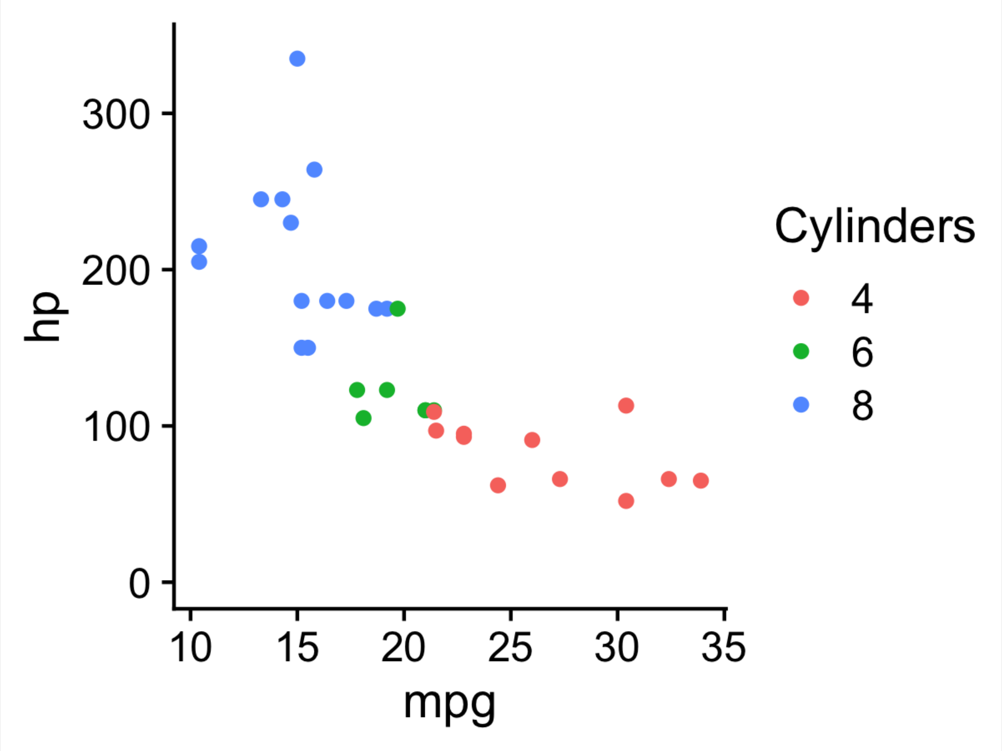
The power of pre-attentive processing

Count all the "5"'s

```
821134907856412043612
304589640981709812734
123450986124790812734
029860192837401489363
123479827961203459816
234009816256908127634
123459087162342015237
123894789237498230192
```

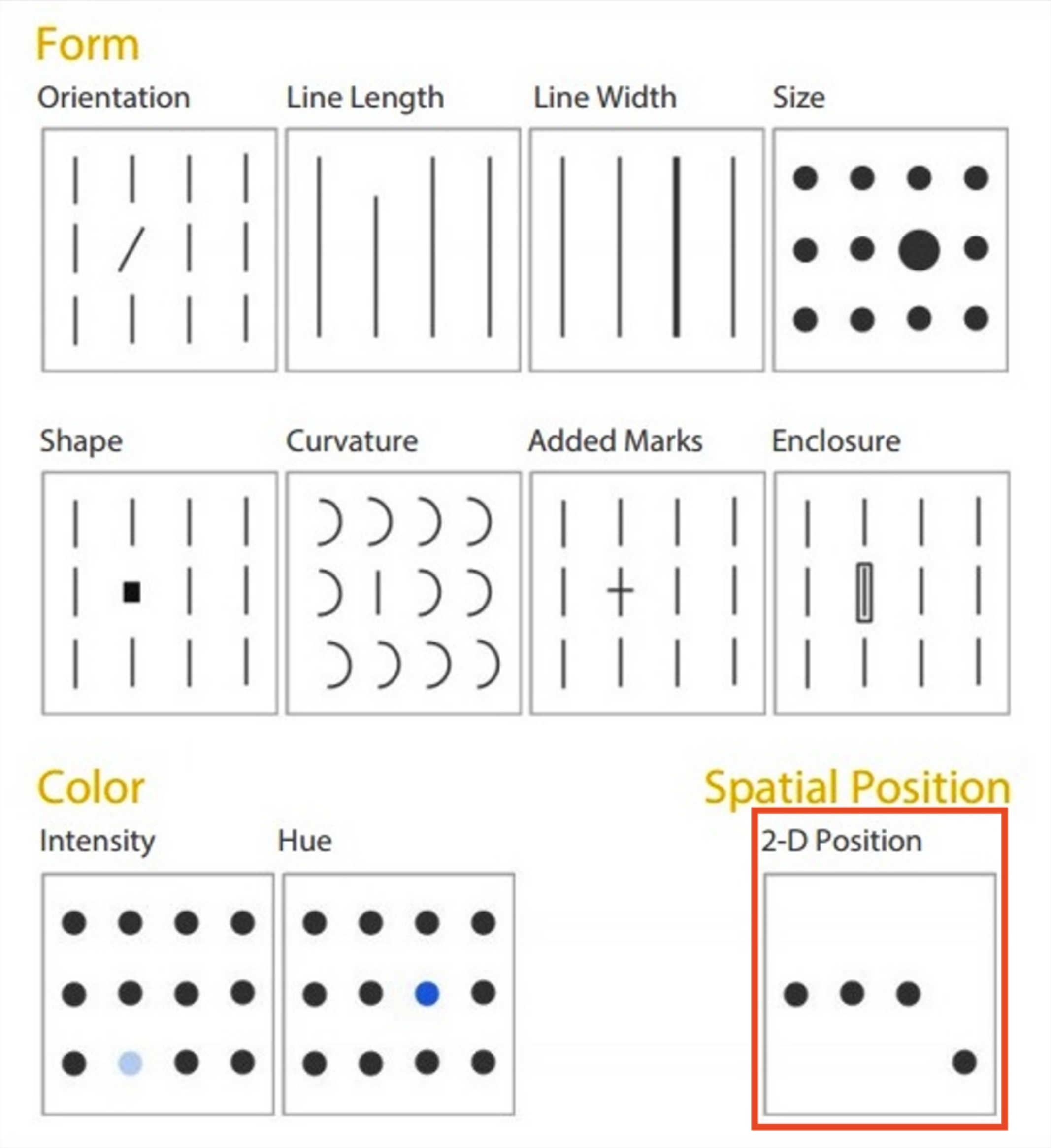
Pre-attentive attributes

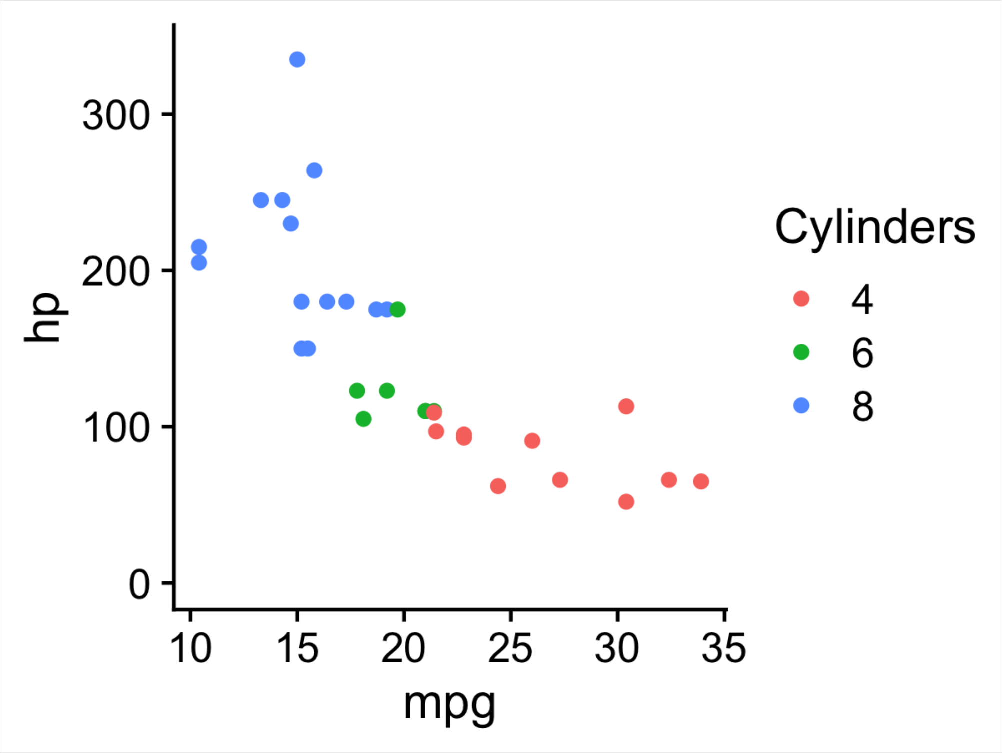




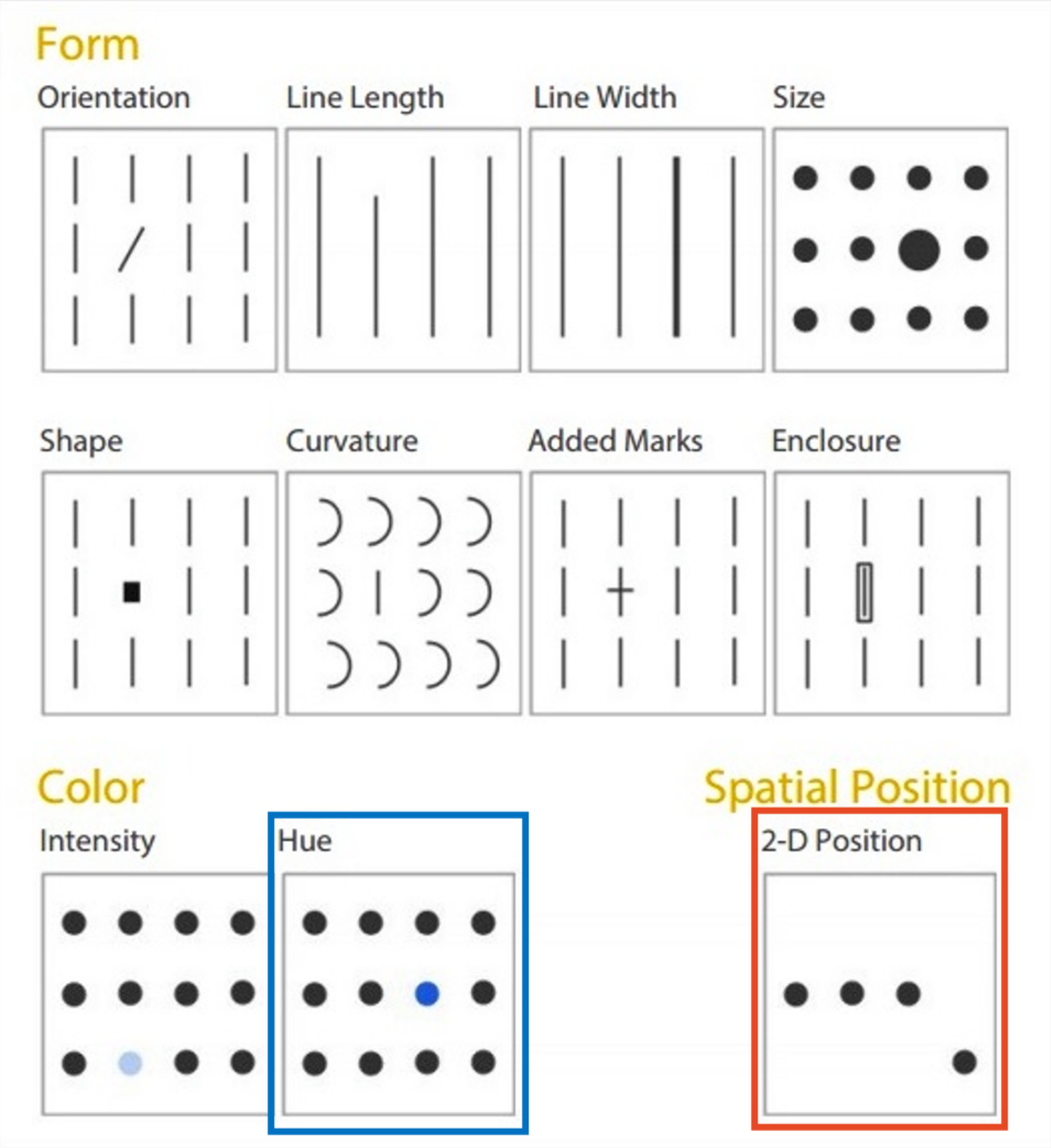
Pre-attentive attributes

Numerical (ratio) data

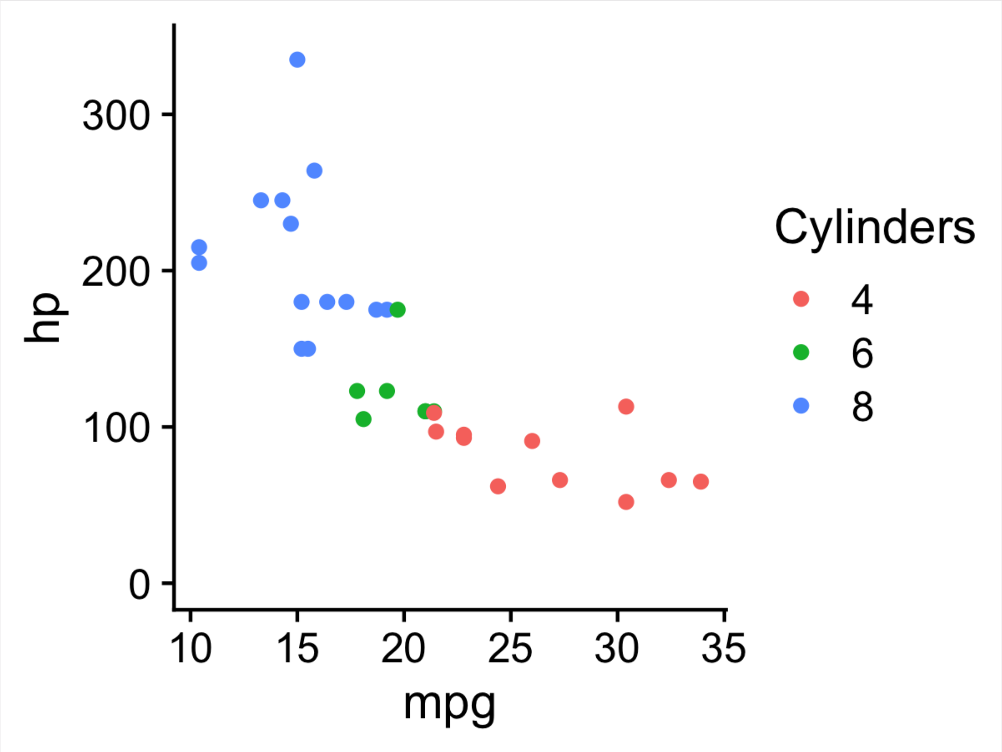




Pre-attentive attributes

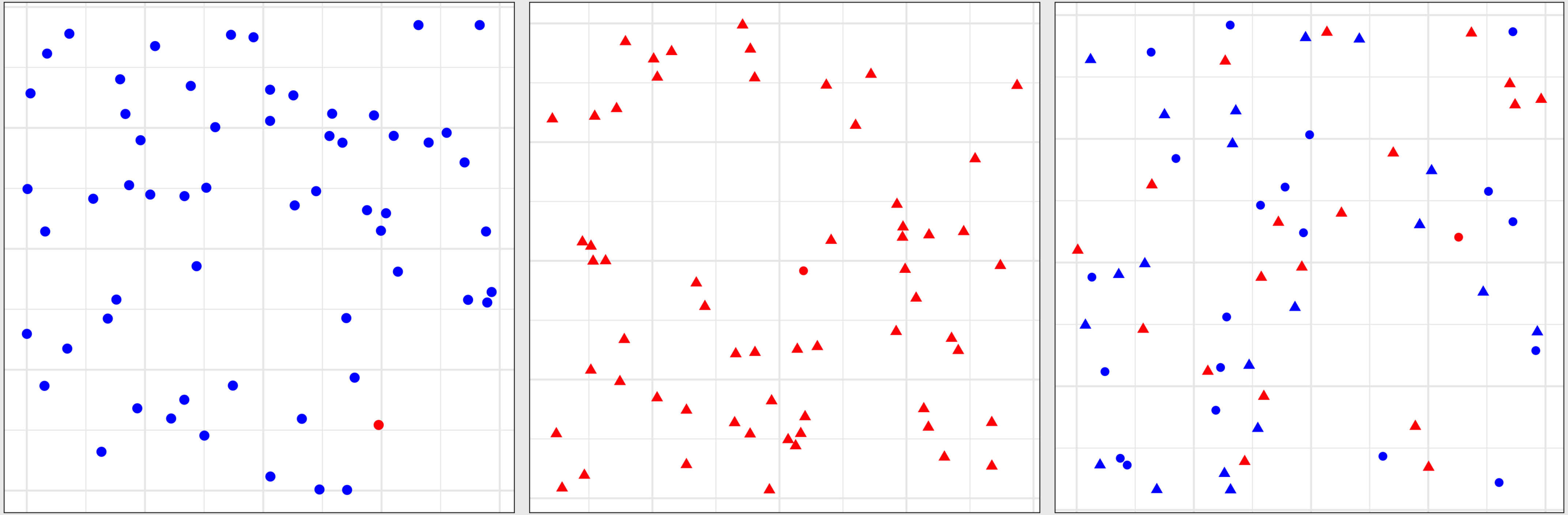


Numerical (ratio) data
Categorical (ordinal) data



Not all pre-attentive attributes are equal

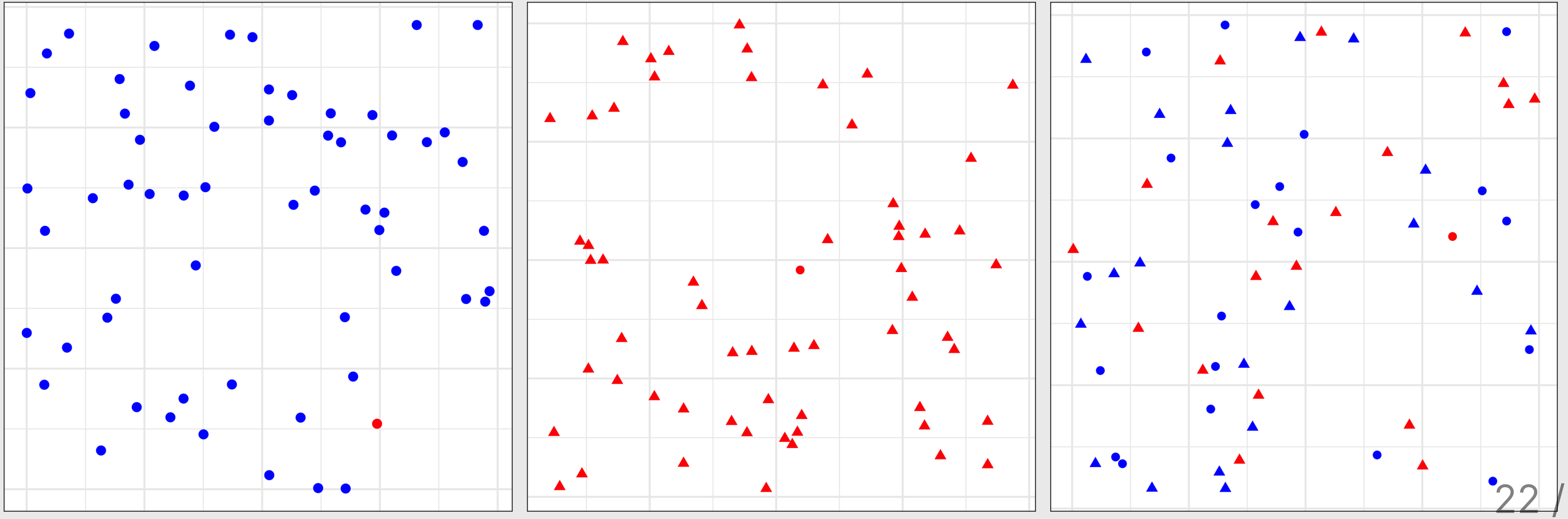
Where is the red dot?

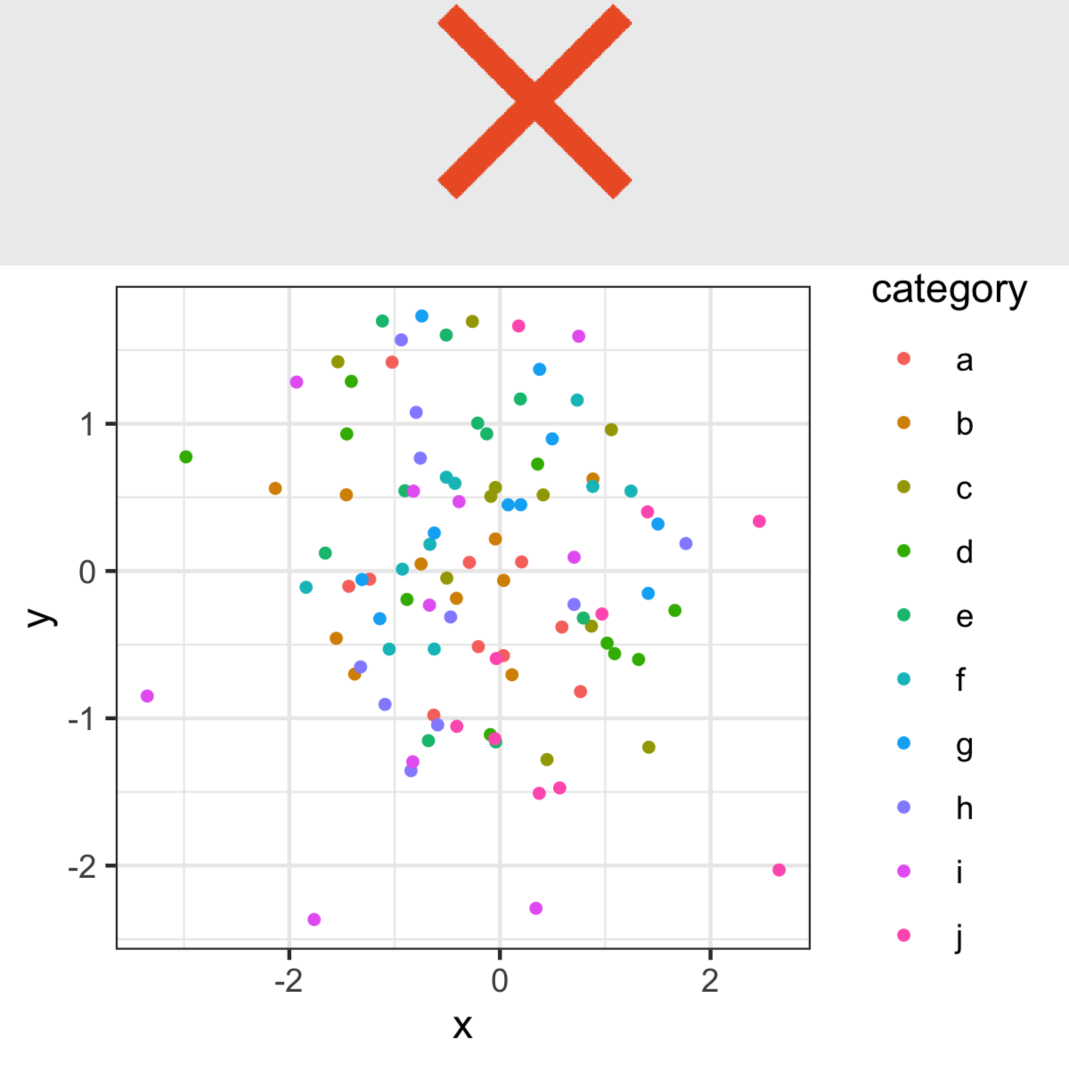


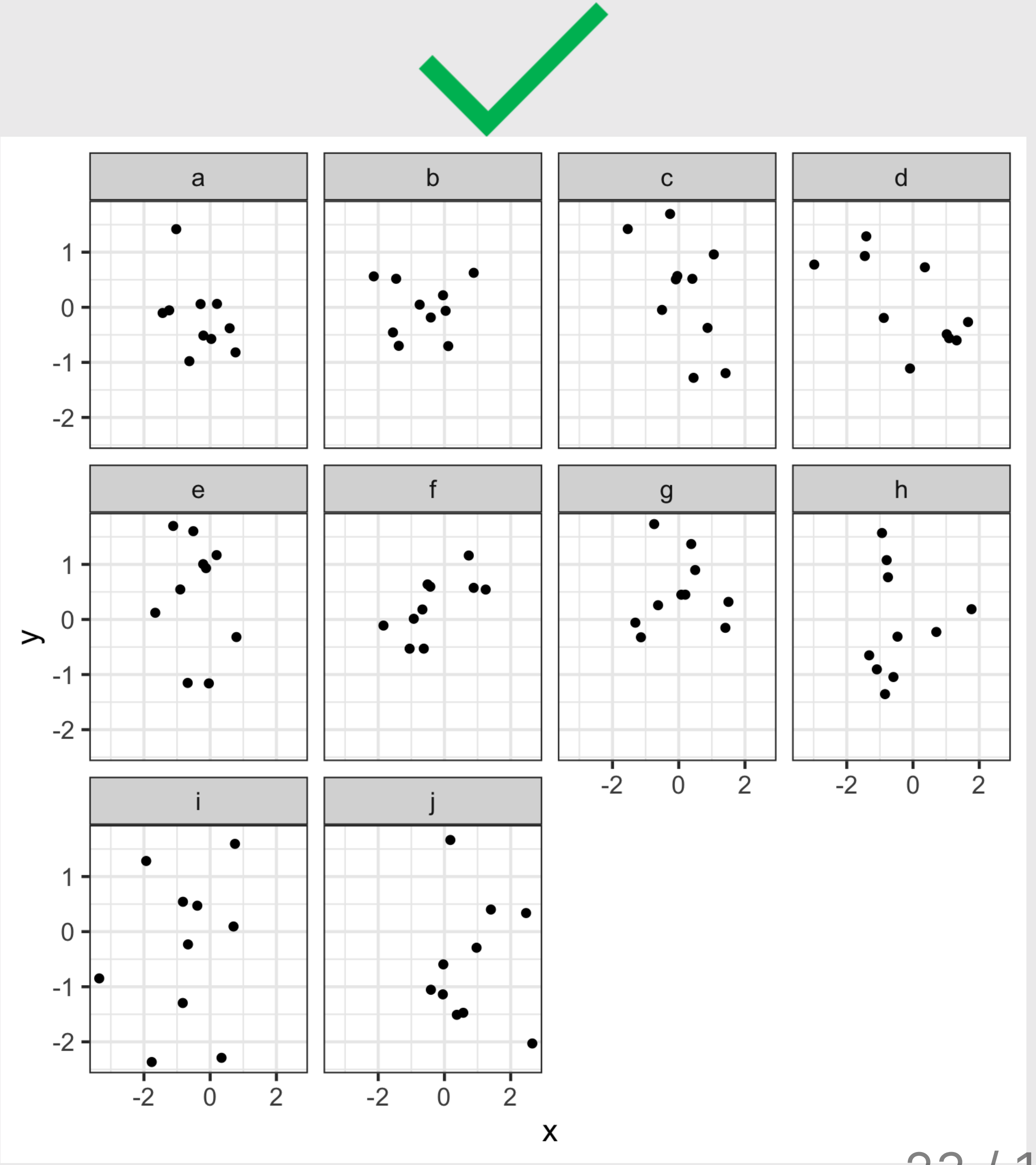
For categorical data:

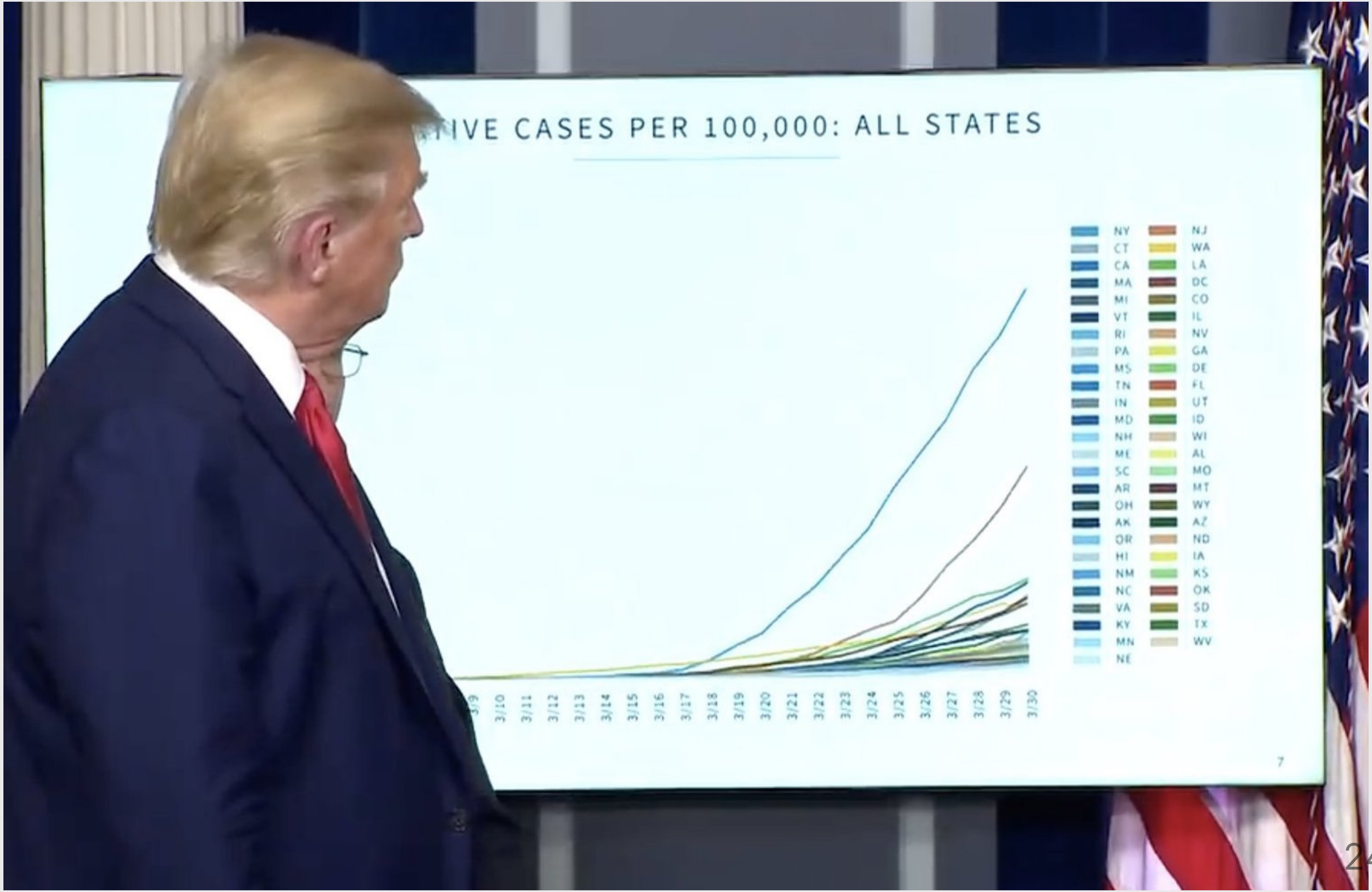
1. Hue (color) > shape

2. Less is more (stay in working memory!)




category
a
b
c
d
e
f
g
h
i
j
x
y


a
b
c
d
e
f
g
h
i
j
x
y


CASES PER 100,000: ALL STATES
NY CT CA MA MI VT RI PA MS TN IN MD NH ME SC AR OH AK OR HI NM NC VA KY MN NE
NJ WA LA DC CO IL NV GA DE FL UT ID WI AL MO MT WY AZ ND IA KS OK SD TX WV
3/10 3/11 3/12 3/13 3/14 3/15 3/16 3/17 3/18 3/19 3/20 3/21 3/22 3/23 3/24 3/25 3/26 3/27 3/28 3/29 3/30
7

Week 5: *Visualizing Information*

1. The Human Visual-Memory System
2. The Psychology of Data Viz

BREAK

3. 10 Data Viz Best Practices
4. Making a (good) ggplot

Much of the content in this section is from John Rauser's talk on YouTube

(Always cite your sources)

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. Science, New Series, 229(4716), 828-833.

Graphical Perception and Graphical Methods for Analyzing Scientific Data

William S. Cleveland and Robert McGill

Summary. Graphical perception is the visual decoding of the quantitative and qualitative information encoded on graphs. Recent investigations have uncovered basic principles of human graphical perception that have important implications for the display of data. The computer graphics revolution has stimulated the invention of many graphical methods for analyzing and presenting scientific data, such as box plots, two-tiered error bars, scatterplot smoothing, dot charts, and graphing on a log base 2 scale.

Graphs provide powerful tools both for analyzing scientific data and for communicating quantitative information. The computer graphics revolution, which began in the 1960's and has intensified during the past several years, stimulated the invention of graphical methods: types of graphs and types of quantitative information to be shown on graphs (*1–4*). One purpose of this article is to describe and illustrate several of these new methods.

What has been missing, until recently, in this period of rapid graphical invention and deployment is the study of graphs and the human visual system. When a graph is constructed, quantitative and categorical information is encoded, chiefly through position, shape, size, symbols, and color. When a person looks at a graph, the information is visually decoded by the person's visual sys-

mation from graphs; theory and experimental data are then used to order the tasks on the basis of accuracy. The ordering has an important application: data should be encoded so that the visual decoding involves tasks as high in the ordering as possible, that is, tasks performed with greater accuracy. This is illustrated by several examples in which some much-used graphical forms are presented, set aside, and replaced by new methods.

Elementary Tasks for the Graphical Perception of Quantitative Information

The first step is to identify elementary graphical-perception tasks that are used to visually extract quantitative information from a graph. (By "quantitative information" we mean numerical values

al field that comes without apparent mental effort. We also perform cognitive tasks such as reading scale information, but much of the power of graphs—and what distinguishes them from tables—comes from the ability of our preattentive visual system to detect geometric patterns and assess magnitudes. We have examined preattentive processes rather than cognition.

We have studied the elementary graphical-perception tasks theoretically, borrowing ideas from the more general field of visual perception (*7, 8*), and experimentally by having subjects judge graphical elements (*1, 5*). The next two sections illustrate the methodology with a few examples.

Study of Graphical Perception: Theory

Figure 2 provides an illustration of theoretical reasoning that borrows some ideas from the field of computational vision (*8*). Suppose that the goal is to judge the ratio, r, of the slope of line segment BC to the slope of line segment AB in each of the three panels. Our visual system tells us that r is greater than 1 in each panel, which is correct. Our visual system also tells us that r is closer to 1 in the two rectangular panels than in the square panel; that is, the slope of BC appears closer to the slope of AB in the two rectangular panels than in the square panel. This, however, is incorrect; r is the same in all three panels.

The reason for the distortion in judging Fig. 2 is that our visual system is geared to judging angle rather than slope. In their work on computational theories of vision in artificial intelligence, Marr (*8*) and Stevens (*9*) have investigated how people judge the slant and tilt (*10*) of the surfaces of three-dimensional objects. They argue that we judge slant and tilt as

Cleveland's operations of pattern perception:

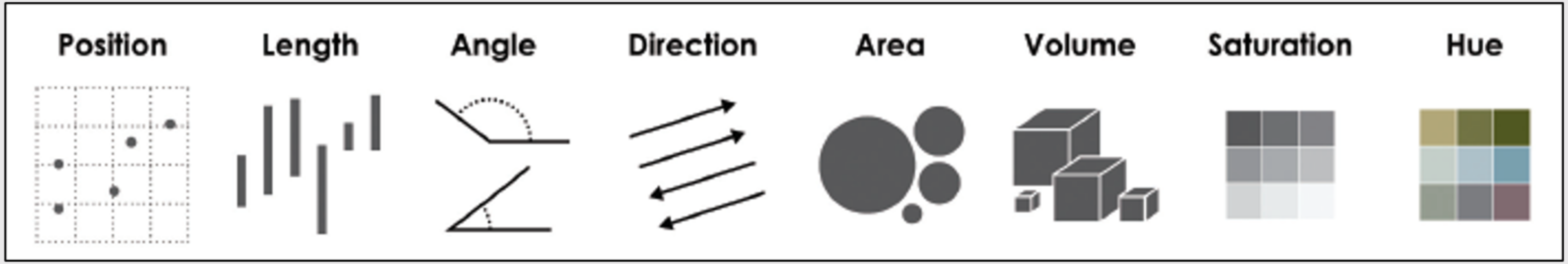
1. Estimation
2. Assembly
3. Detection

Cleveland's operations of pattern perception:

1. Estimation --------->
2. Assembly
3. Detection

- **Discrimination** (X equal to Y?)
- **Ranking** (X greater than Y?)
- **Ratioing** (X double Y?)

Estimation: Hierarchy for *numerical* data



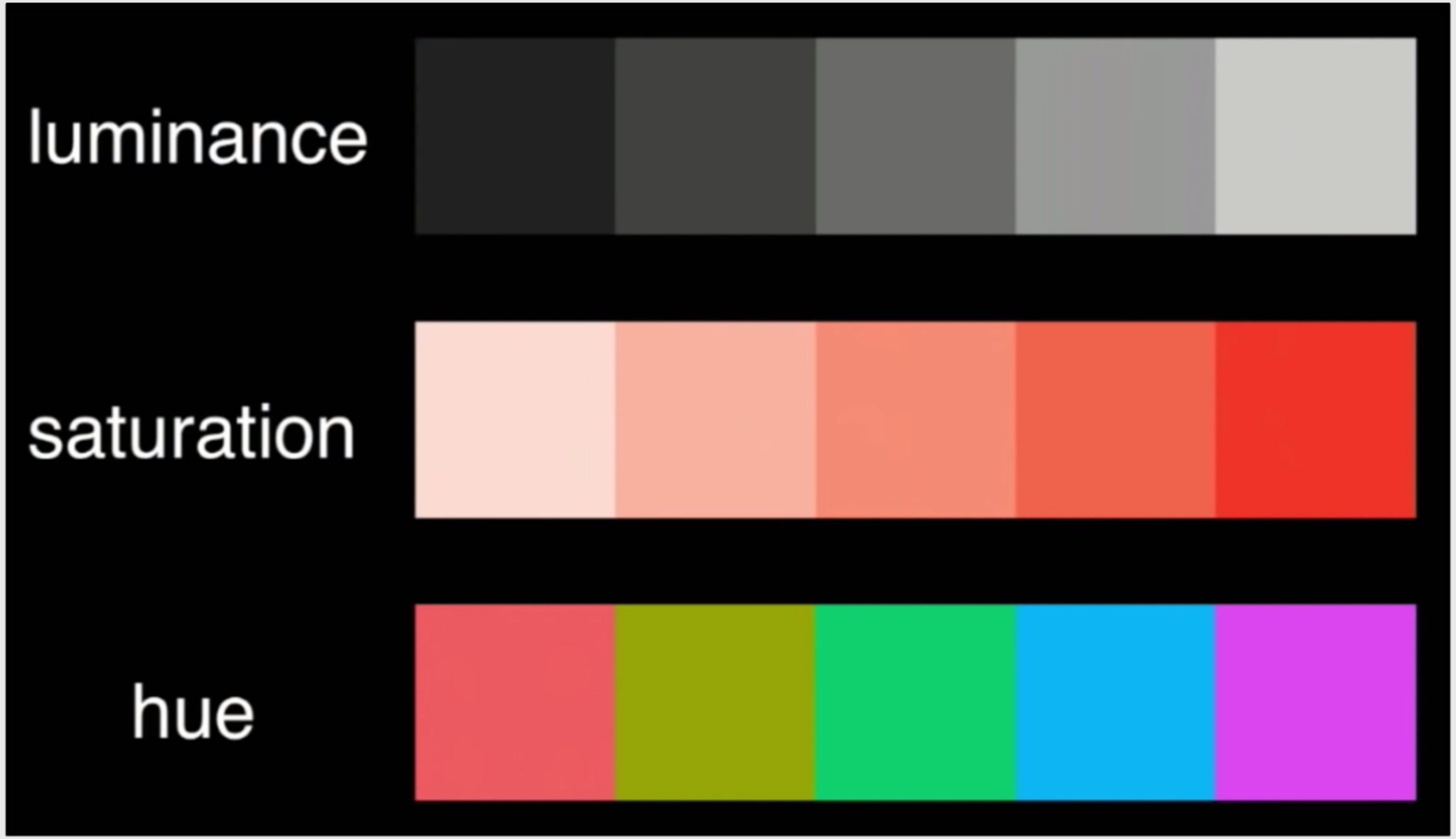
More Accurate

Less Accurate

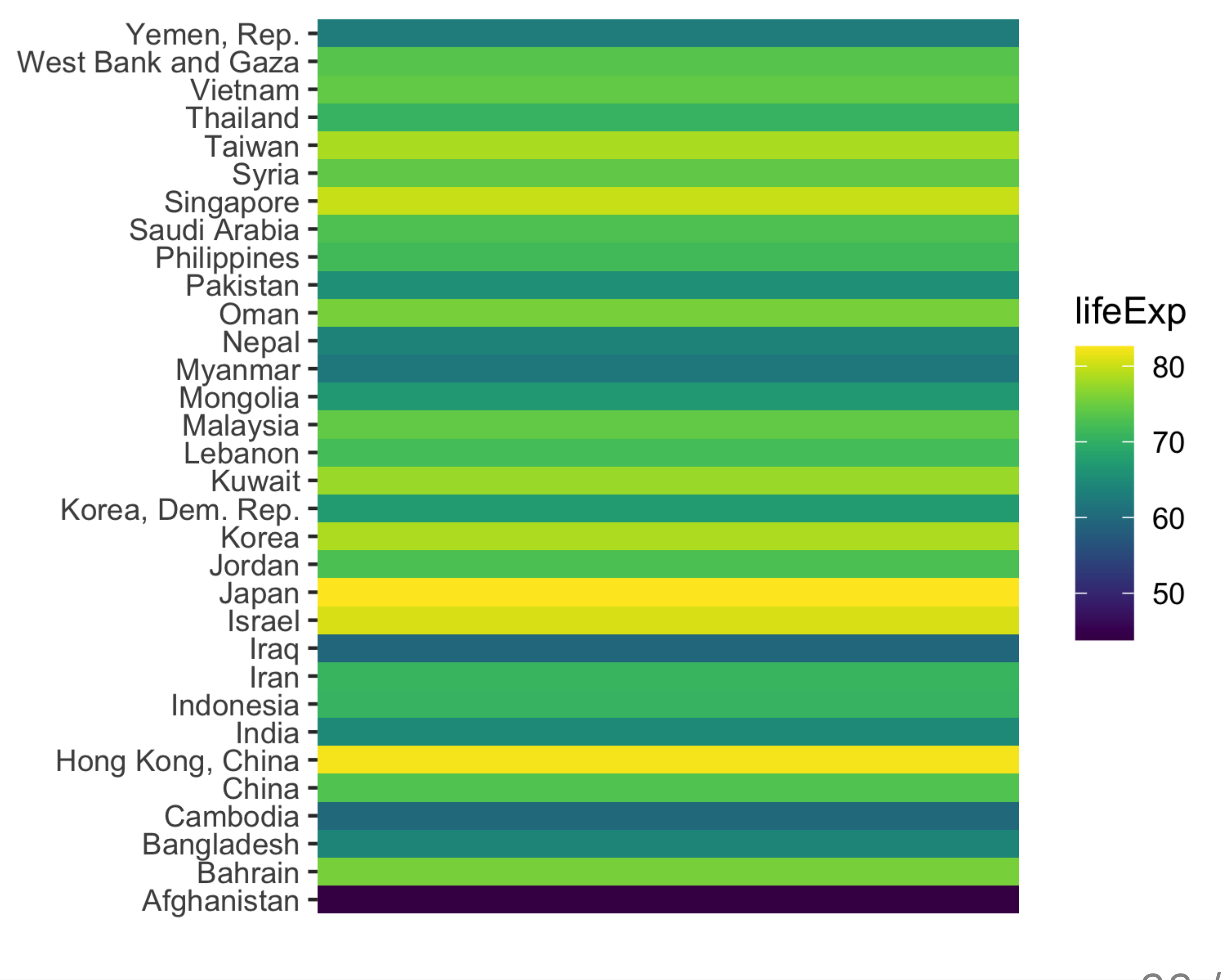
Example: Life expectancy in countries in Asia

```
#>                 country lifeExp
#> 1           Afghanistan  43.828
#> 2                  Iraq  59.545
#> 3              Cambodia  59.723
#> 4               Myanmar  62.069
#> 5           Yemen, Rep.  62.698
#> 6                 Nepal  63.785
#> 7            Bangladesh  64.062
#> 8                 India  64.698
#> 9              Pakistan  65.483
#> 10             Mongolia  66.803
#> 11   Korea, Dem. Rep.  67.297
#> 12             Thailand  70.616
#> 13            Indonesia  70.650
#> 14                 Iran  70.964
#> 15          Philippines  71.688
#> 16              Lebanon  71.993
#> 17               Jordan  72.535
#> 18         Saudi Arabia  72.777
#> 19                China  72.961
```

1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. Color saturation
7. **Color hue**



1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. Color saturation
7. **Color hue**

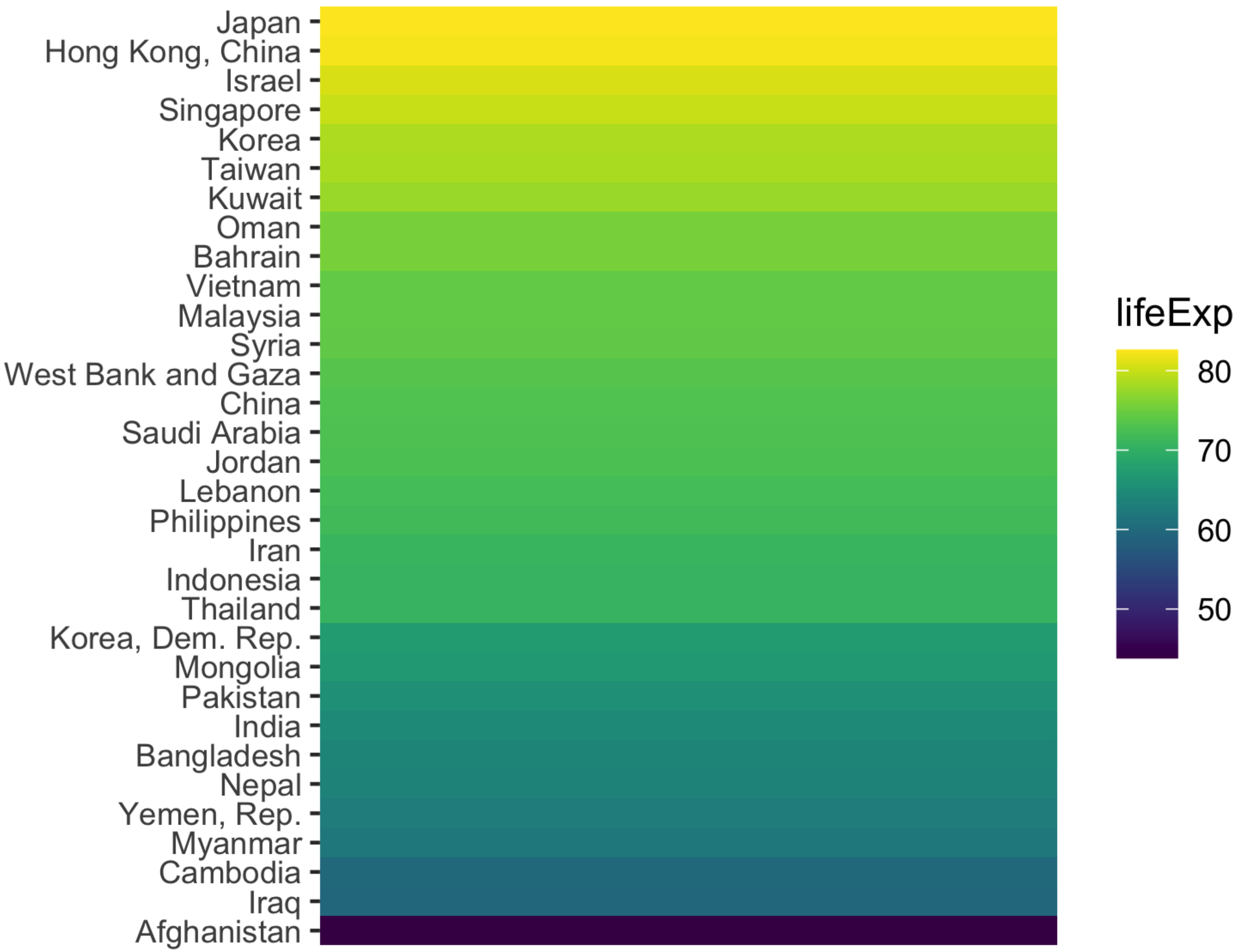


- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ❌ **Ratio**

1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. Color saturation
7. **Color hue**

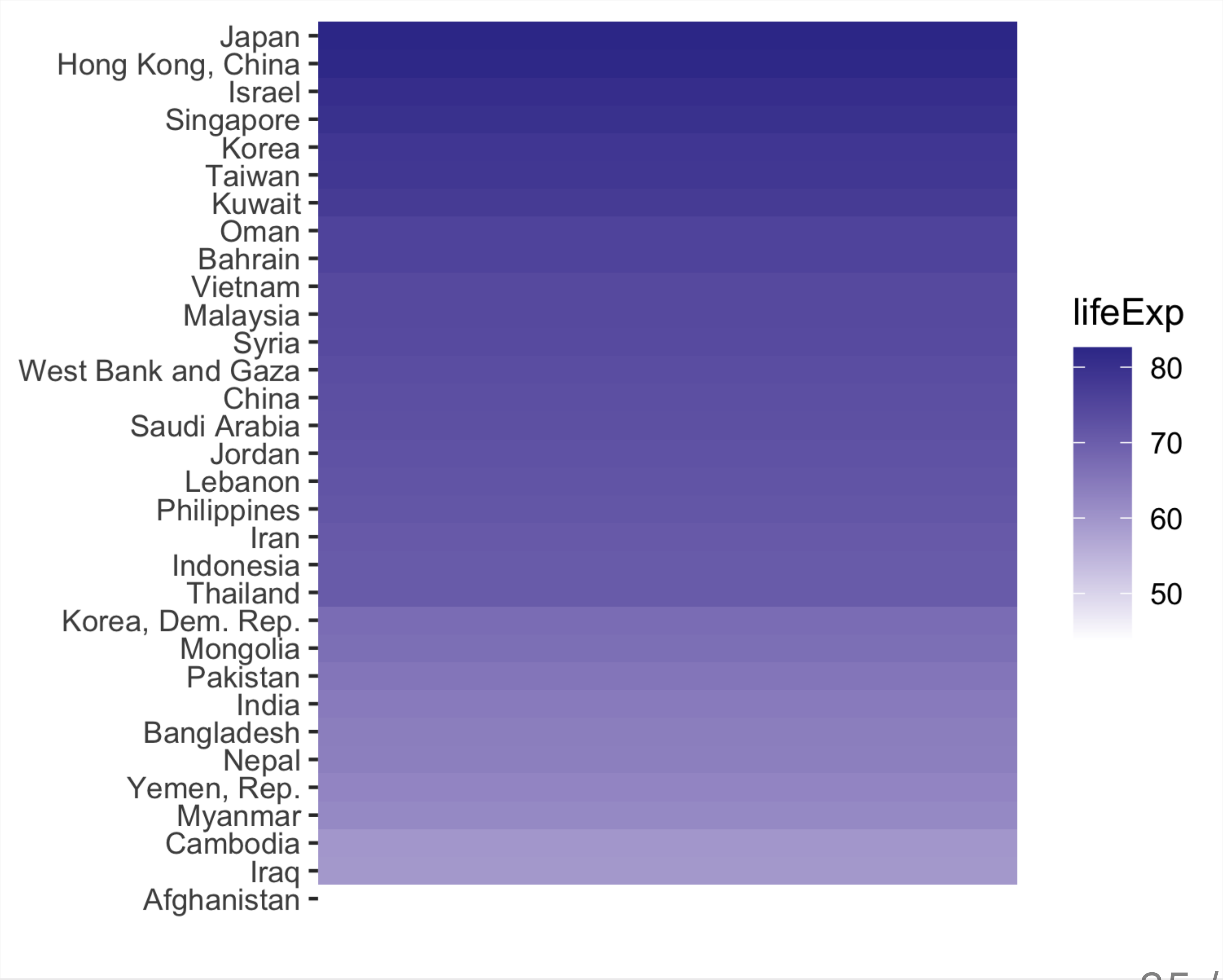
Sorting helps a bit...

- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ❌ **Ratio**



1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. **Color saturation**
7. Color hue

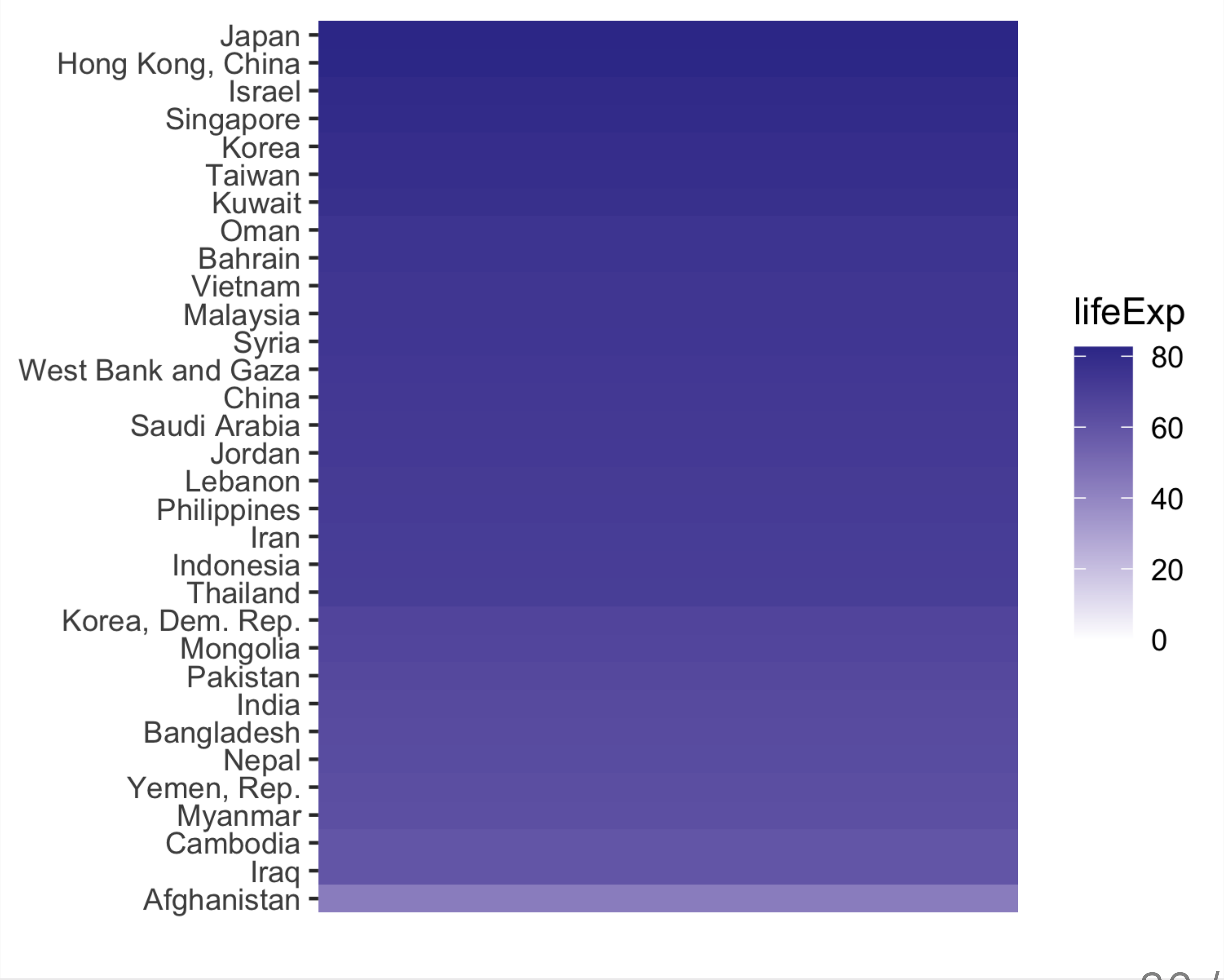
- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ❌ **Ratio**



1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. **Color saturation**
7. Color hue

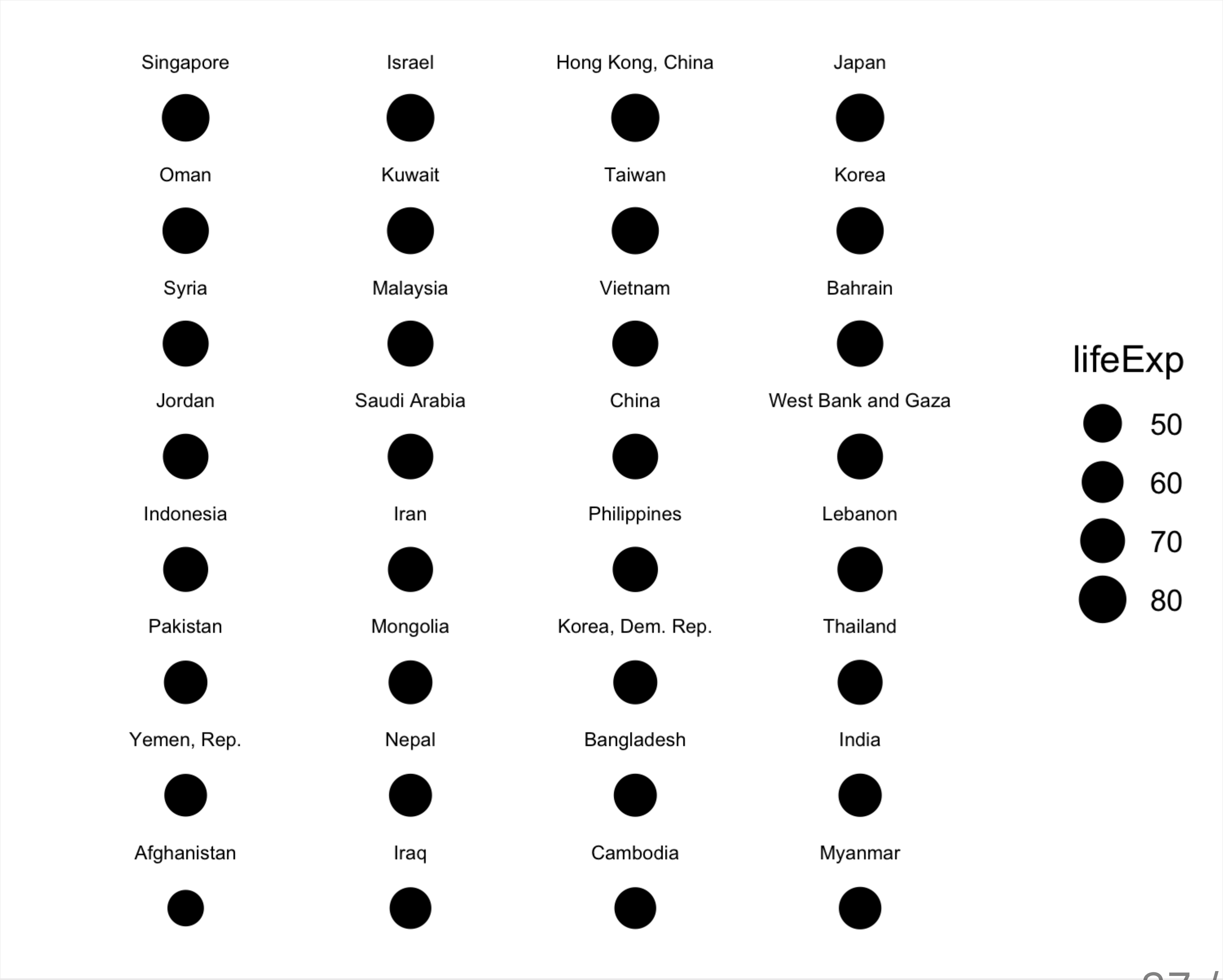
Align to 0 scale:

- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ✅ / ❌ **Ratio**



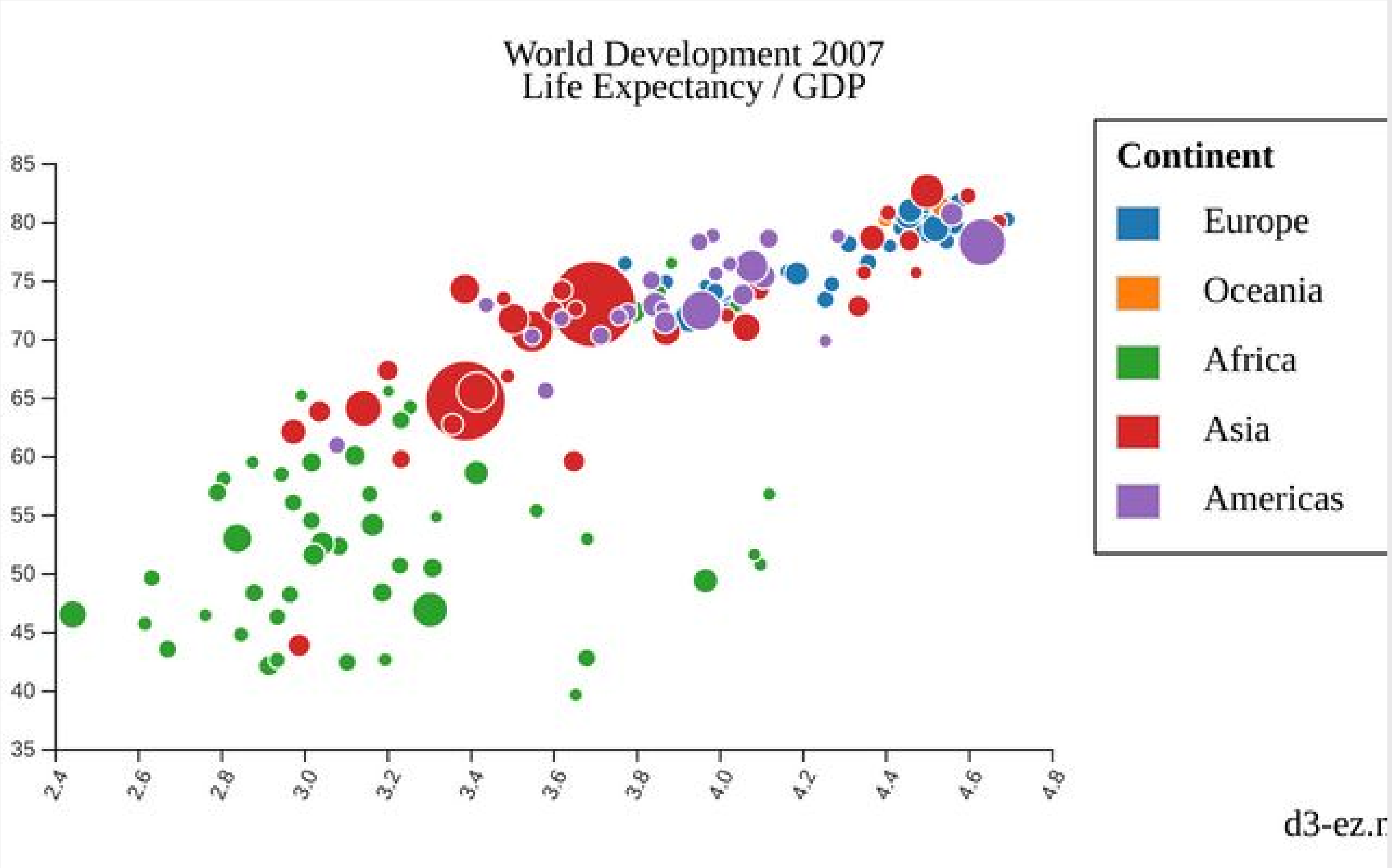
1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. **Area**
6. Color saturation
7. Color hue

- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ✅ / ❌ **Ratio**



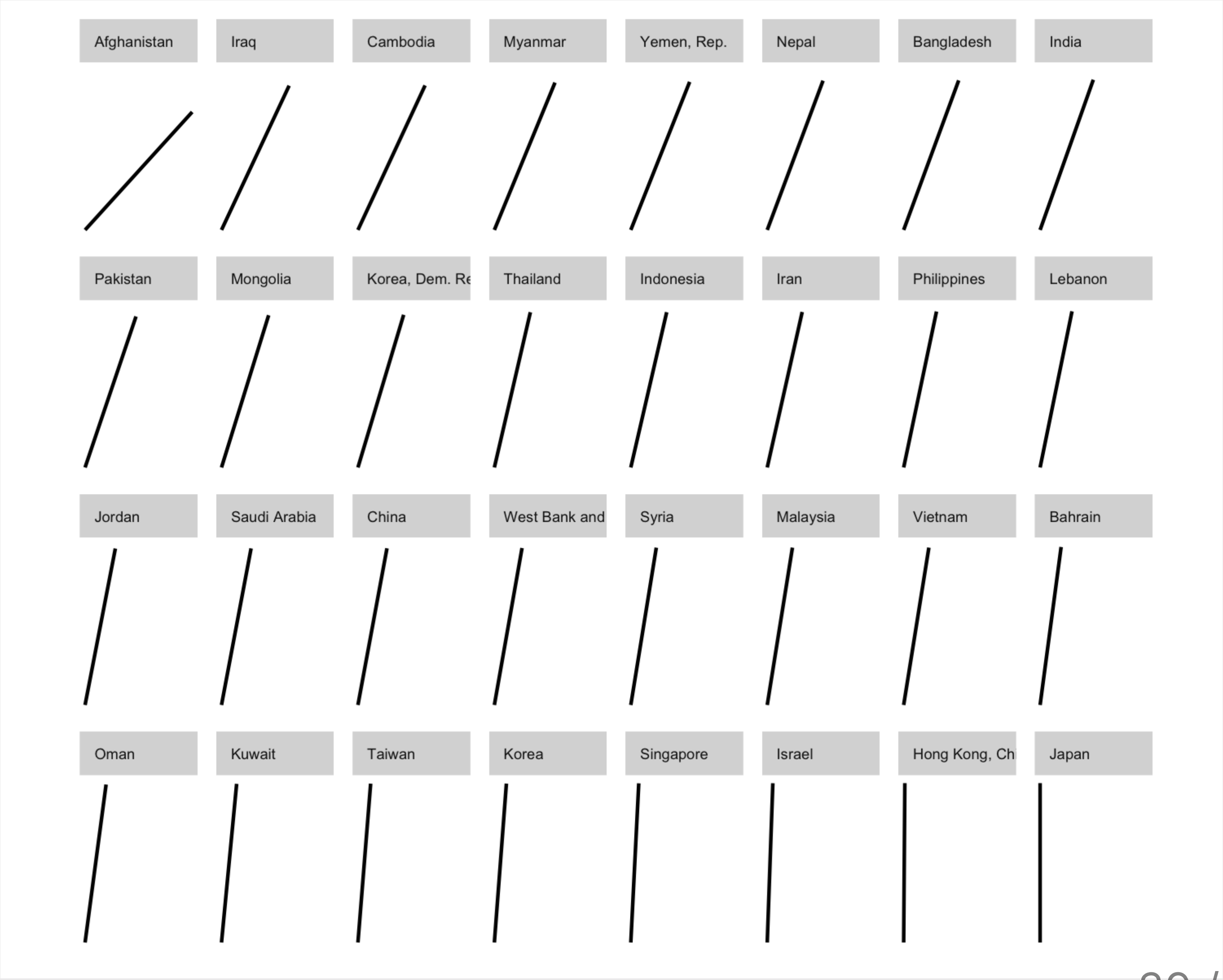
1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. Angle
5. **Area**
6. Color saturation
7. Color hue

Area works okay for "bubble" charts

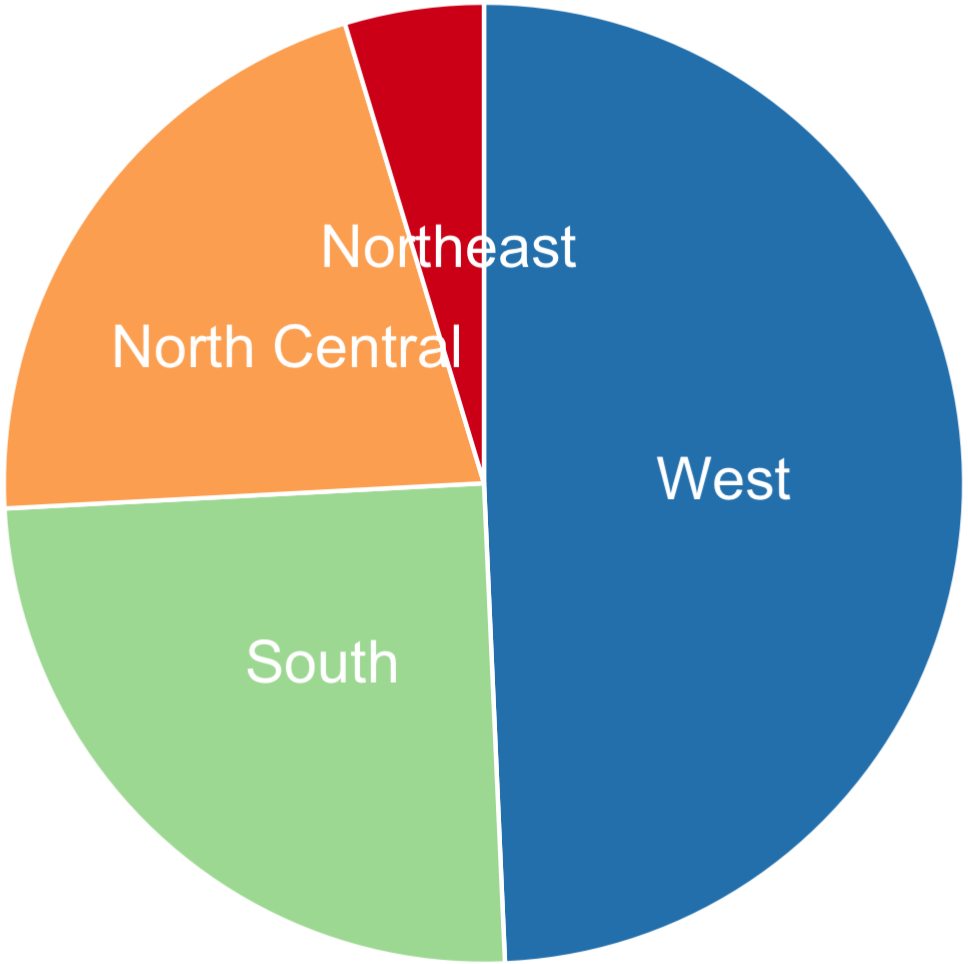


1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. **Angle**
5. Area
6. Color saturation
7. Color hue

- ✅ / ❌ **Discriminate**
- ✅ **Rank**
- ✅ / ❌ **Ratio**



1. Position on a common scale
2. Position on non-aligned scales
3. Length
4. **Angle**
5. Area
6. Color saturation
7. Color hue



1. Position on a common scale
2. Position on non-aligned scales
3. **Length**
4. Angle
5. Area
6. Color saturation
7. Color hue

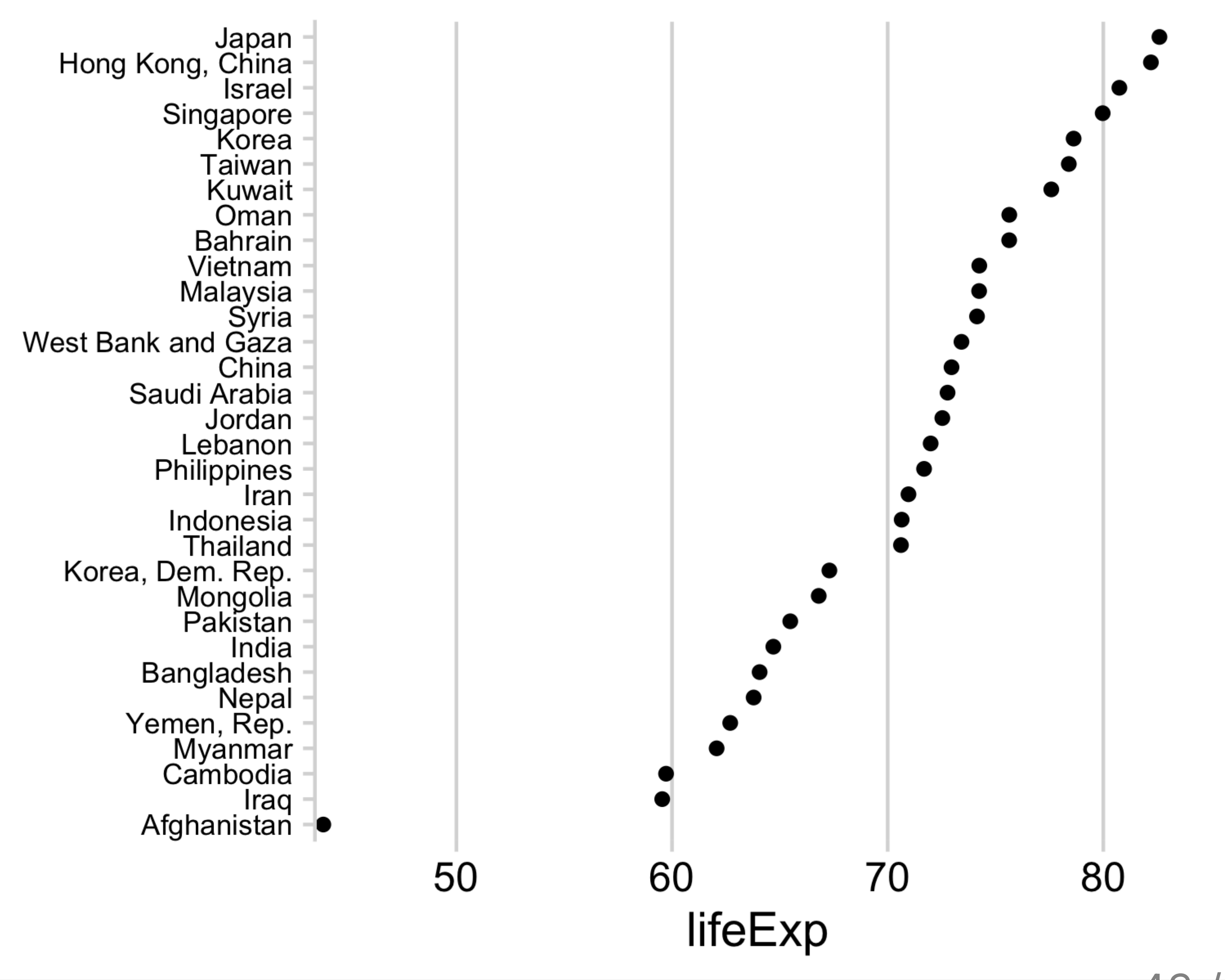
- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ✅ **Ratio**

1. Position on a common scale
2. **Position on non-aligned scales**
3. Length
4. Angle
5. Area
6. Color saturation
7. Color hue

- ✅ / ❌ **Discriminate**
- ✅ / ❌ **Rank**
- ✅ **Ratio**

1. **Position on a common scale**
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. Color saturation
7. Color hue

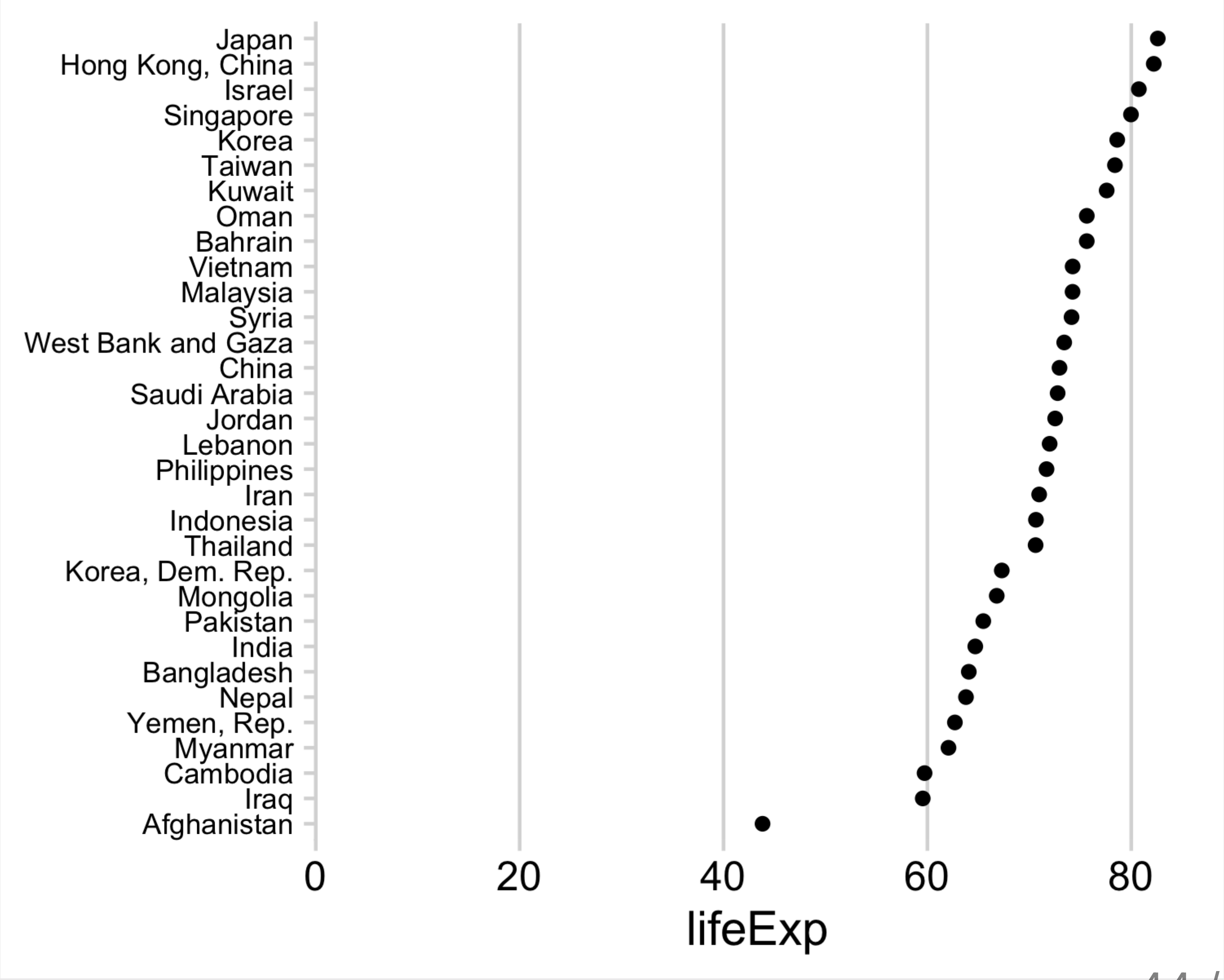
- ✅ **Discriminate**
- ✅ **Rank**
- ✅ **Ratio**



1. **Position on a common scale**
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. Color saturation
7. Color hue

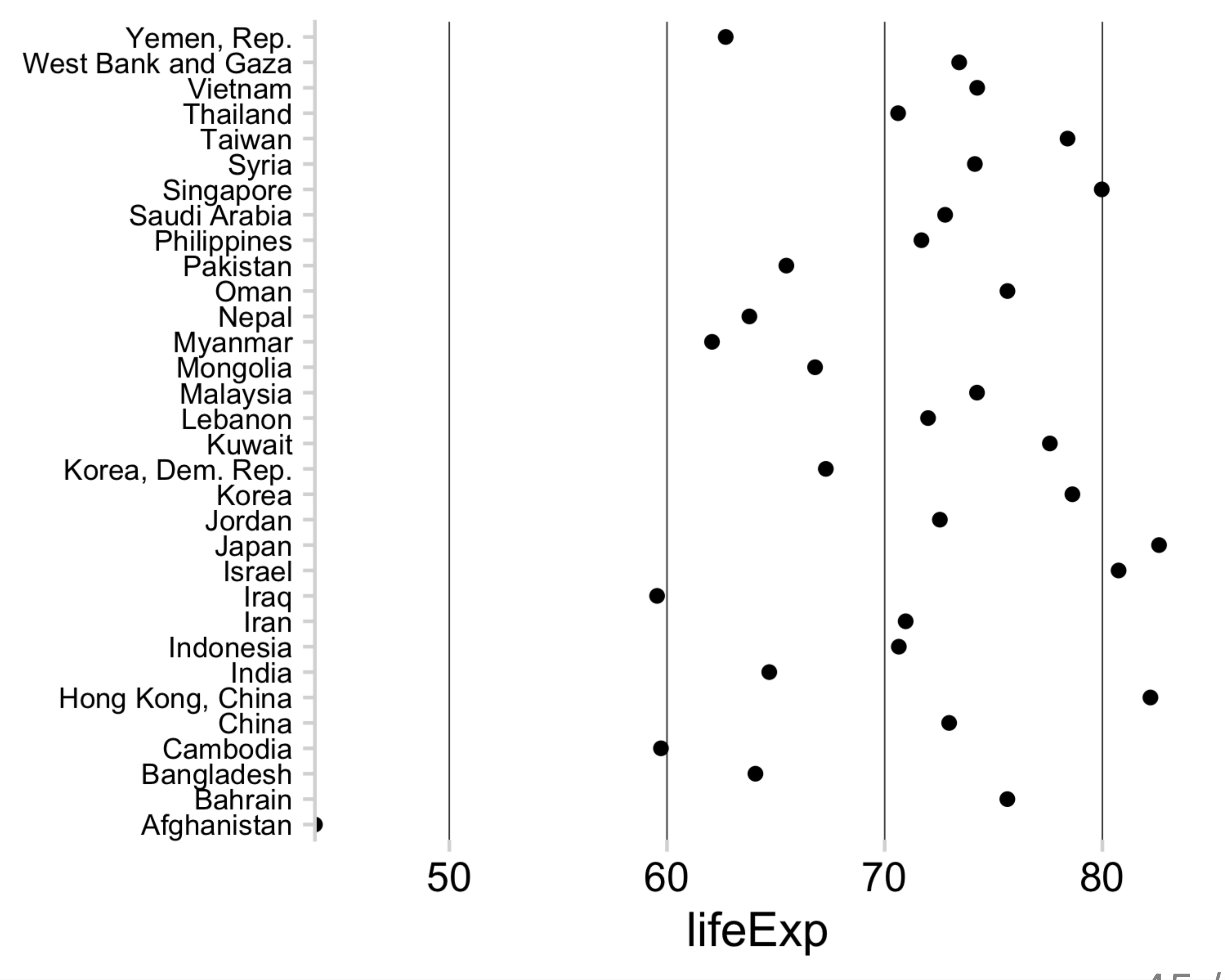
No need to scale to 0:

- Lowers resolution
- Isn't needed for accurate ratioing



1. **Position on a common scale**
2. Position on non-aligned scales
3. Length
4. Angle
5. Area
6. Color saturation
7. Color hue

Sorting still matters!



Cleveland's operations of pattern perception:

1. Estimation
2. Assembly
3. Detection

Cleveland's operations of pattern perception:

1. Estimation
2. Assembly ---------> **The grouping of graphical elements**
3. Detection

Assembly: Gestalt Psychology

The whole has a reality that is entirely separate from the parts

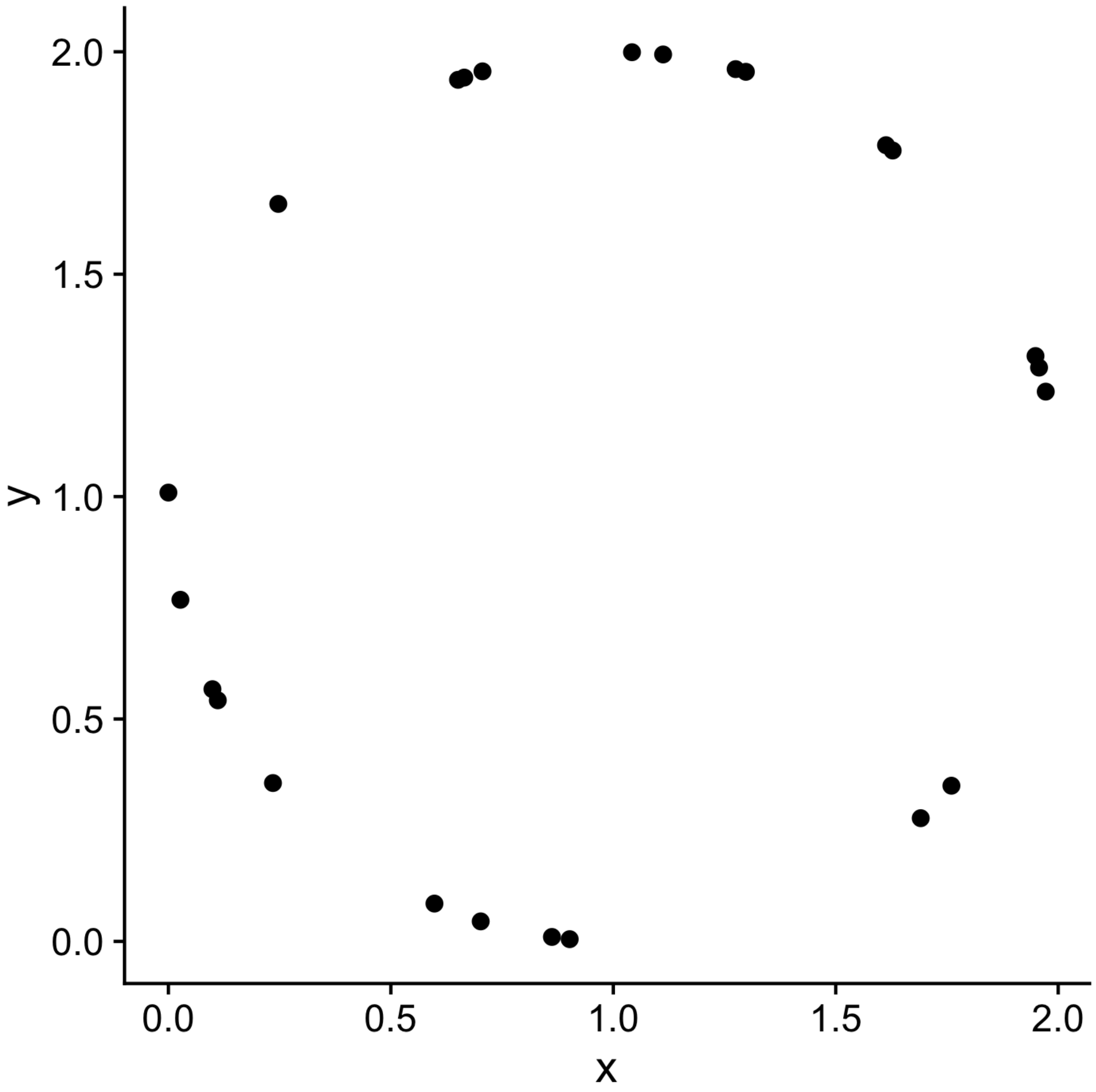
WWF

Reification

Emergence

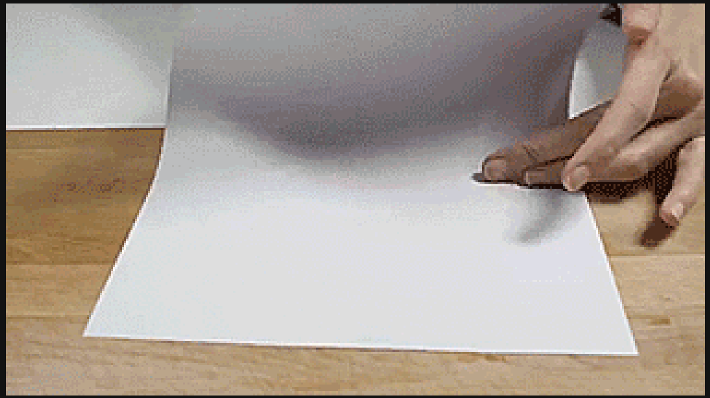
Law of Closure

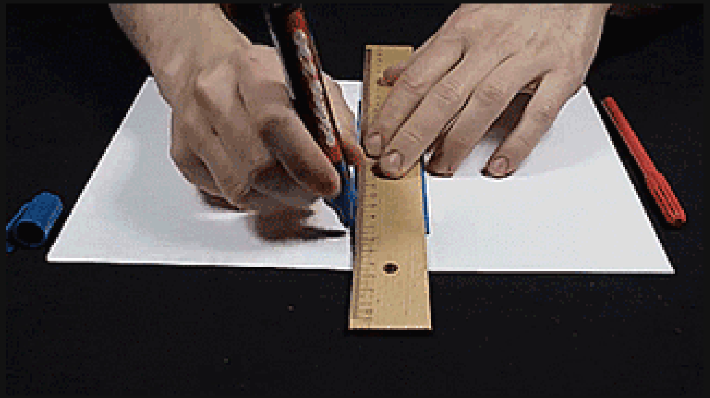
Our minds fill in the missing information



Prägnanz

We strongly prefer to interpret stimuli as regular, simple, and orderly

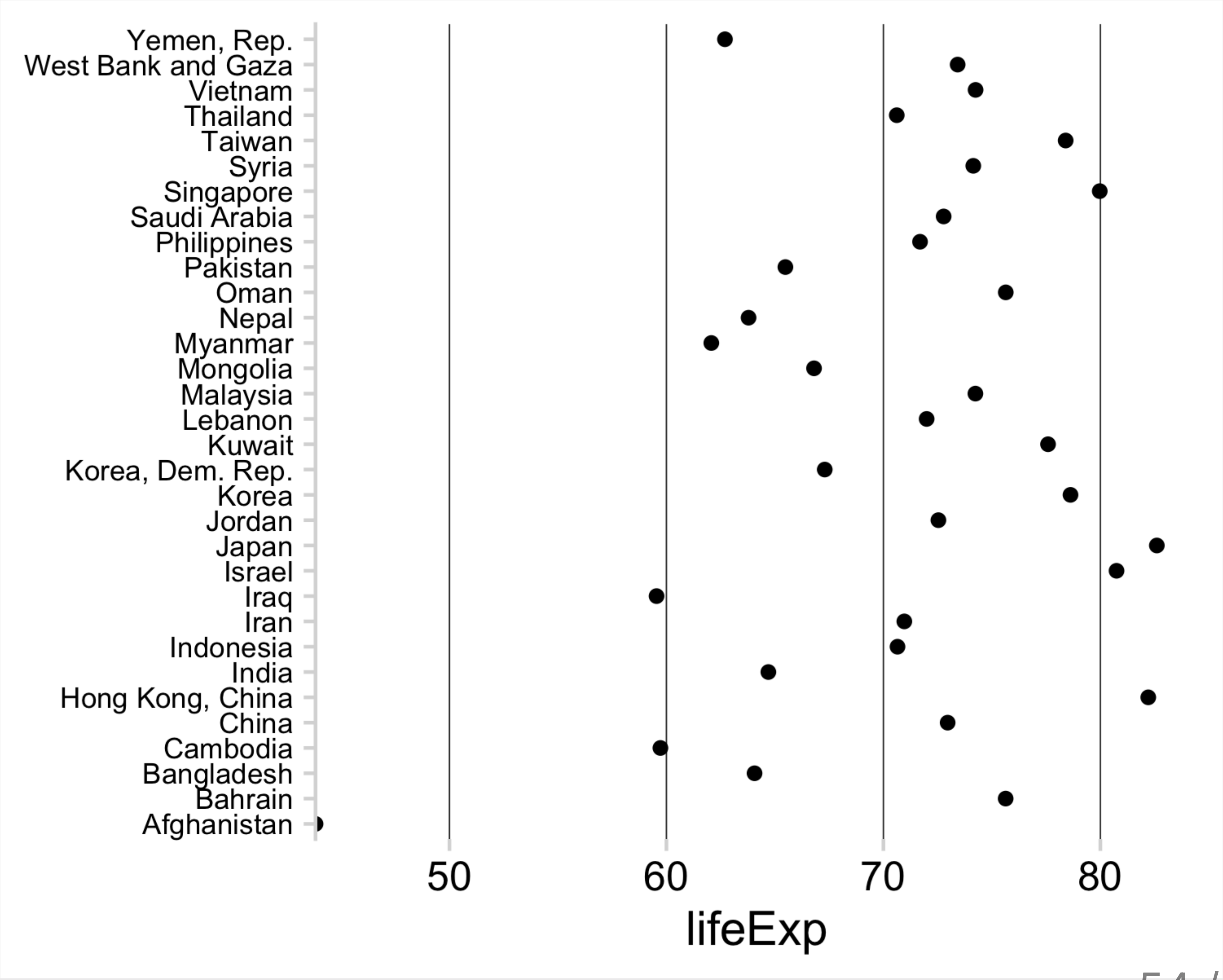




Prägnanz

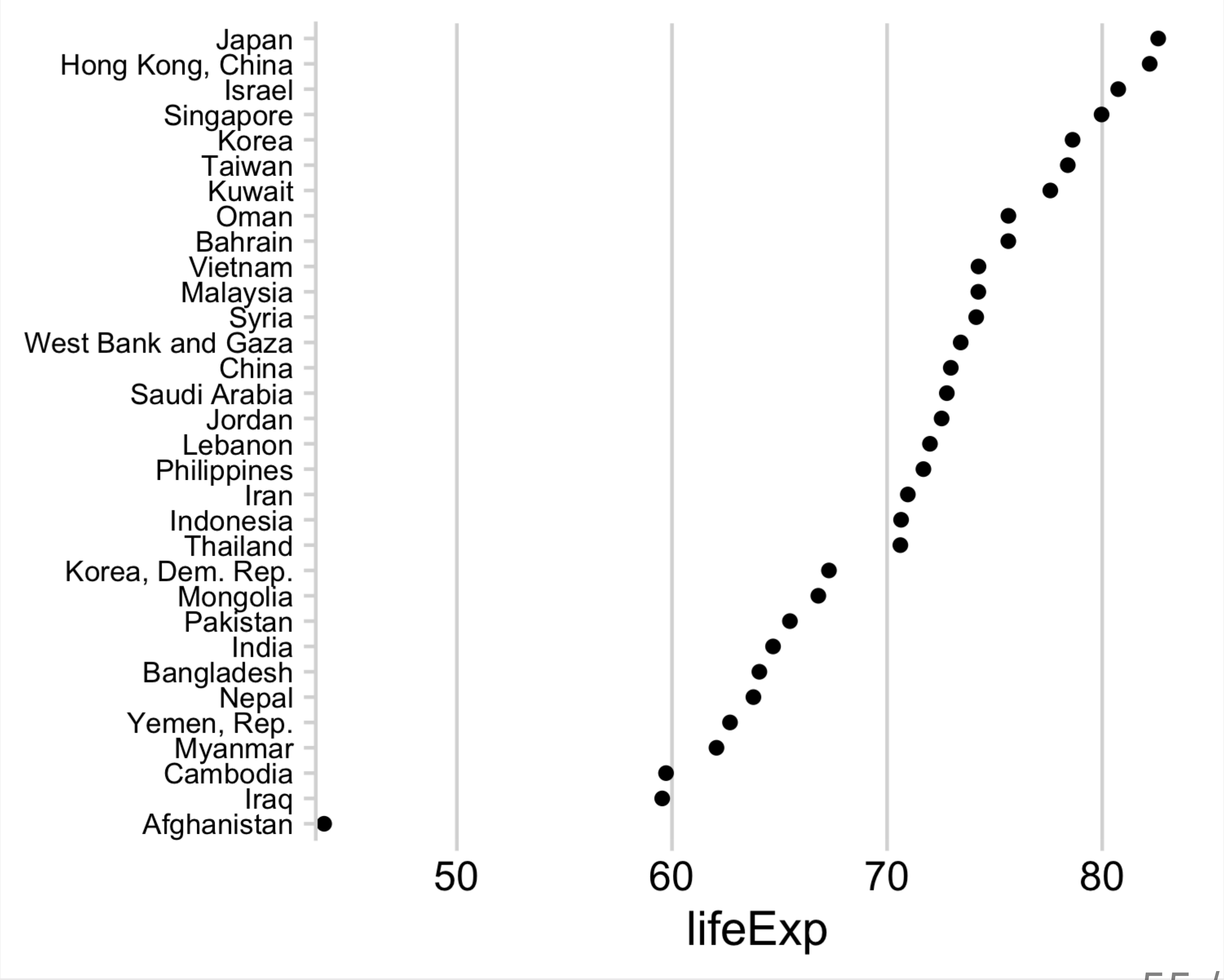
This should cause you cognitive pain

It's the graphical equivalent of this:



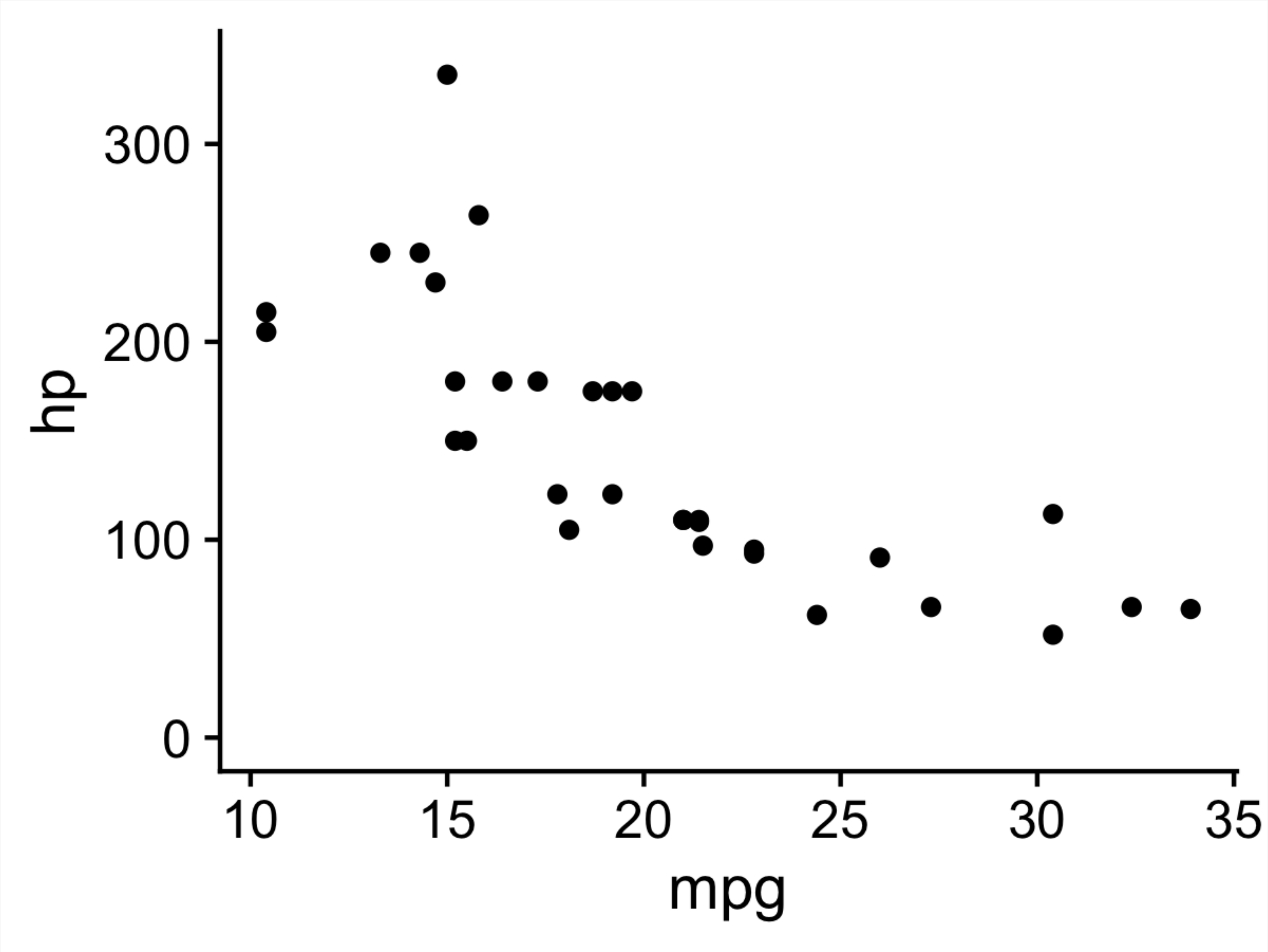
Prägnanz

This makes our brains happy



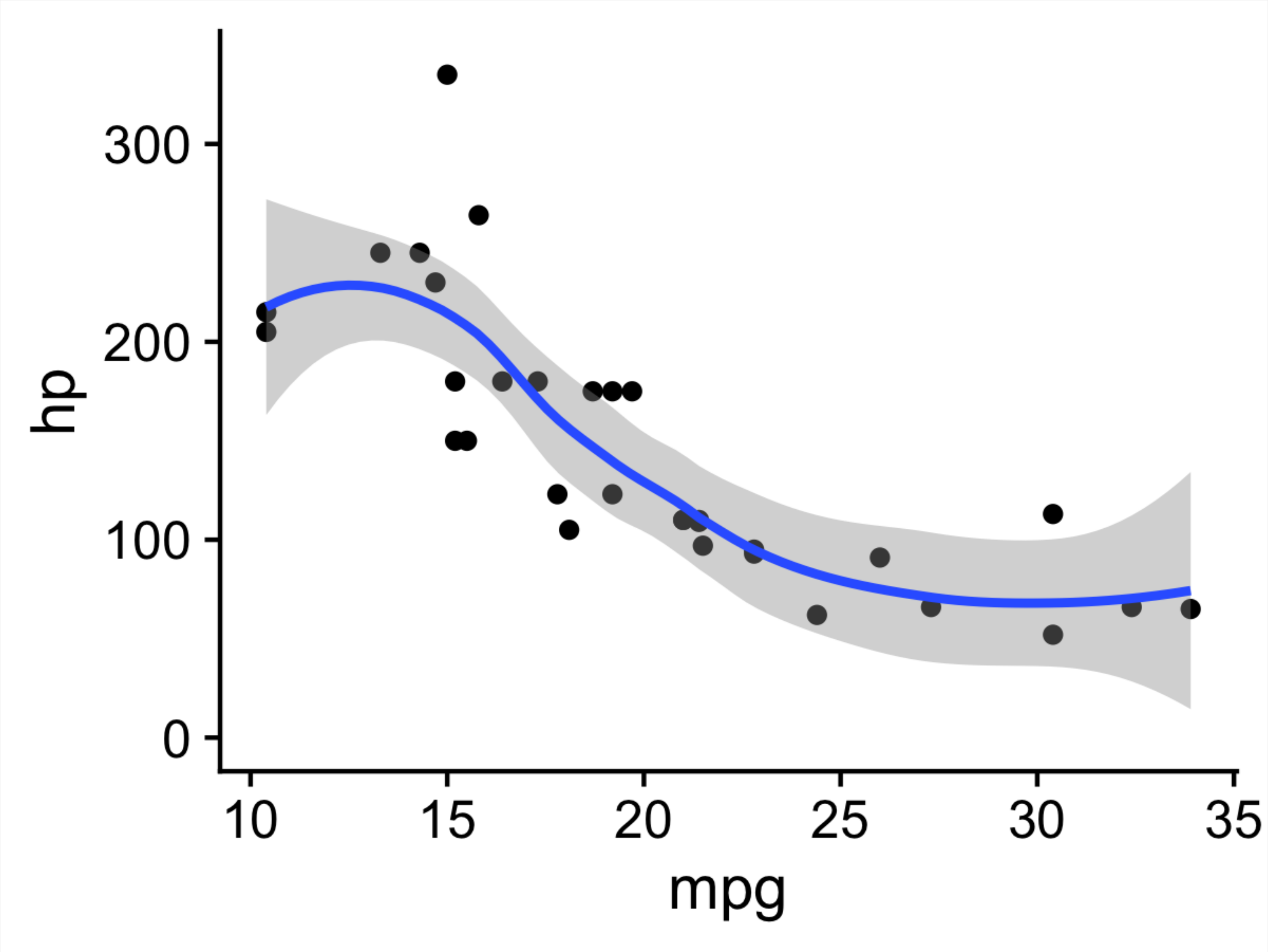
Law of Continuity

We will group together objects that follow an established direction



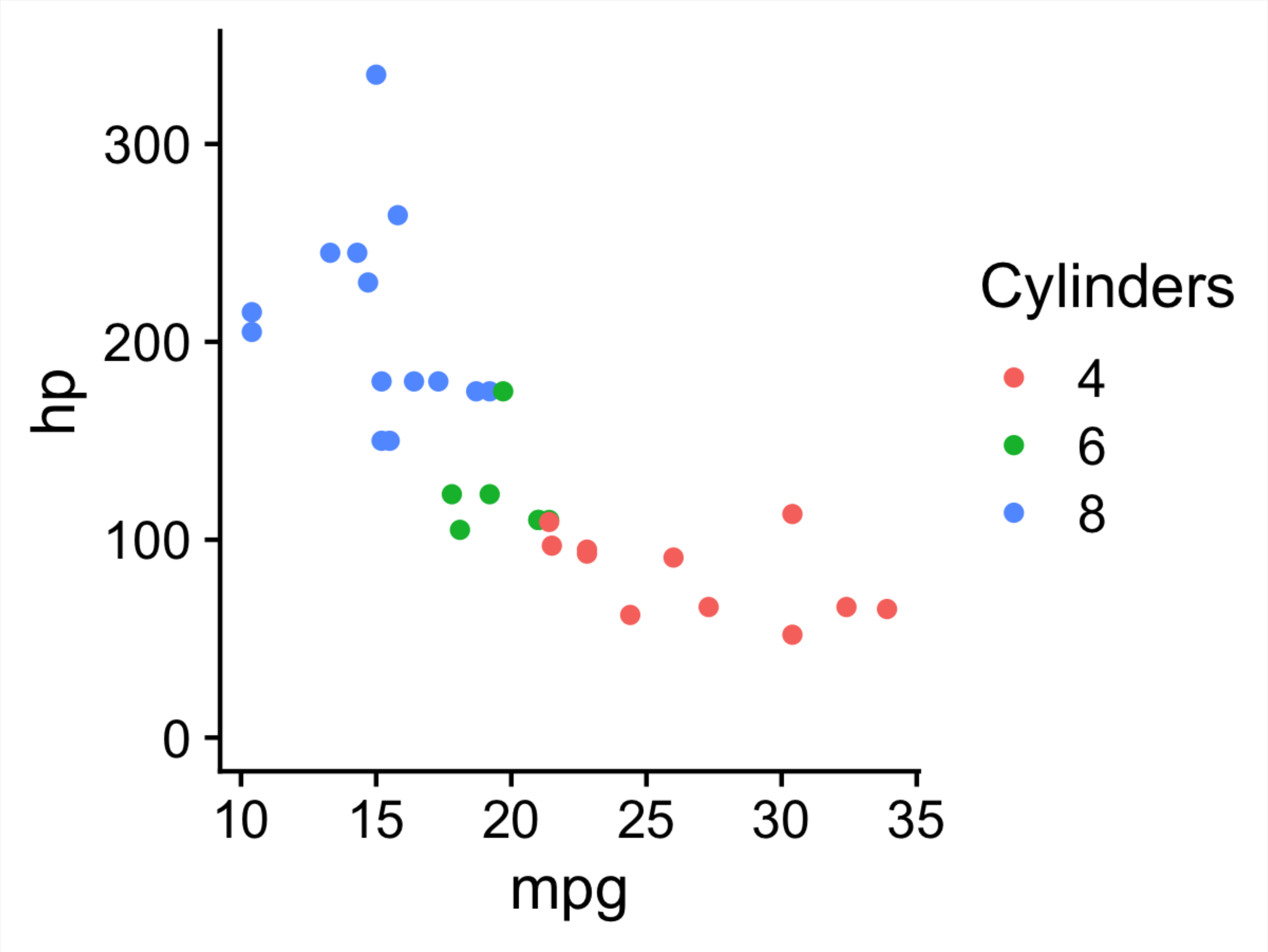
Law of Continuity

We will group together objects that follow an established direction



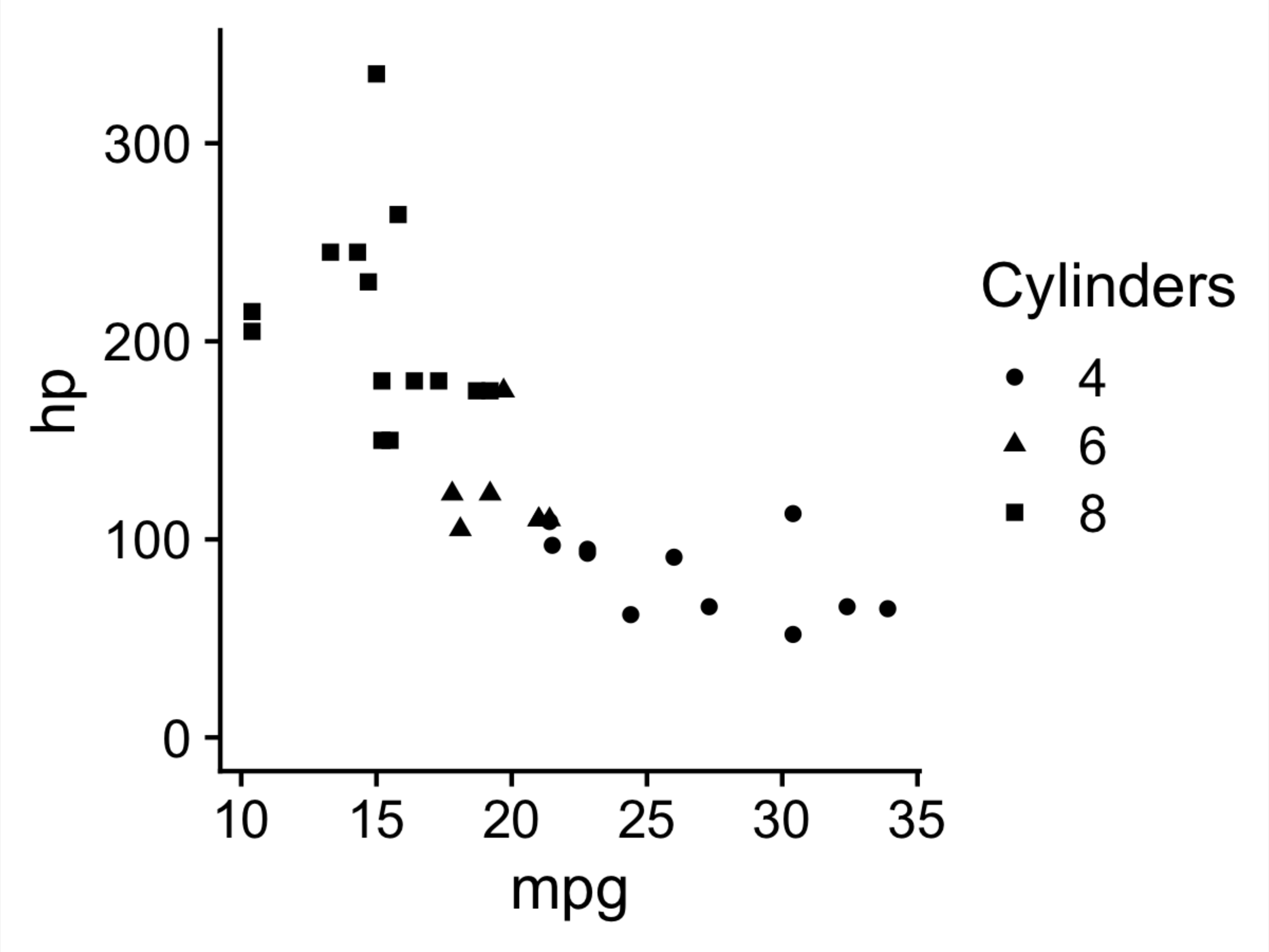
Law of Similarity

We see elements that are *physically similar* as part of the same object



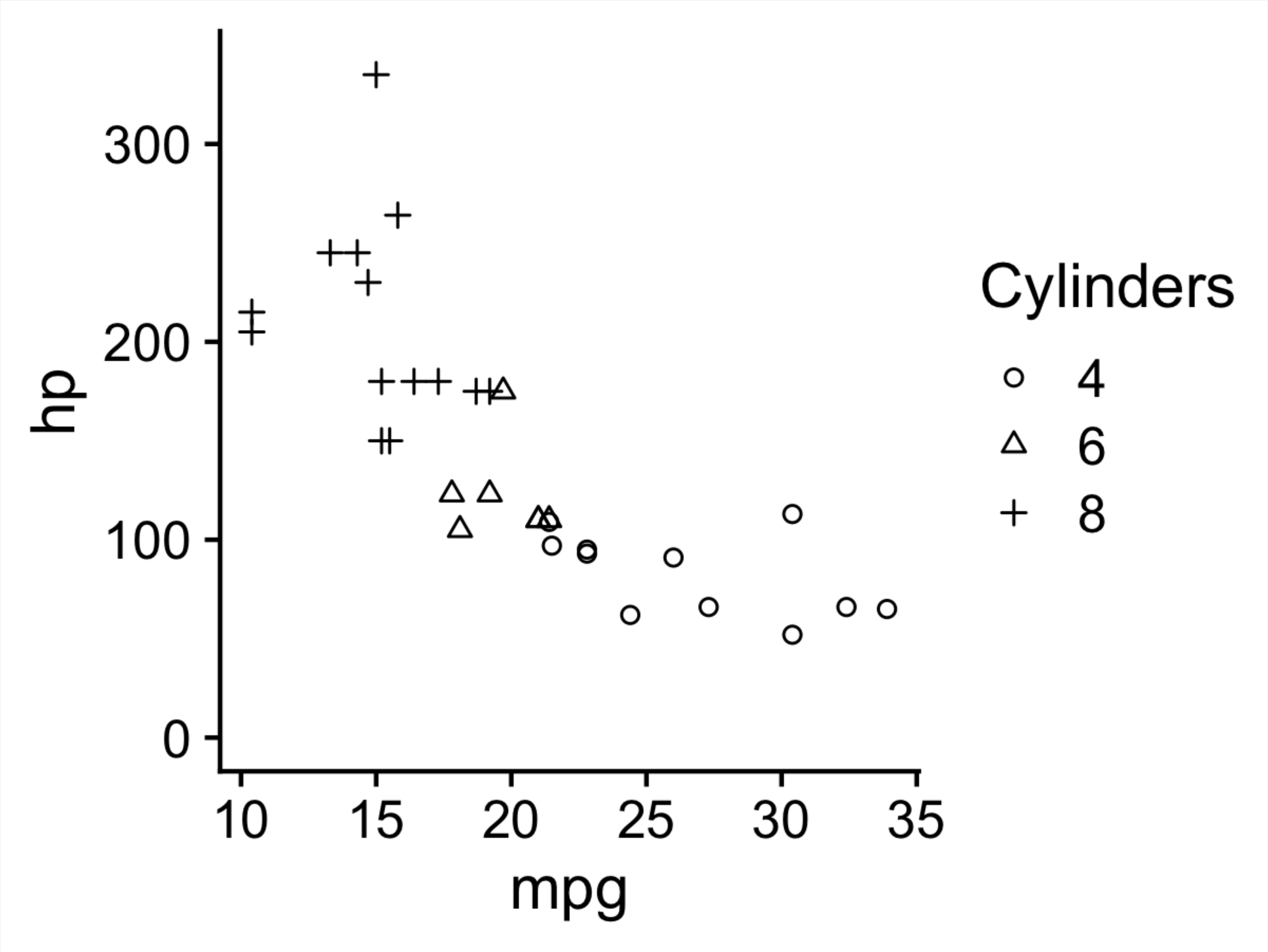
Law of Similarity

We see elements that are *physically similar* as part of the same object



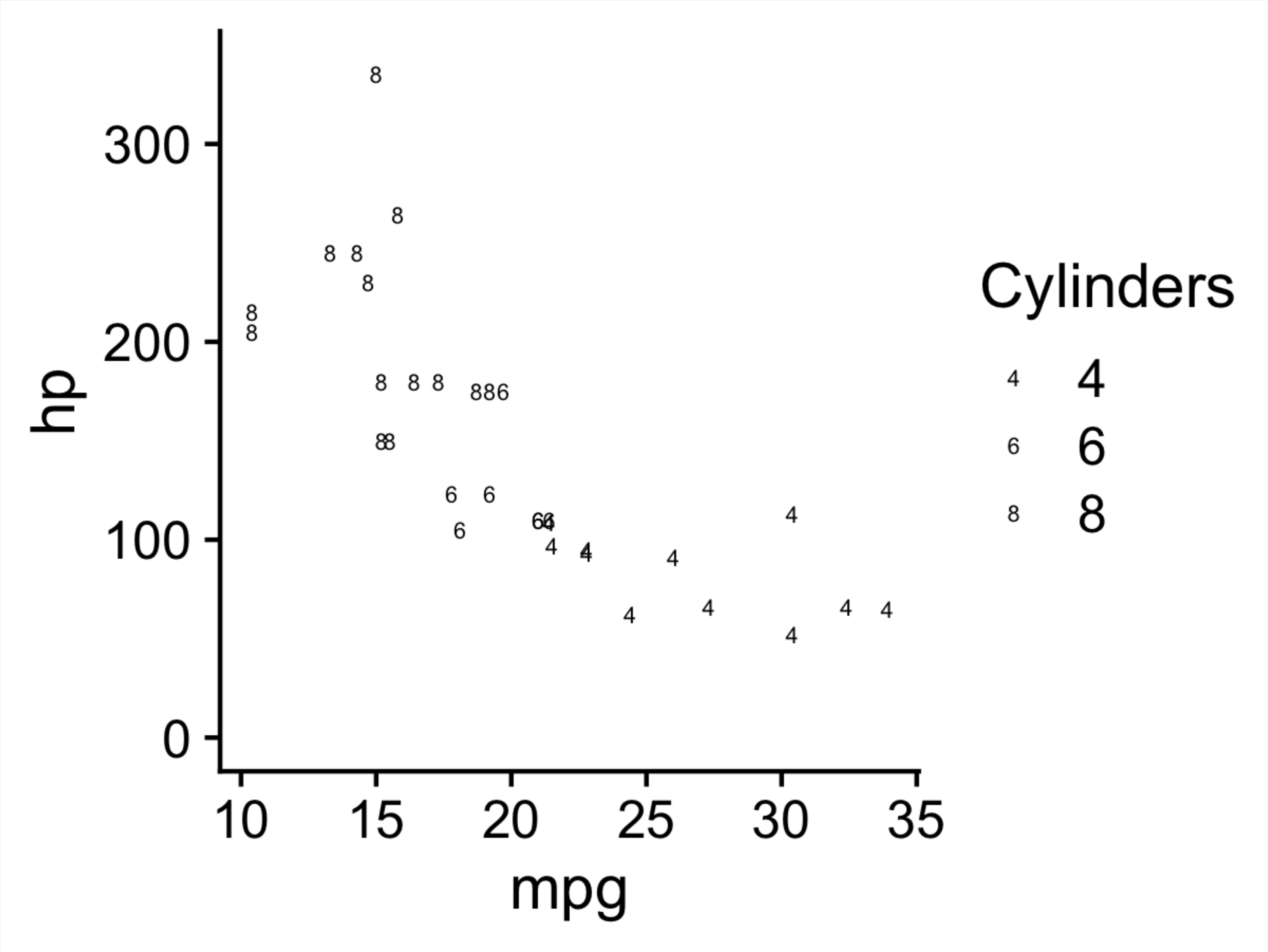
Law of Similarity

We see elements that are *physically similar* as part of the same object



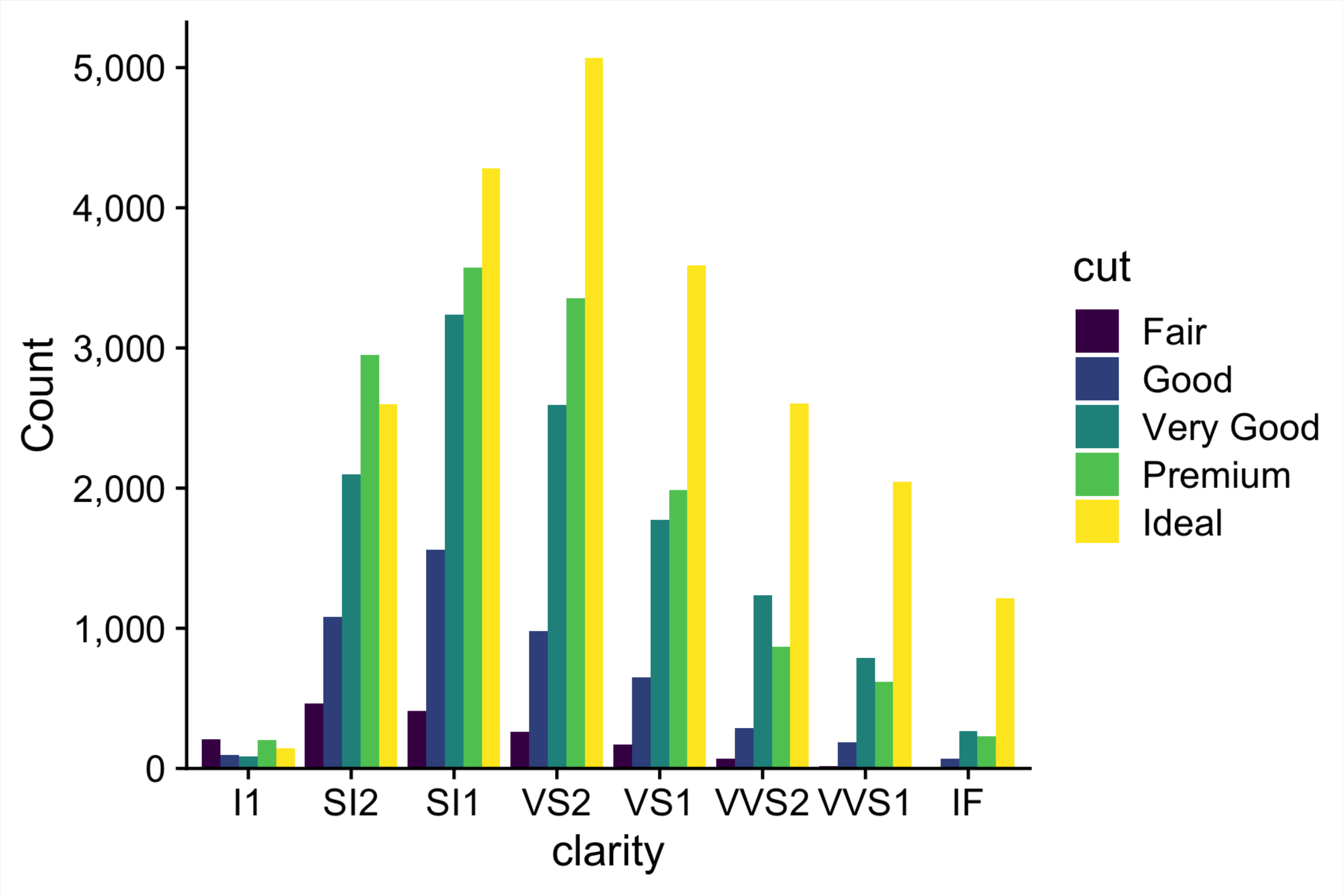
Law of Similarity

We see elements that are *physically similar* as part of the same object



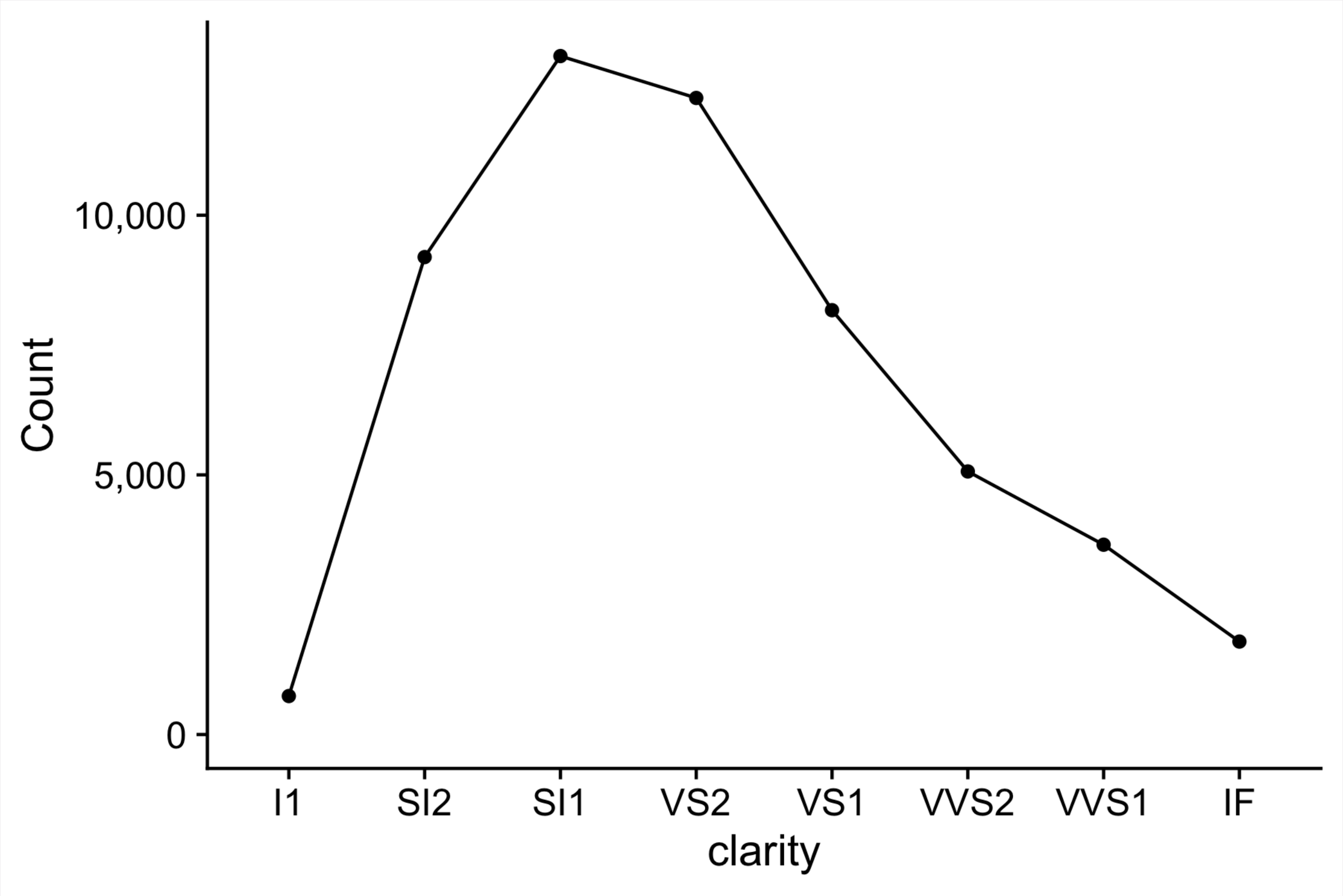
Law of Proximity

We tend to see elements that are *physically near* each other as part of the same object



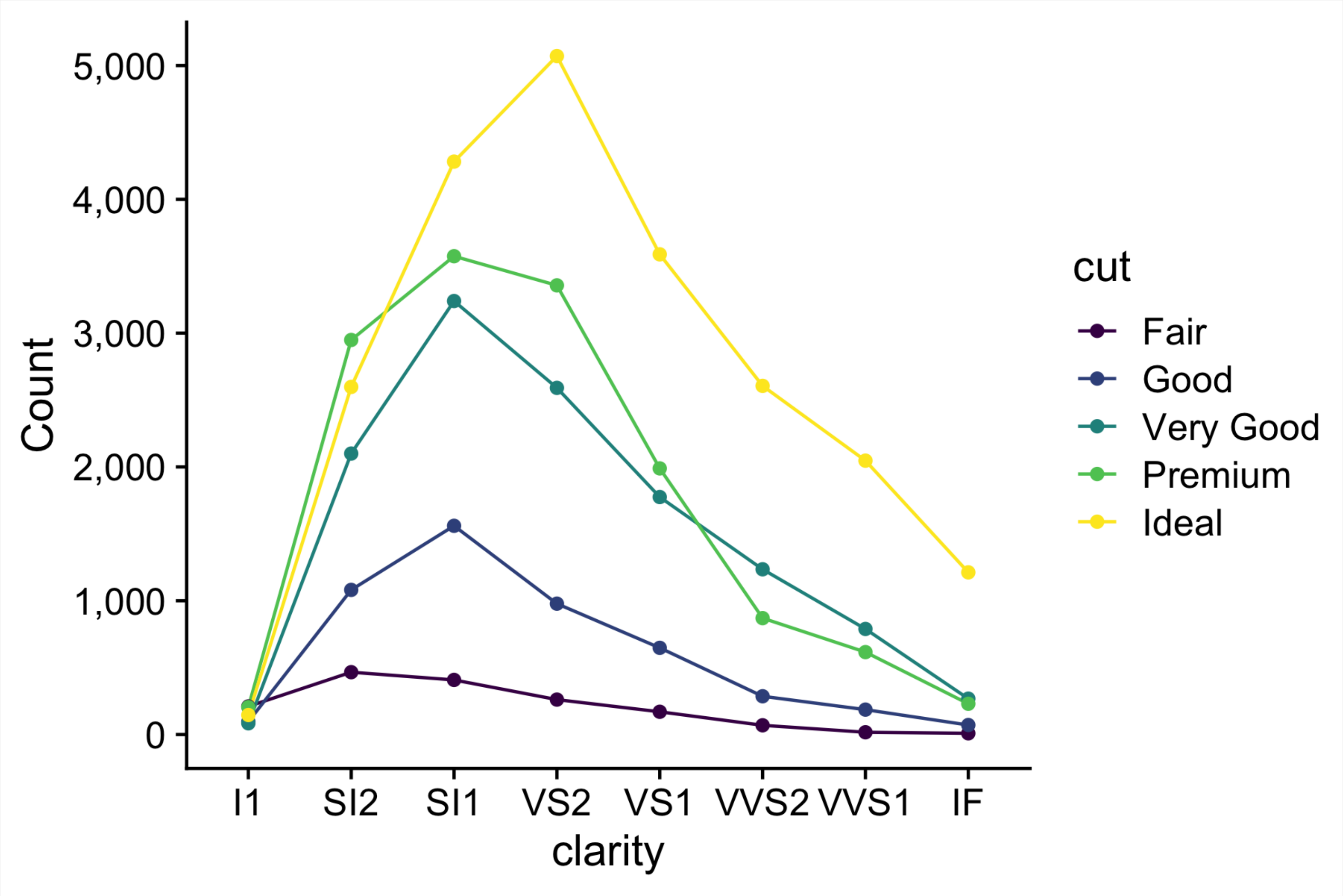
Law of Proximity

We tend to see elements that are *physically near* each other as part of the same object



Law of Proximity

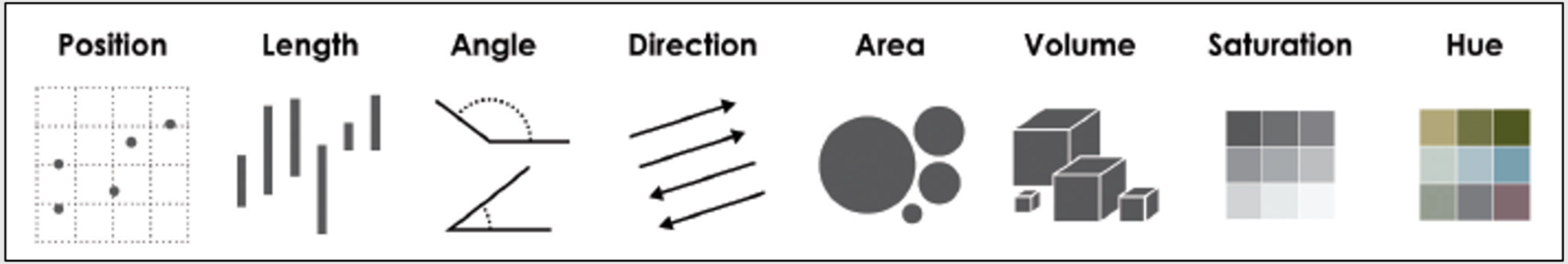
We tend to see elements that are *physically near* each other as part of the same object



Cleveland's operations of pattern perception:

1. Estimation
2. Assembly
3. Detection

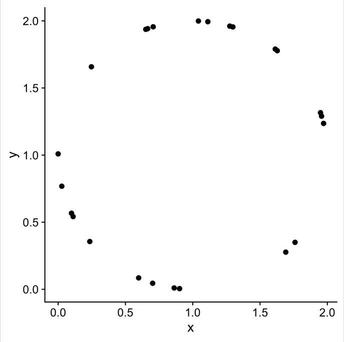
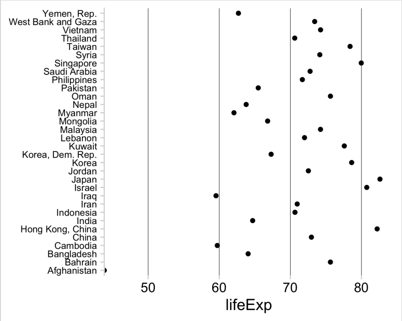
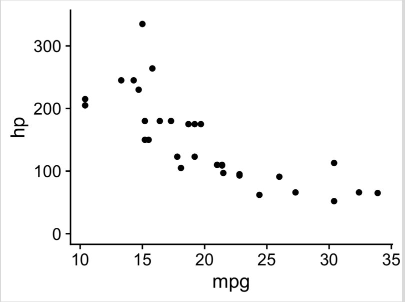
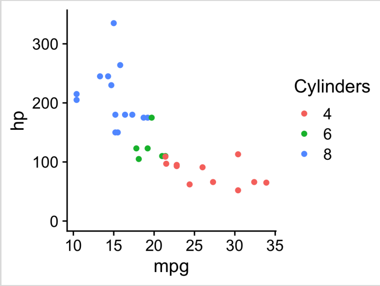
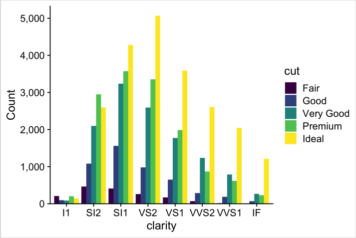
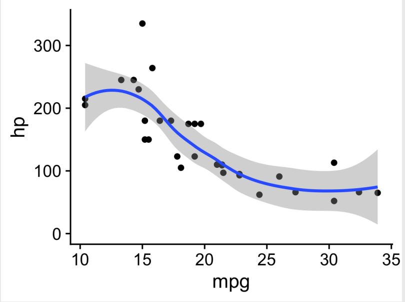
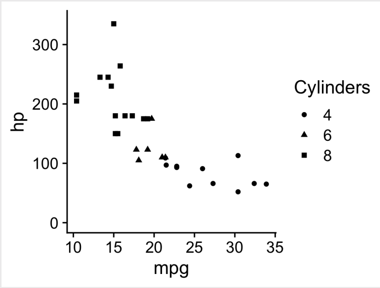
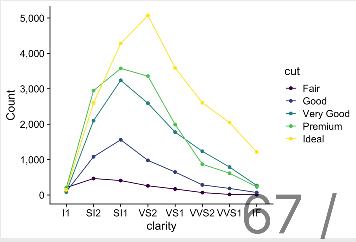
Estimation: Hierarchy for *numerical* data



More Accurate

Less Accurate

Assembly: Gestalt Psychology

Law of Closure	Prägnanz	Law of Continuity	Law of Similarity	Law of Proximity
Fill in the missing information	We like regular, simple, and orderly	Group together objects with established direction	Physically *similar* = same object	Physically *near* = same object
				
				

Cleveland's operations of pattern perception:

1. Estimation

2. Assembly

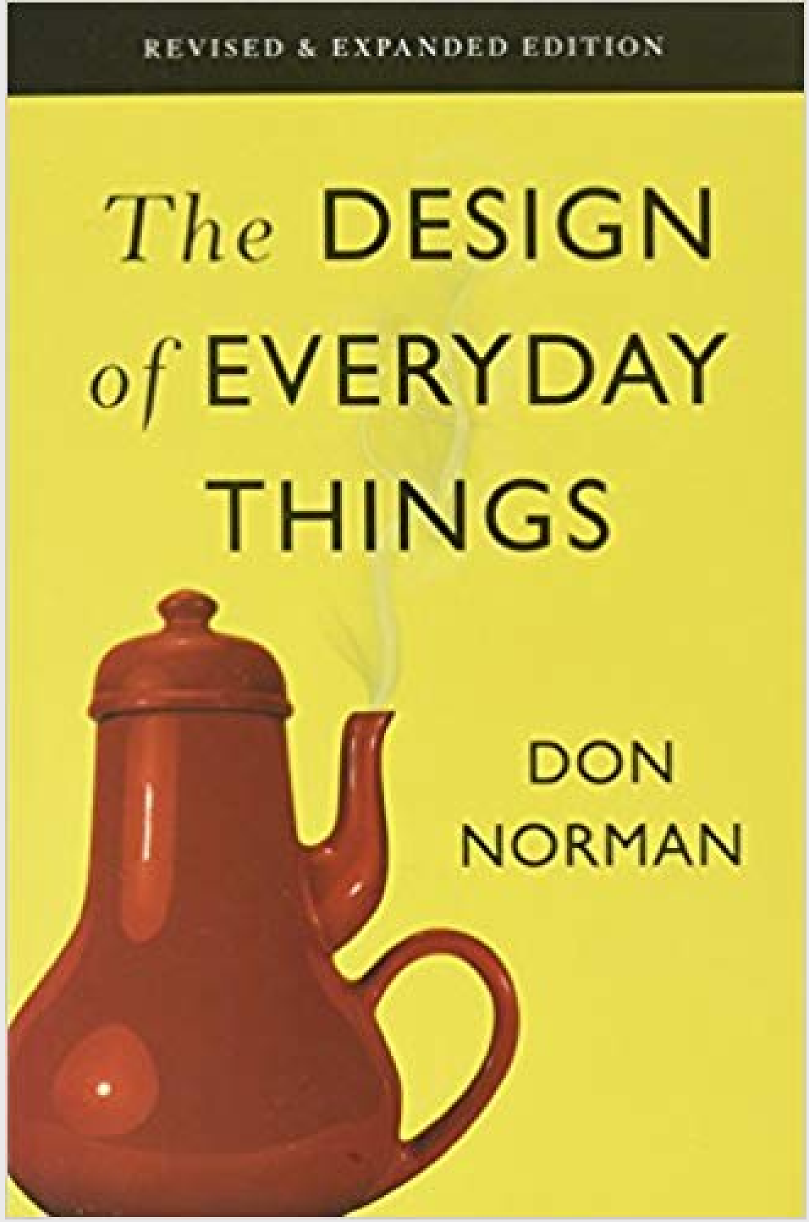
3. Detection -------->

Recognizing that a geometric object encodes a physical value

Norman door (n.):

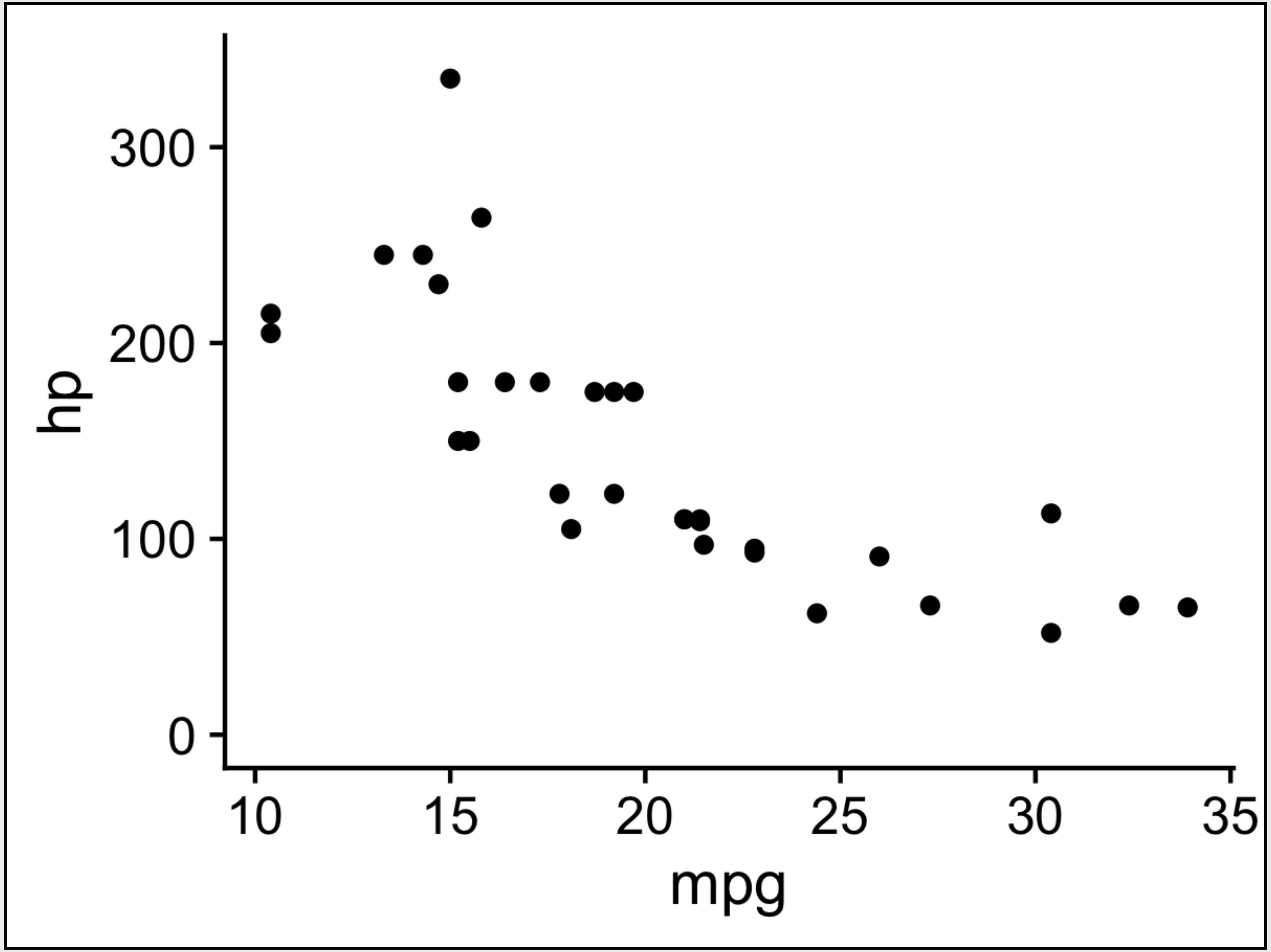
1. A door where the design tells you to do the opposite of what you're actually supposed to do.

2. A door that gives the wrong signal and needs a sign to correct it.

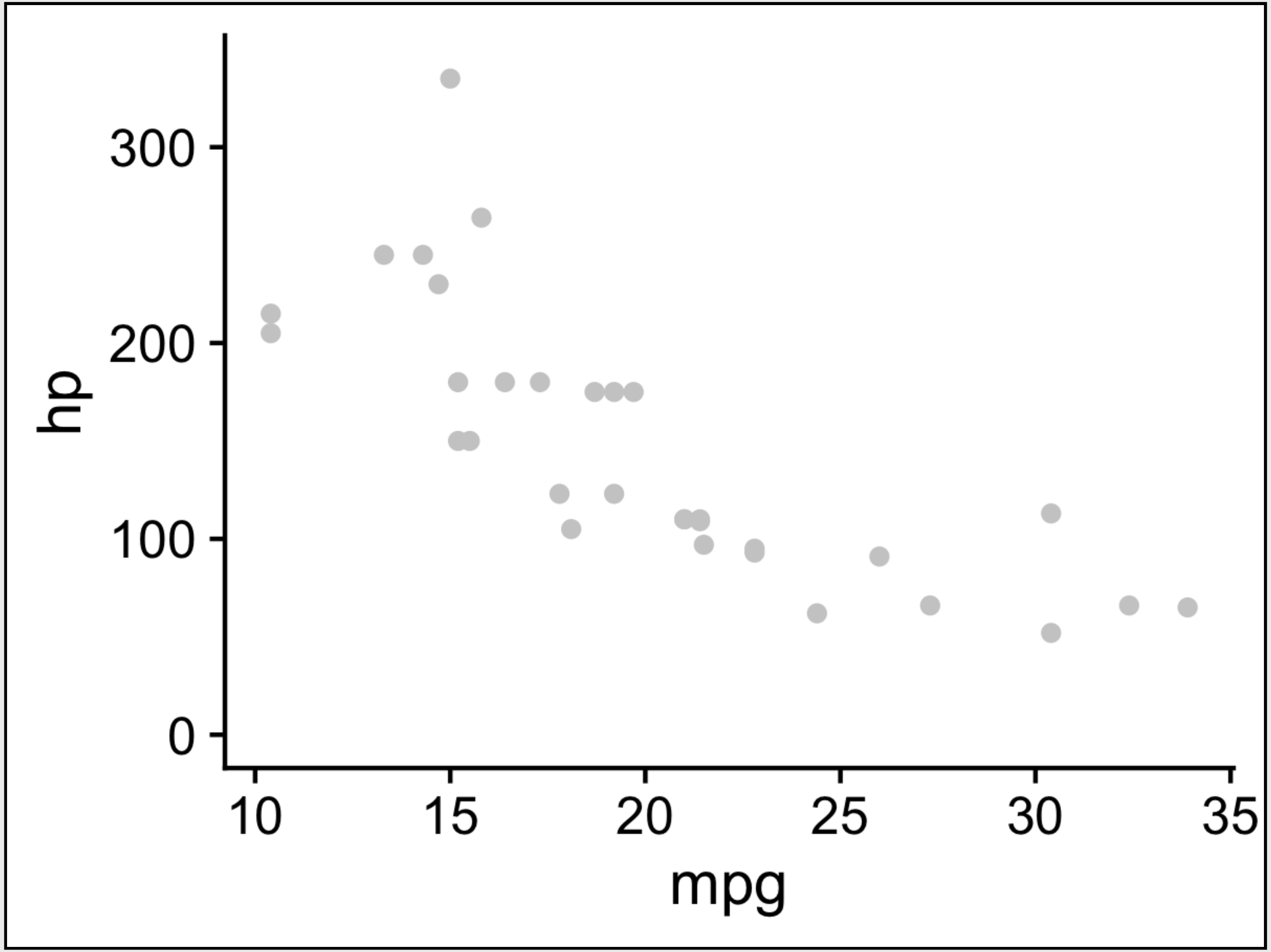


Norman door

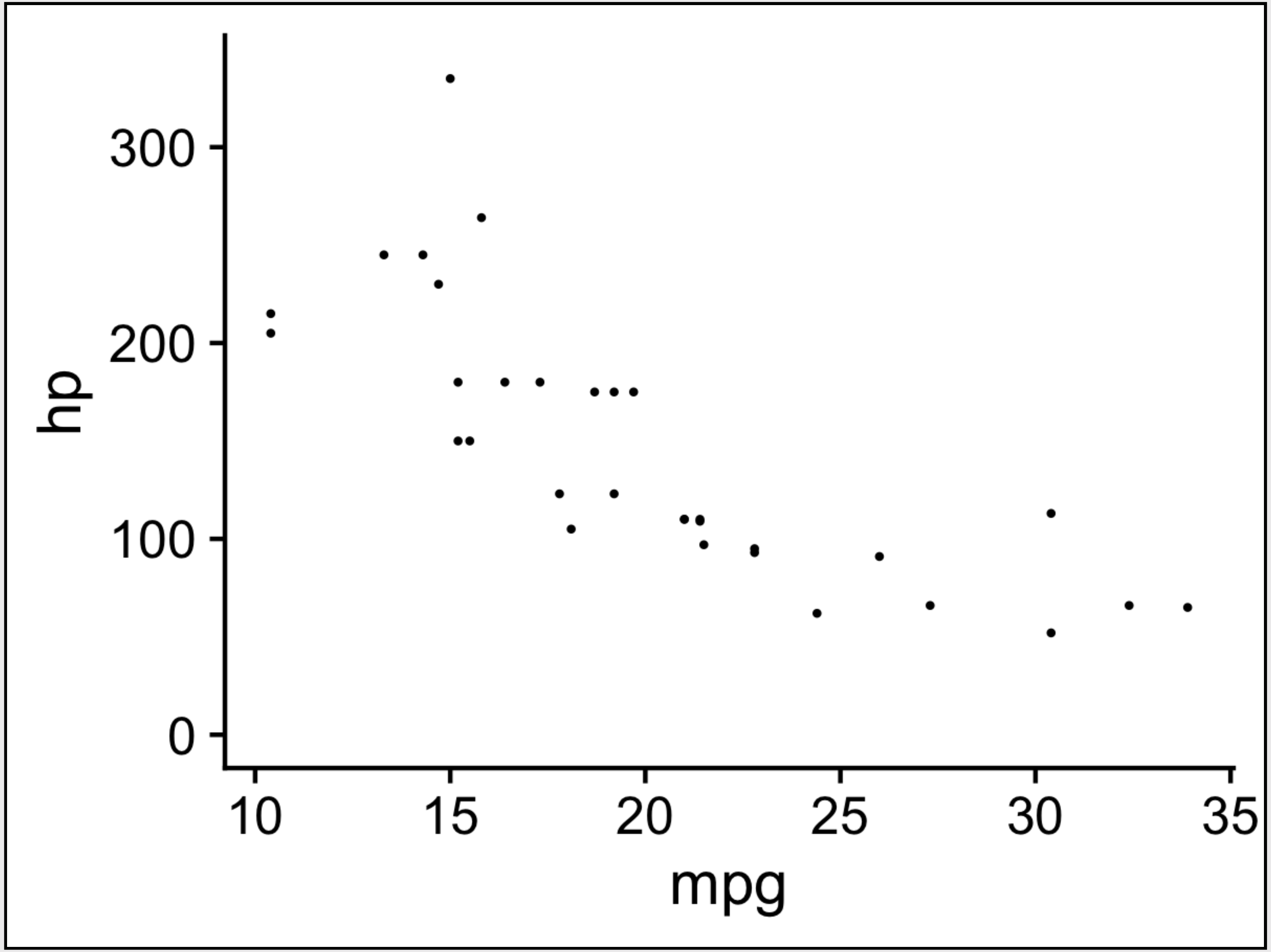
Non-Norman door



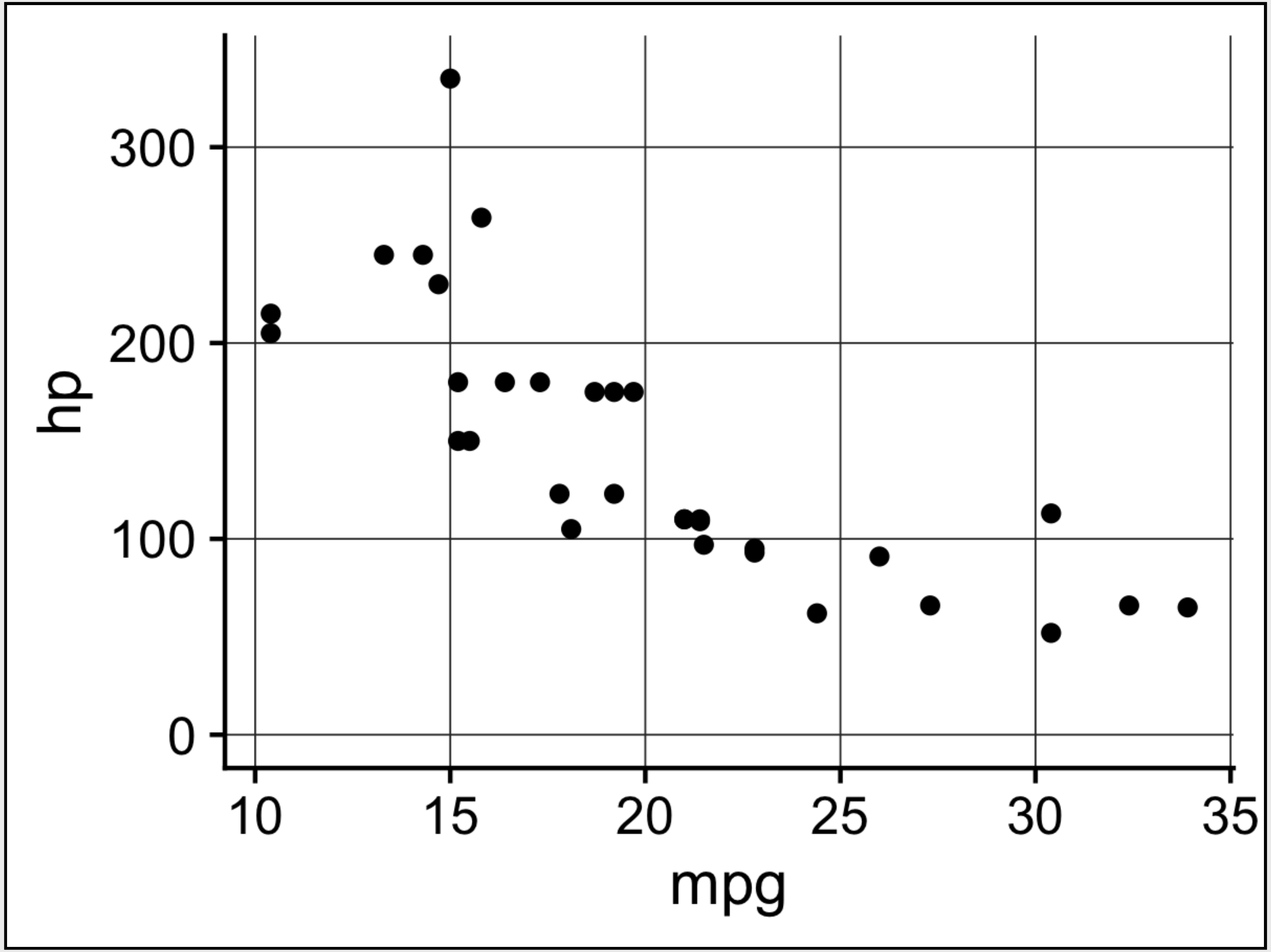
hp
300
200
100
0
10
15
20
25
30
35
mpg

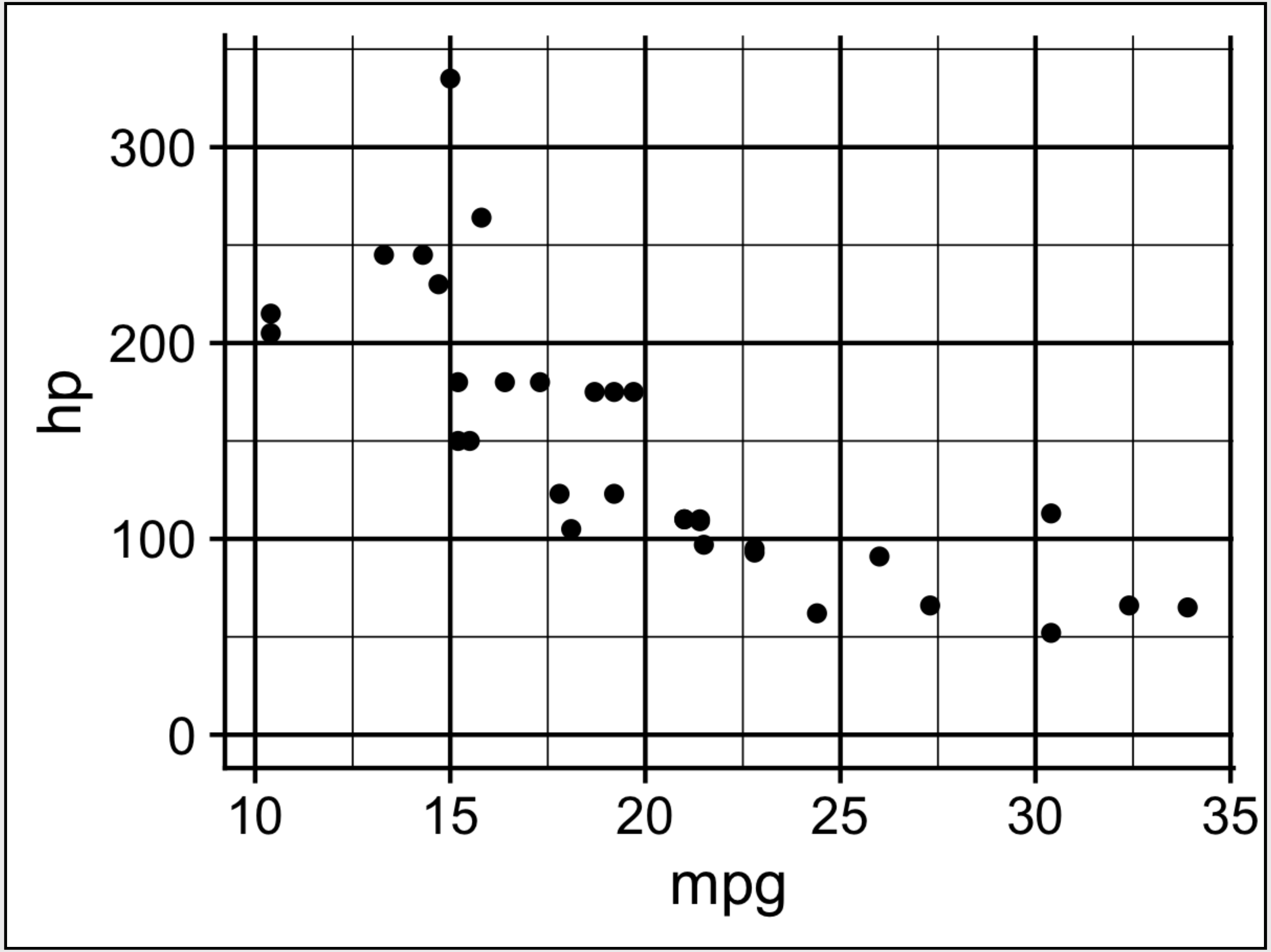


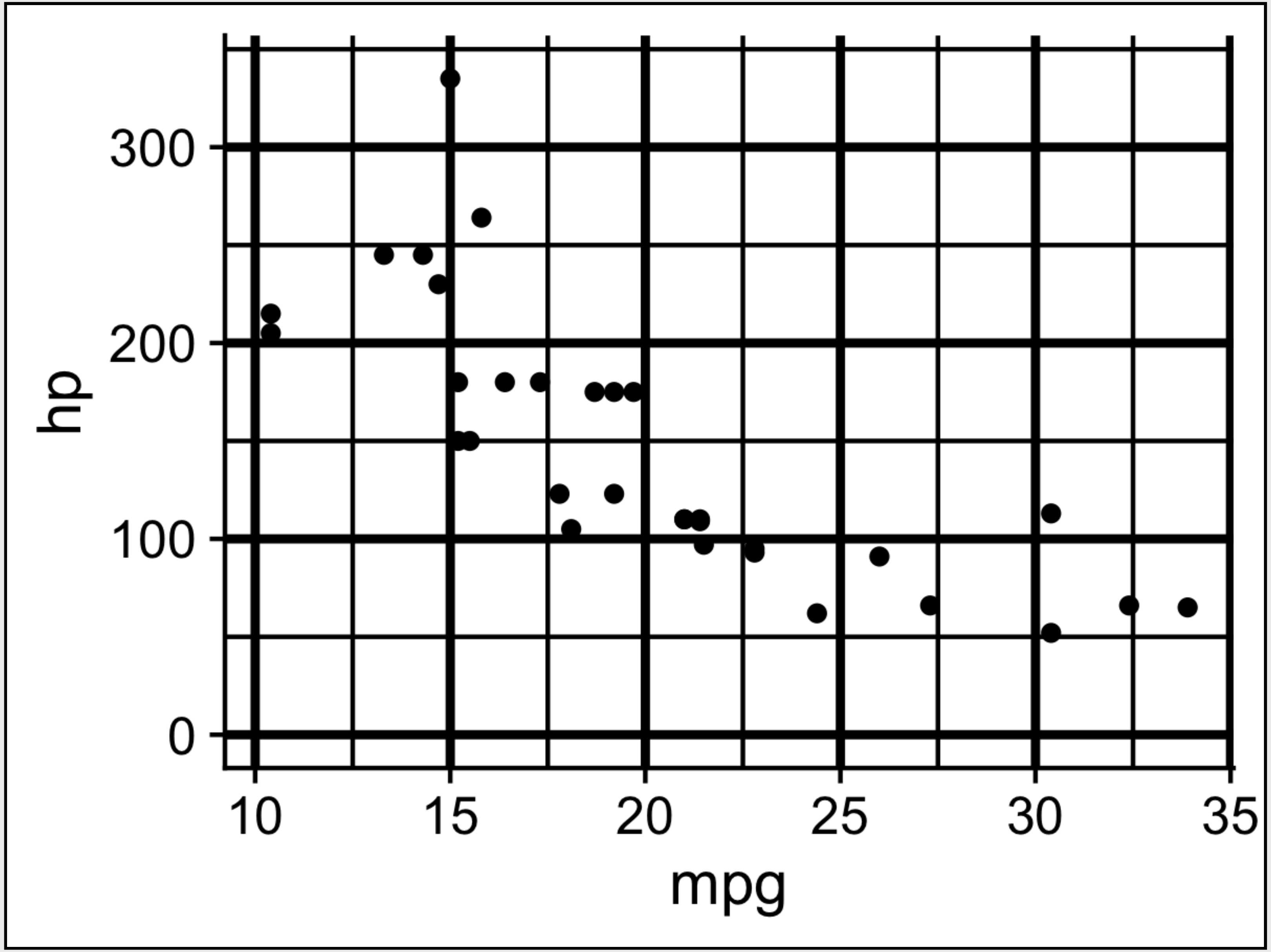
300
200
100
0
hp
10
15
20
25
30
35
mpg



hp
300
200
100
0
10
15
20
25
30
35
mpg

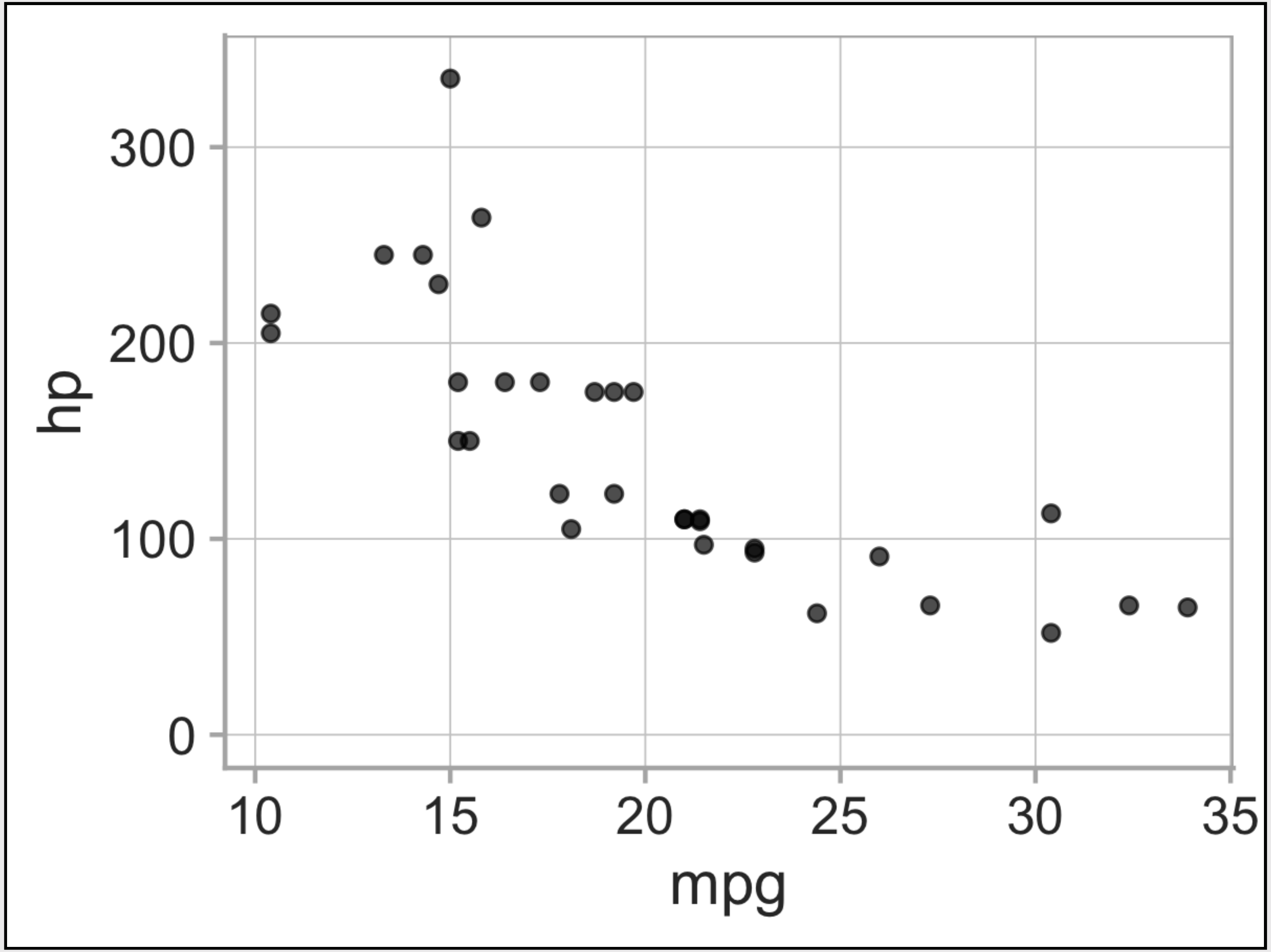

300
200
100
0
hp
10
15
20
25
30
35
mpg


300
200
100
0
hp
10
15
20
25
30
35
mpg



300
200
100
0
hp
10
15
20
25
30
35
mpg

The white circles you see at the intersections is called the **"Hermann Grid illusion"**



300
200
100
0
hp
10
15
20
25
30
35
mpg

Break!

Stand up, Move around, Stretch!

05:00

Week 5: *Visualizing Information*

1. The Human Visual-Memory System

2. The Psychology of Data Viz

BREAK

3. 10 Data Viz Best Practices

4. Making a (good) ggplot

10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

*most of the time

10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

*most of the time

"*Erase non-data ink.*"

— Ed Tufte

Figure 1.6: `Monstrous Costs' by Nigel Holmes, in Healy, 2018

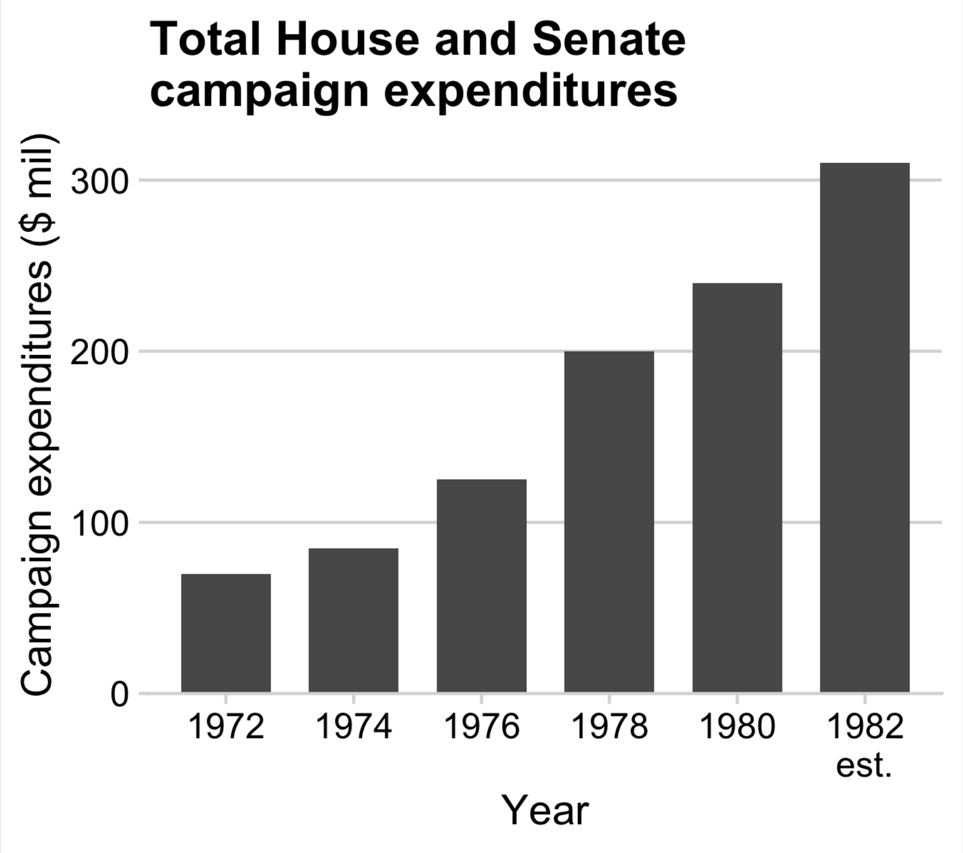
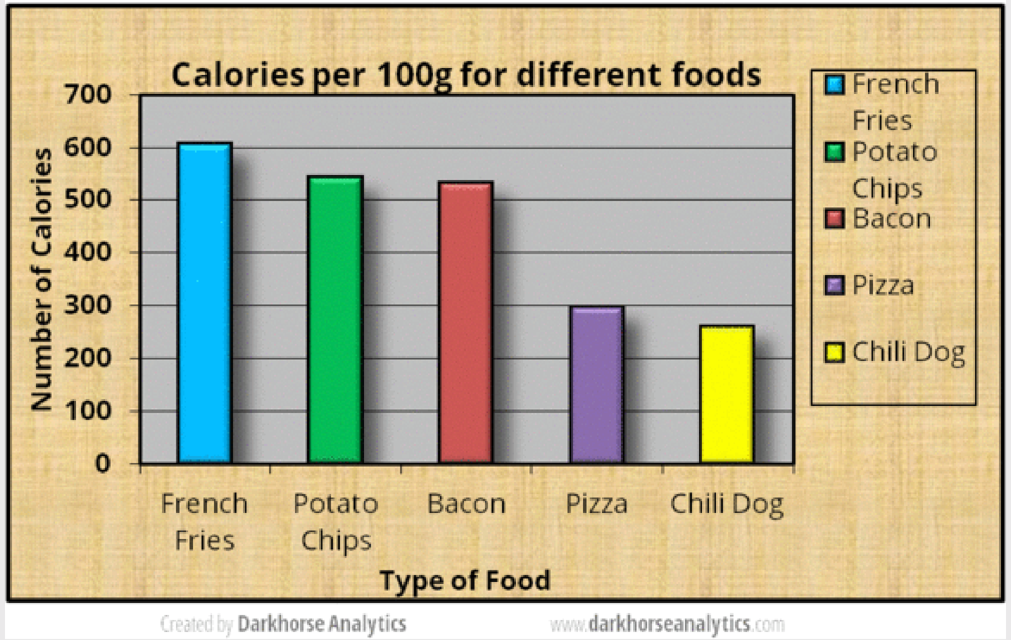
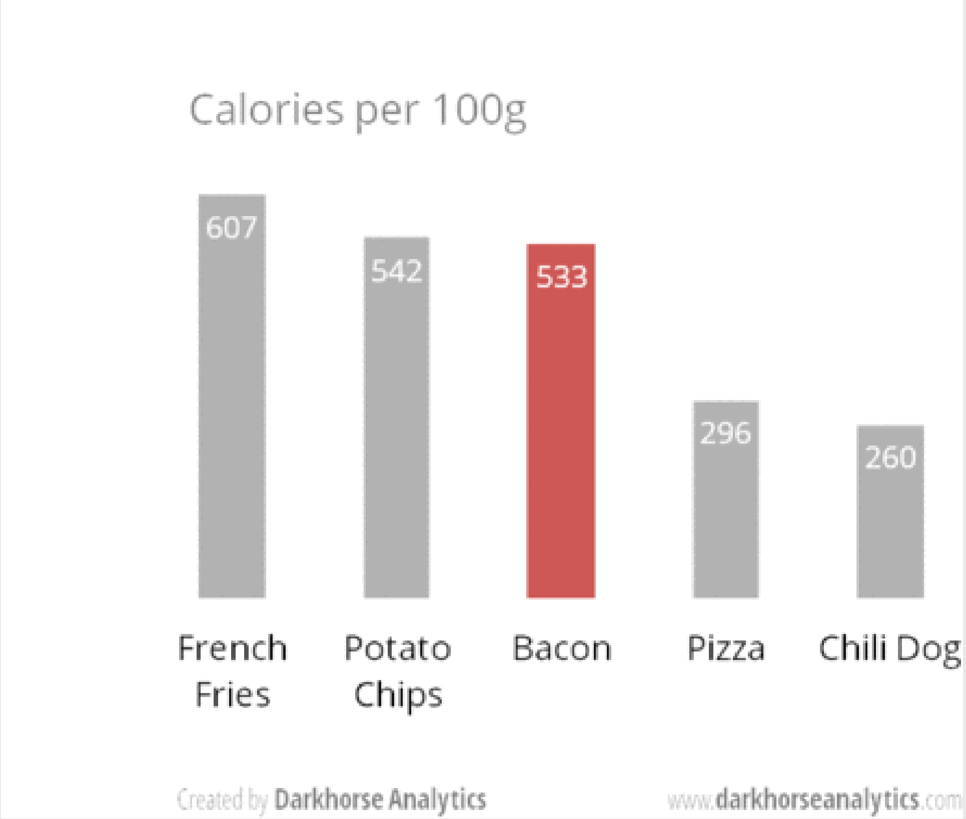


Figure 24.1: From Data Looks Better Naked by Darkhorse Analytics

Before



After

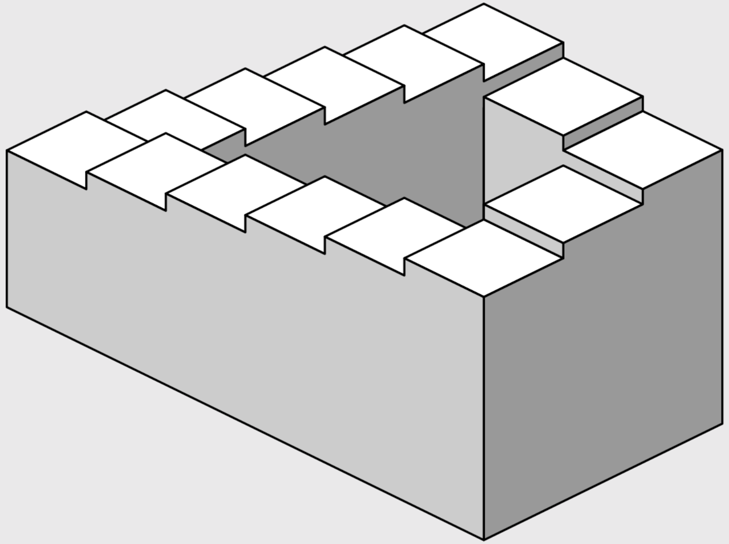


10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

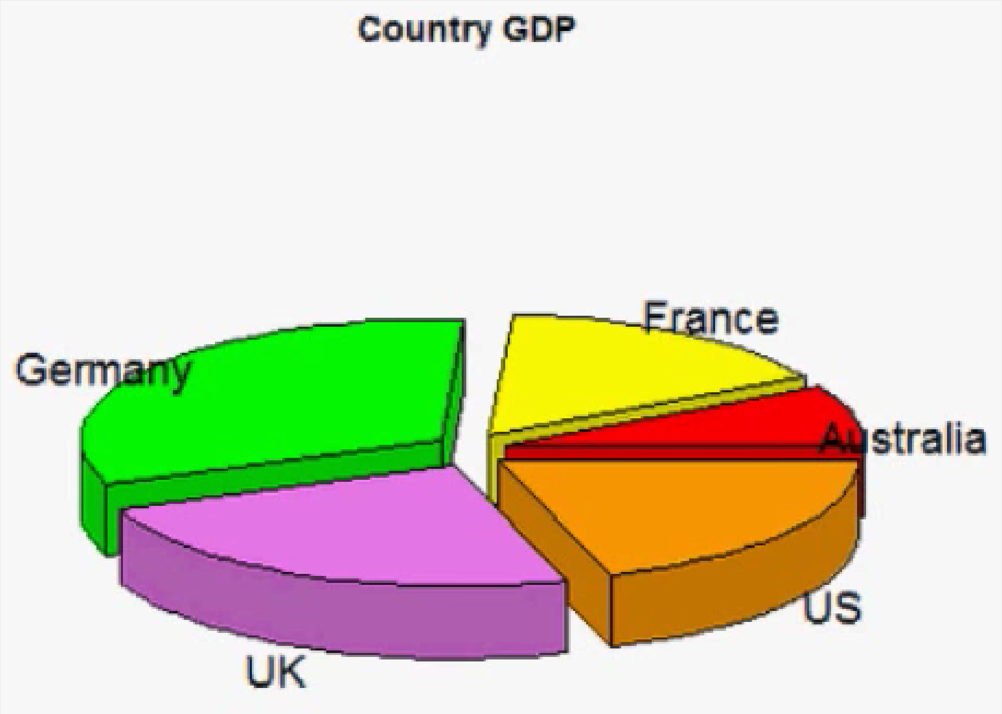
*most of the time

Humans aren't good at distinguishing 3D space

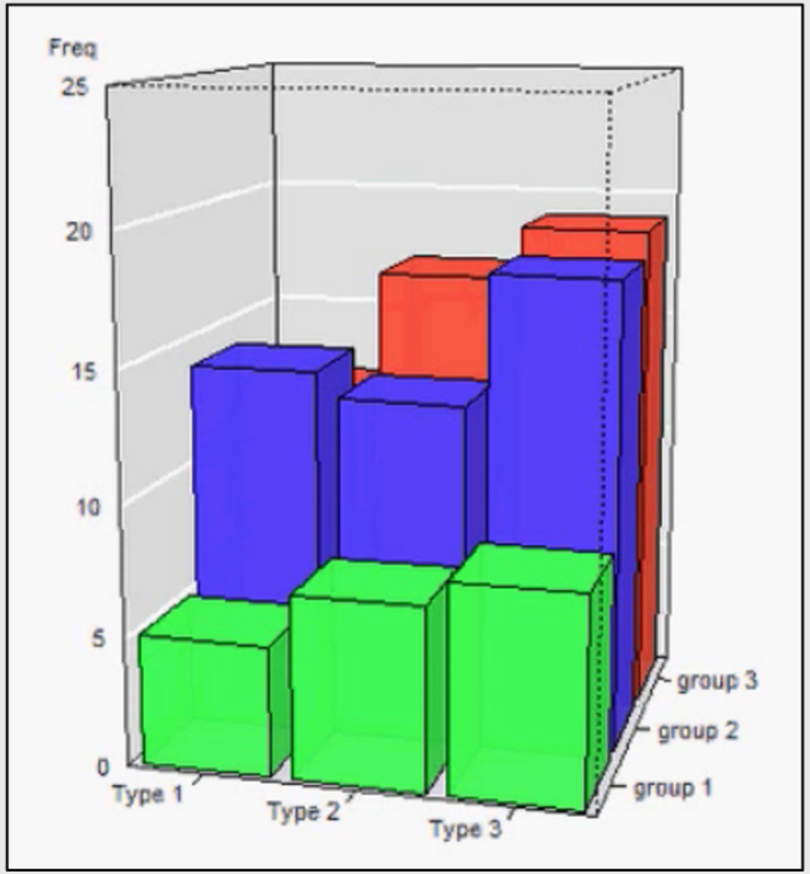


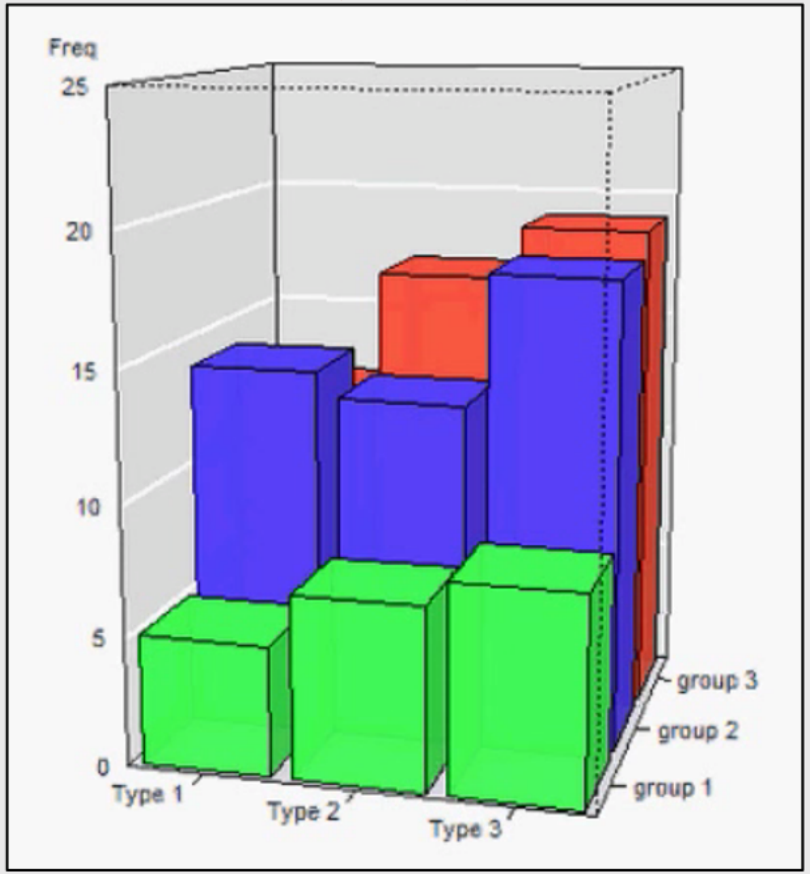
Penrose Stairs, made famous by M.C. Escher (1898-1972)

Ink proportions != true proportions

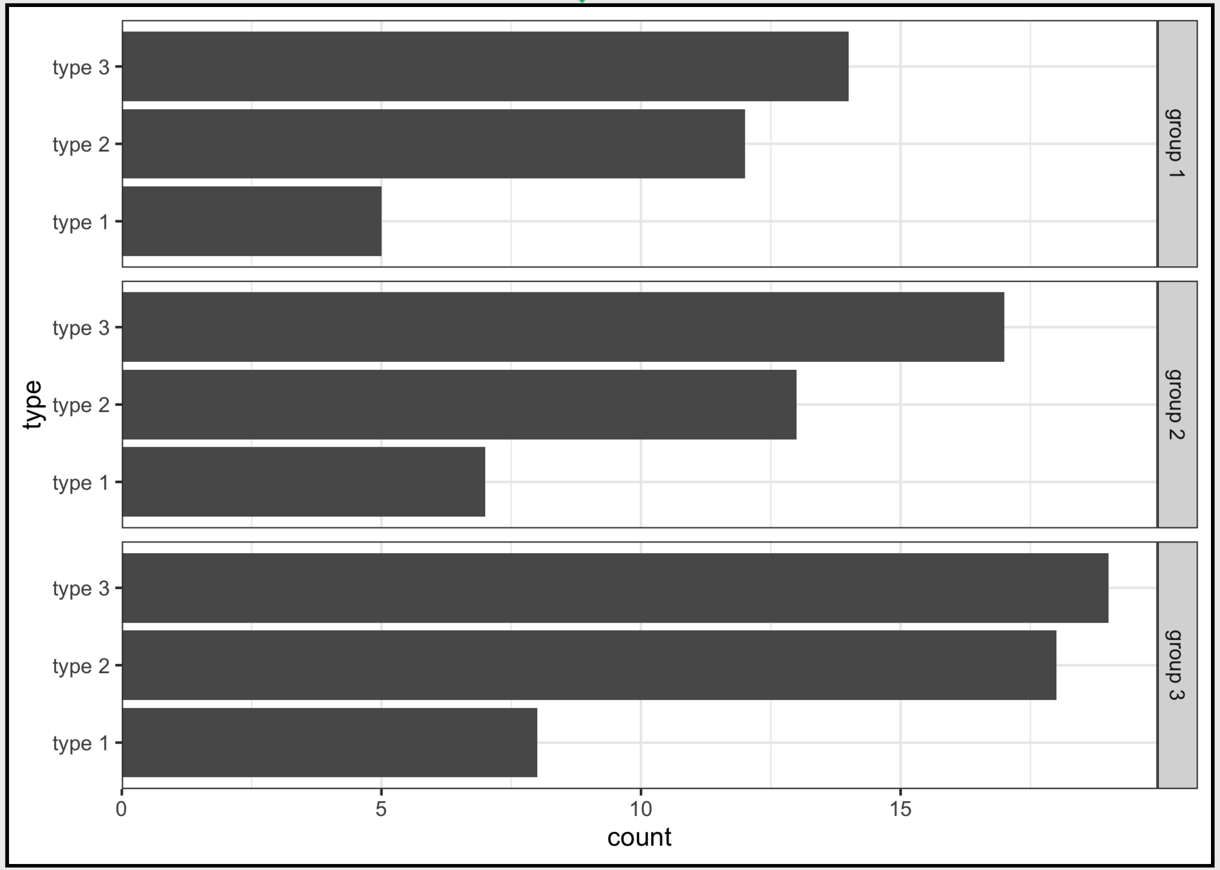


Occlusion: geoms are obscured





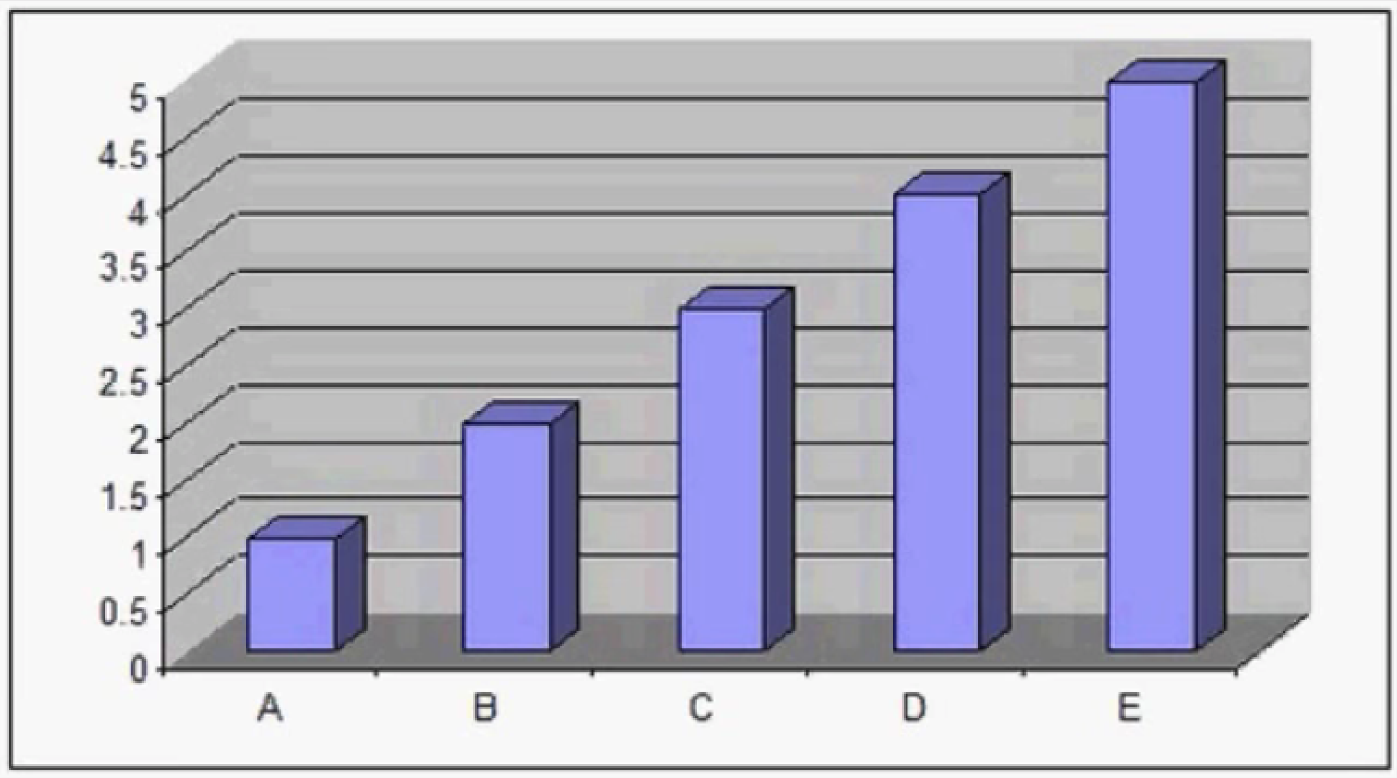
Freq
25
20
15
10
5
0
Type 1
Type 2
Type 3
group 3
group 2
group 1

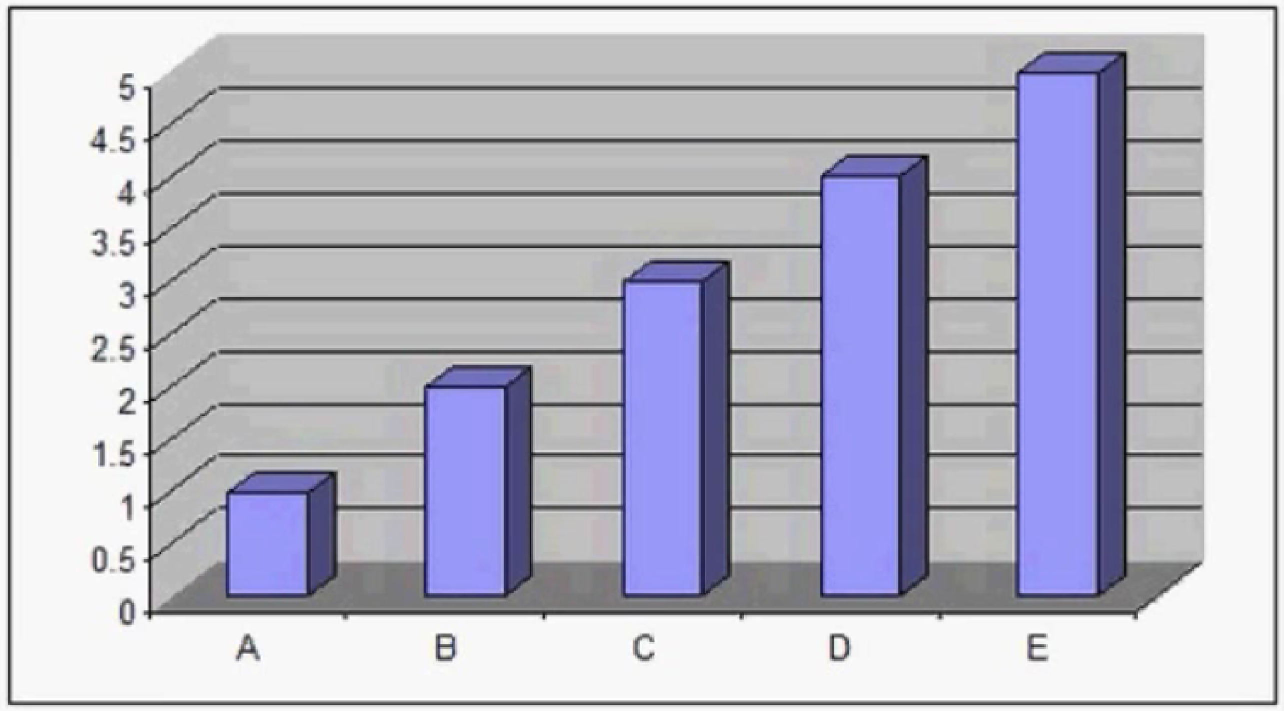


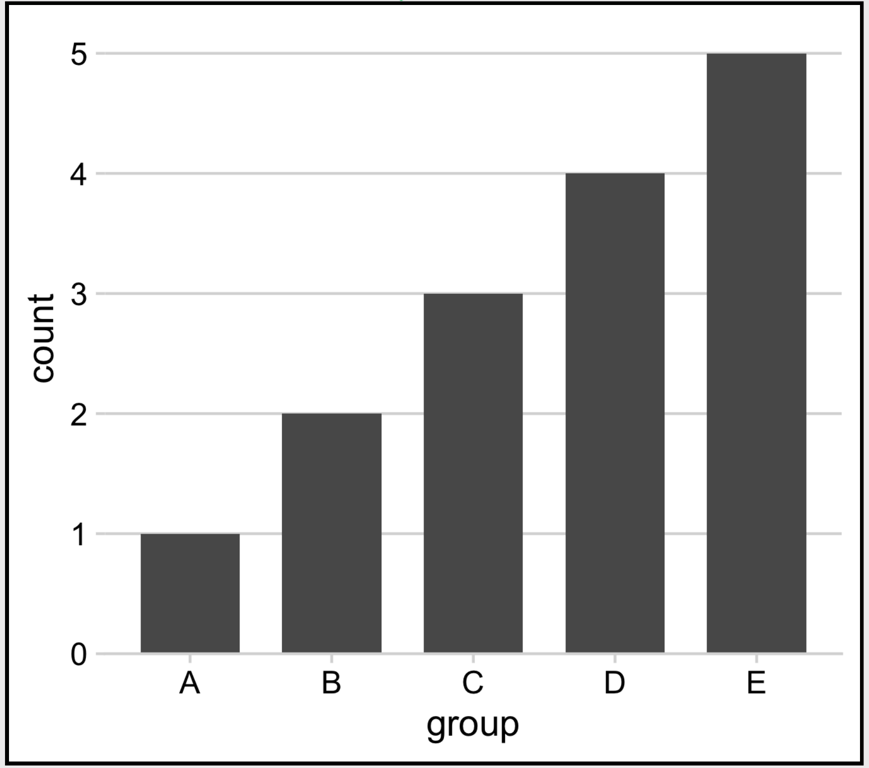
type 3
type 2
type 1
group 1
group 2
group 3
type
0
5
10
15
count

The third dimension distracts from the data

(this is what Tufte calls "chart junk")




5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
A
B
C
D
E


5
4
3
2
1
0
count
A
B
C
D
E
group

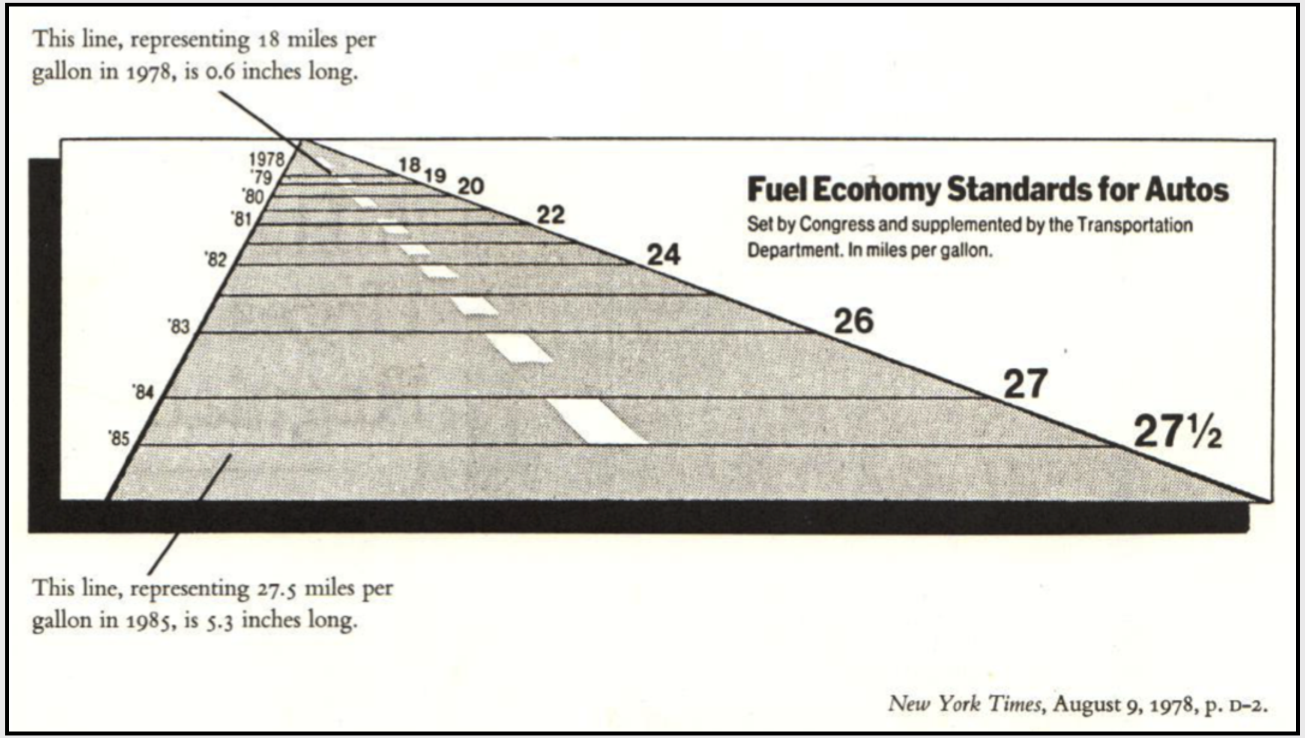
10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

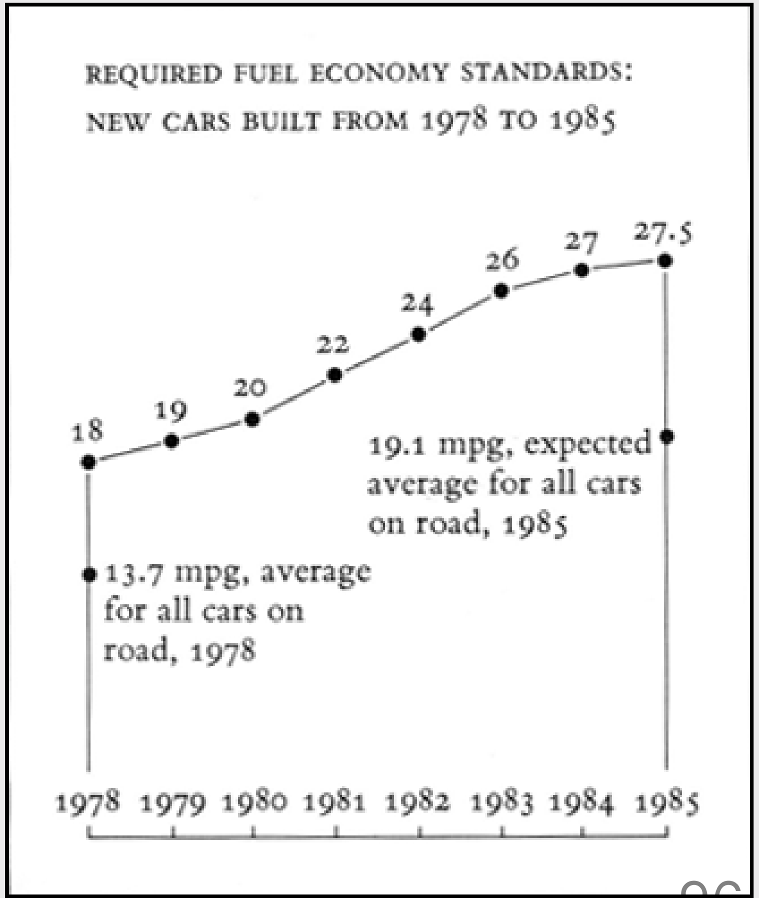
*most of the time

$$\text{"Lie Factor"} = \frac{\text{Size of effect in graphic}}{\text{Size of effect in data}}$$

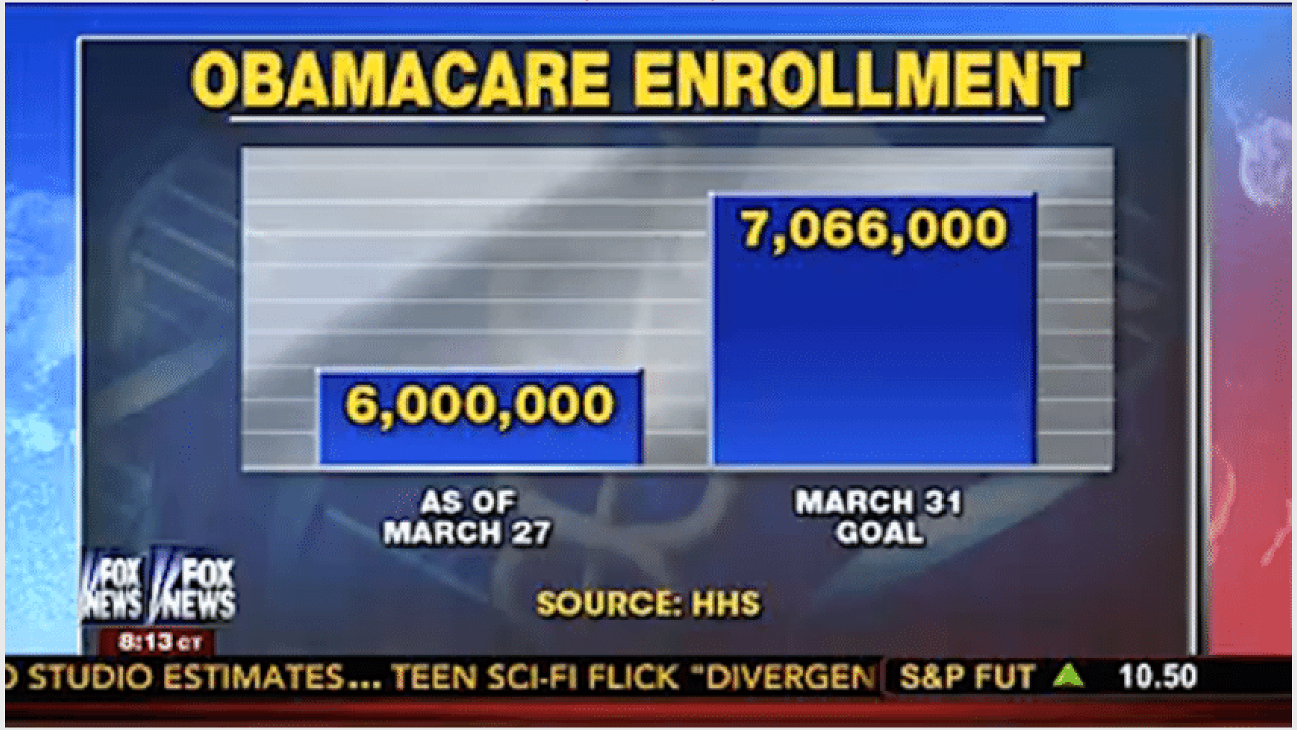
$$\text{"Lie Factor"} = \frac{\text{Size of effect in graphic}}{\text{Size of effect in data}} = \frac{\frac{5.3-0.6}{0.6}}{\frac{27.5-18}{18}} = \frac{7.83}{0.53} = 14.8$$



Edward Tufte (2001) "The Visual Display of Quantitative Information", 2nd Edition, pg. 57-58.



Bar charts should always start at 0



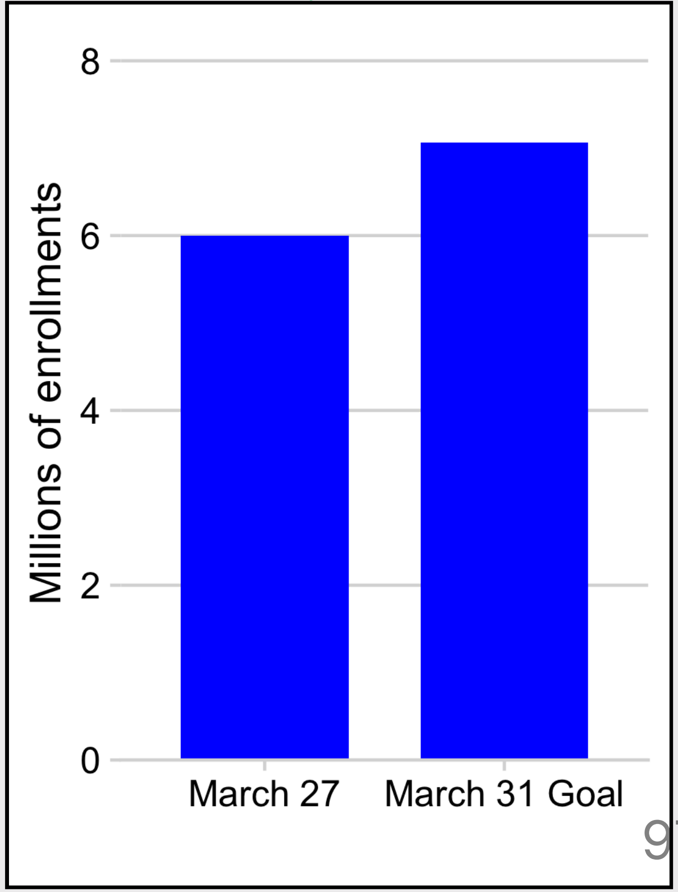
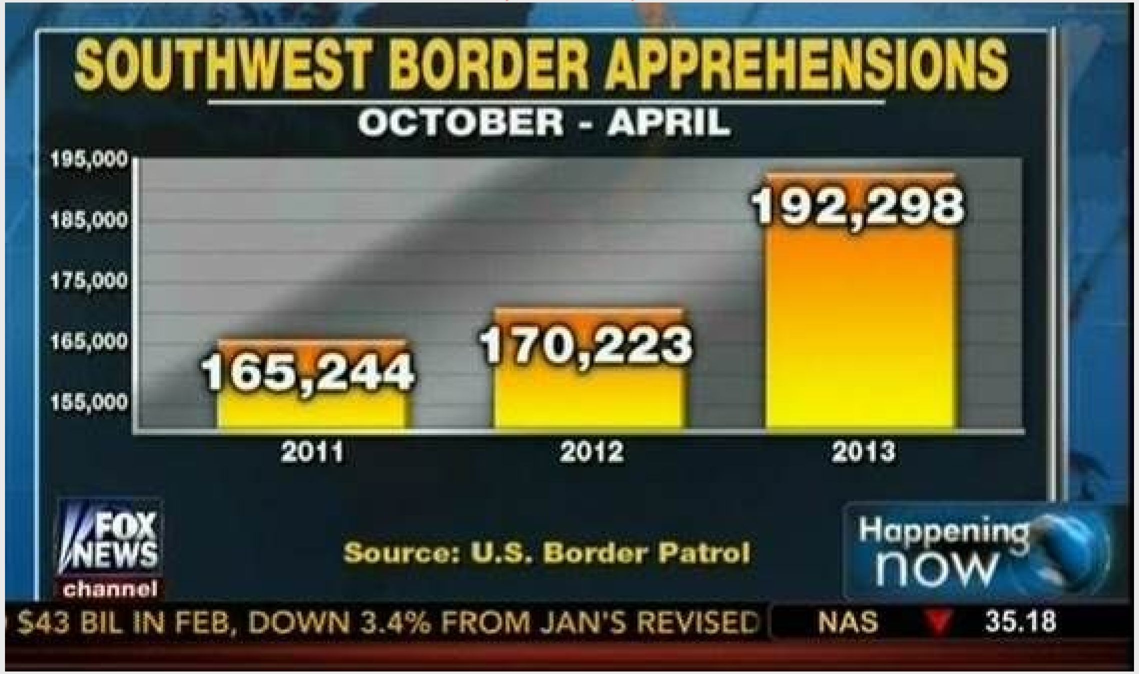


Image from http://livingqlikview.com/the-9-worst-data-visualizations-ever-created/

Don't cherry-pick your data



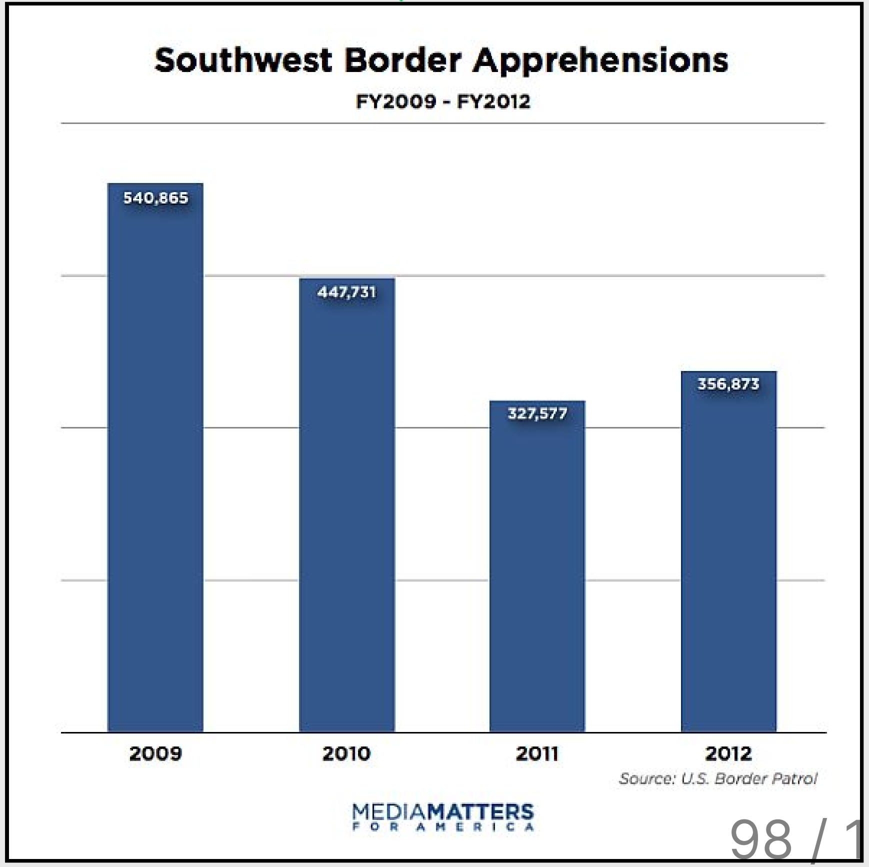


Image from https://www.mediamatters.org/fox-news/fox-news-newest-dishonest-chart-immigration-enforcement

Make sure your chart makes sense

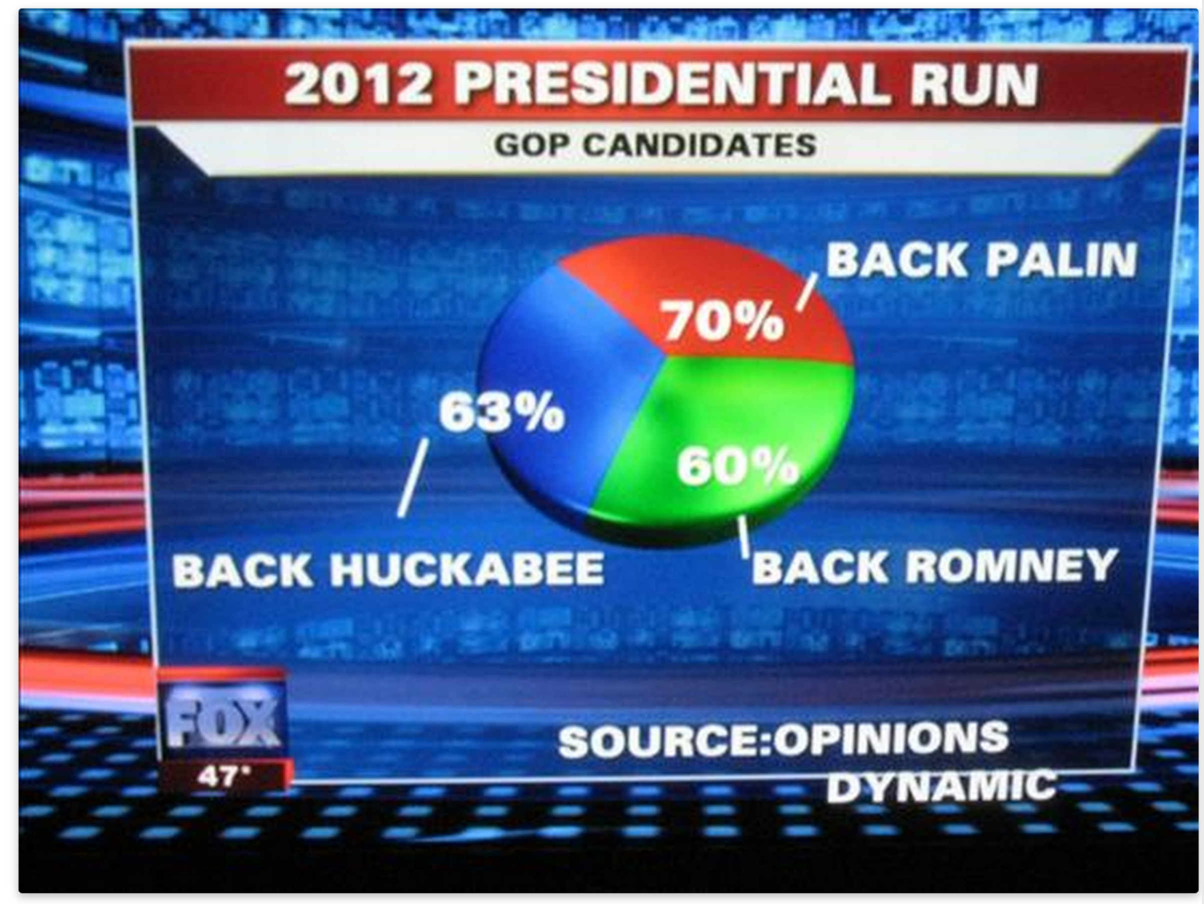
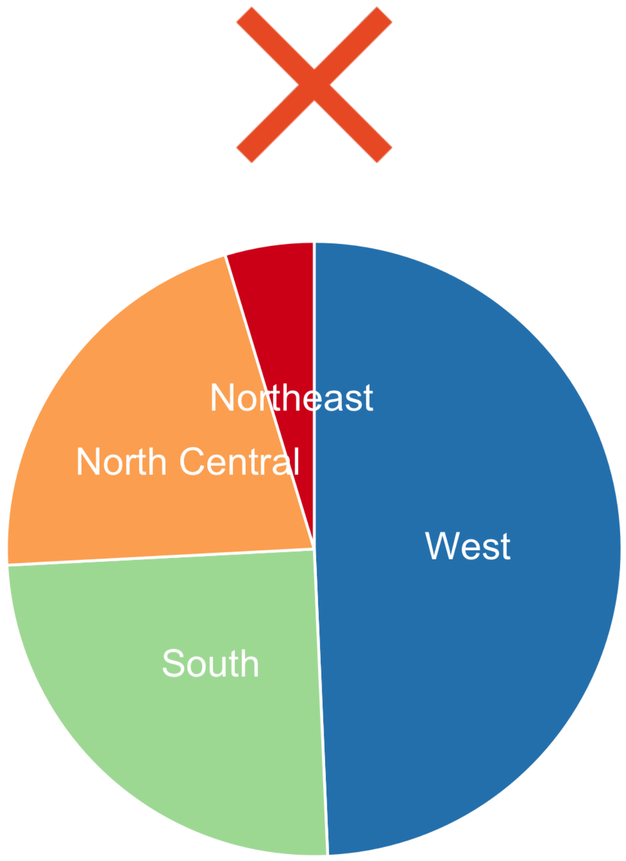


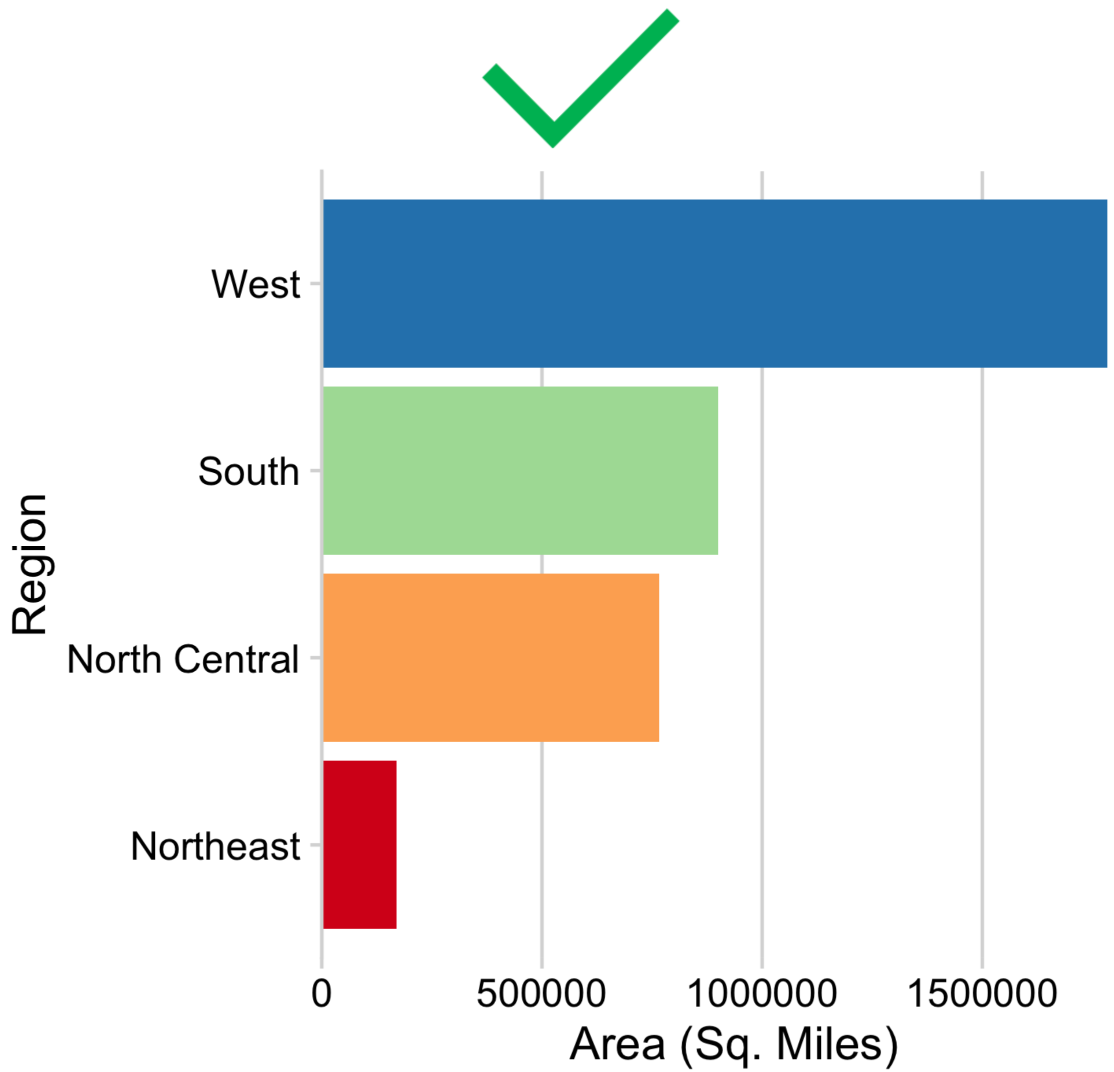
Image from Image from http://livingqlikview.com/the-9-worst-data-visualizations-ever-created/

10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

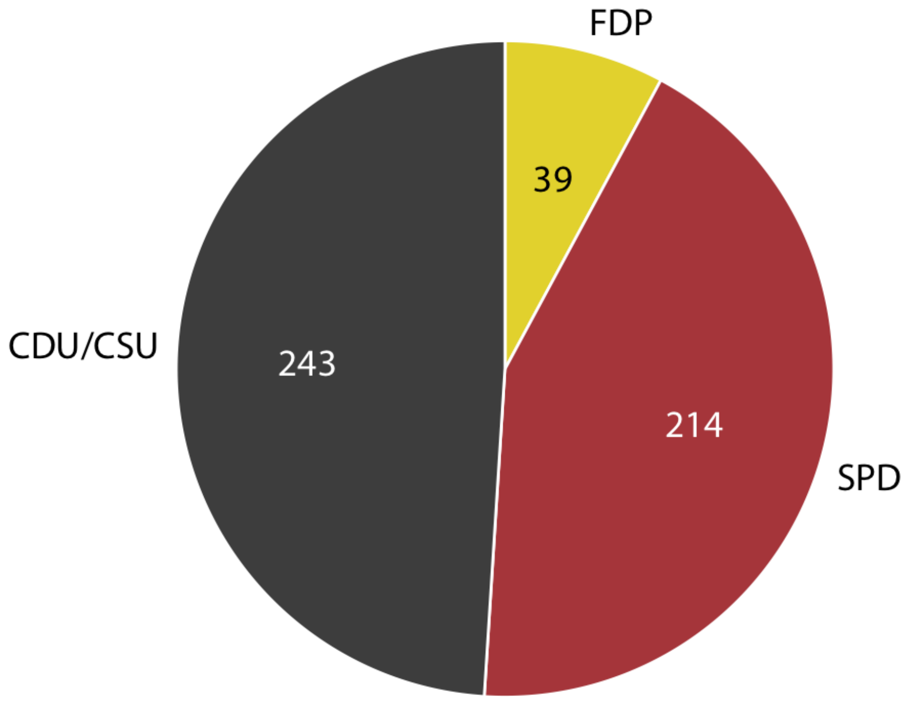
*most of the time


Northeast
North Central
West
South

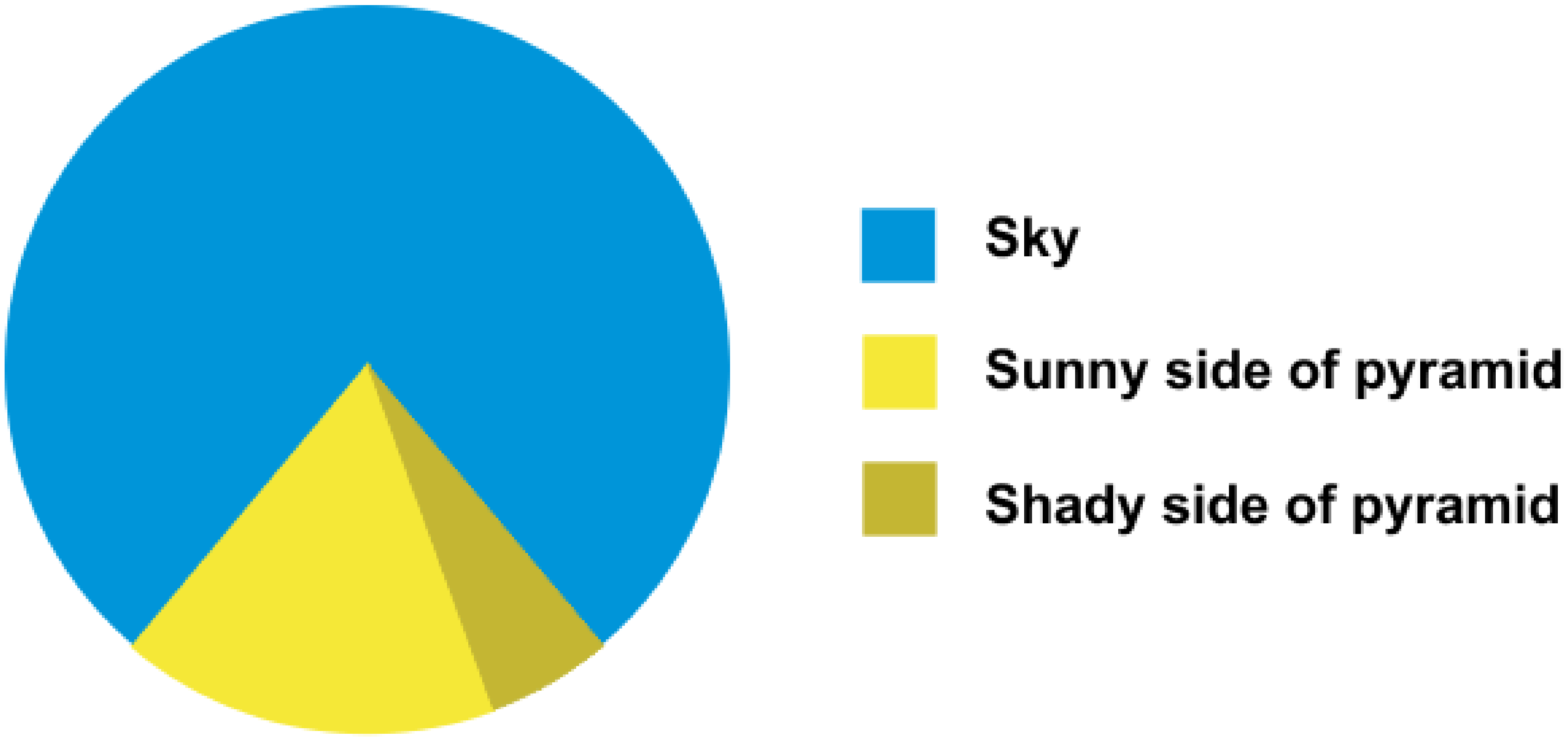

West
South
North Central
Northeast
Region
0
500000
1000000
1500000
Area (Sq. Miles)

Exceptions:

- Small data
- Simple fractions
- If sum of parts matters



Best pie chart of all time

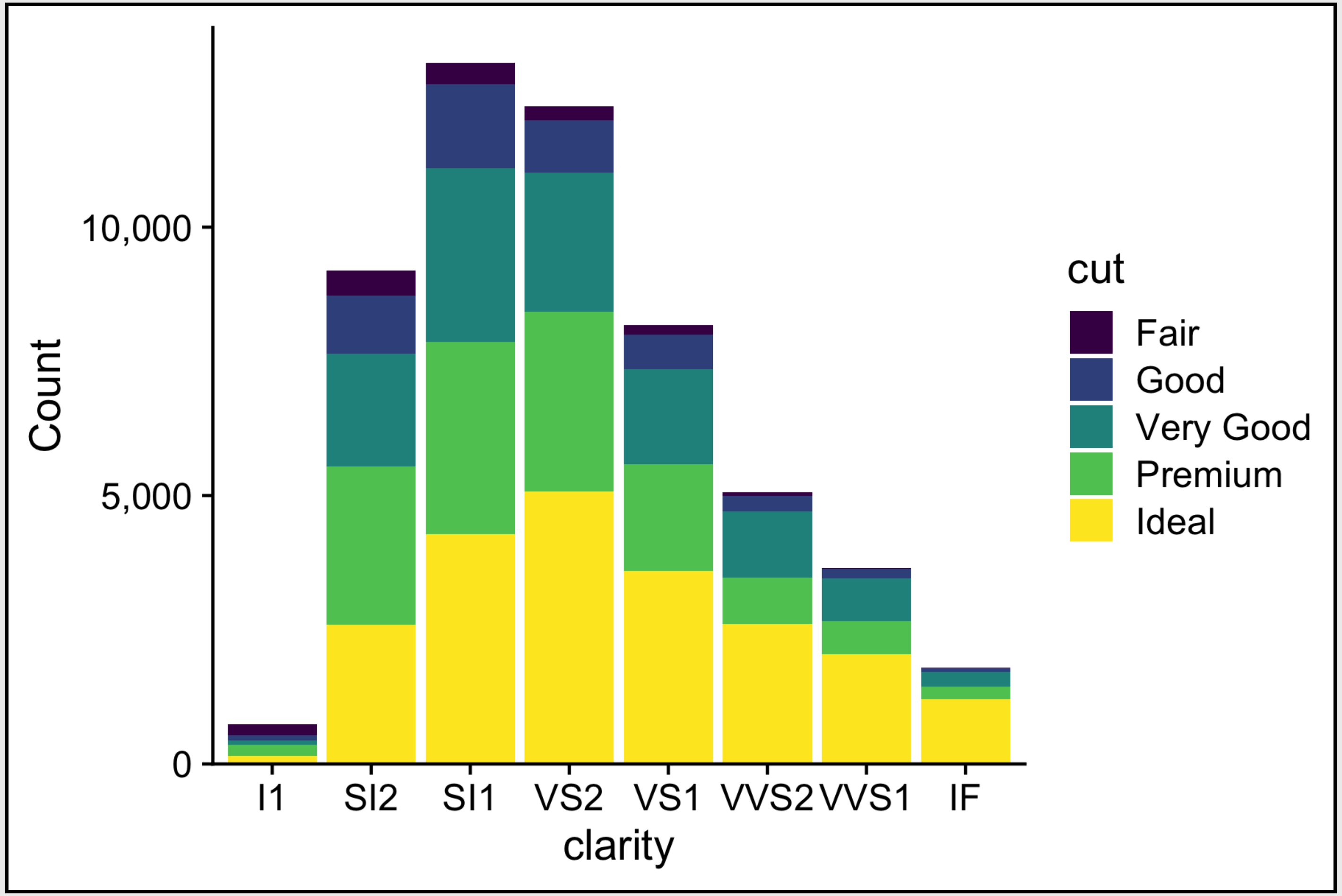


10 Data Viz Best Practices

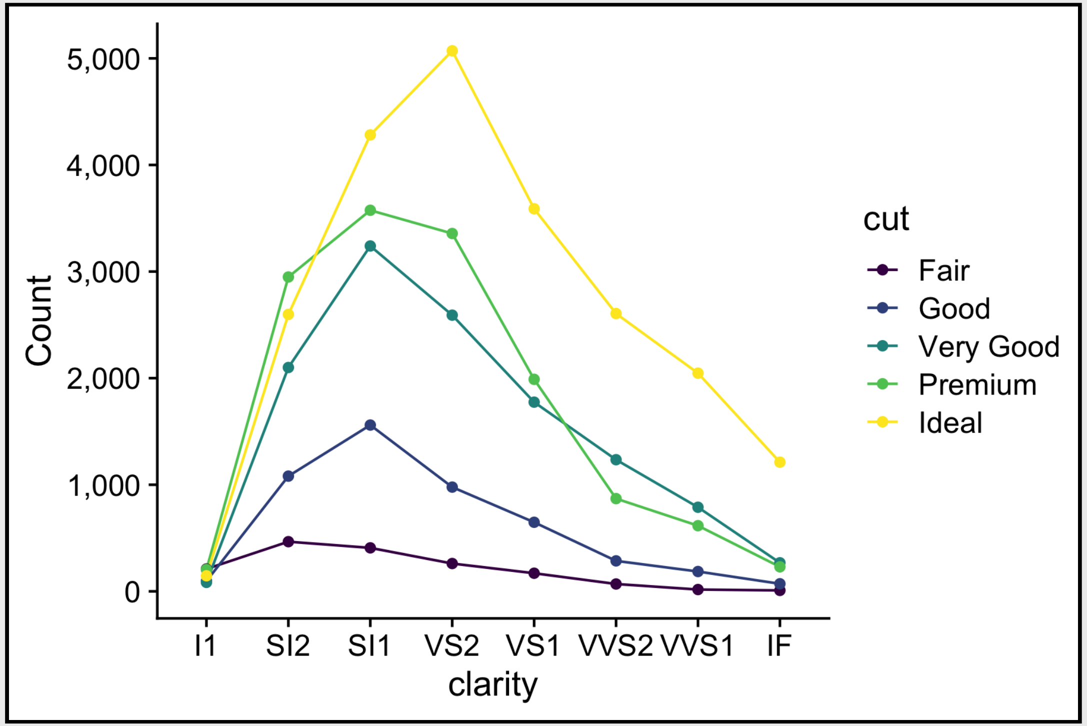
1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

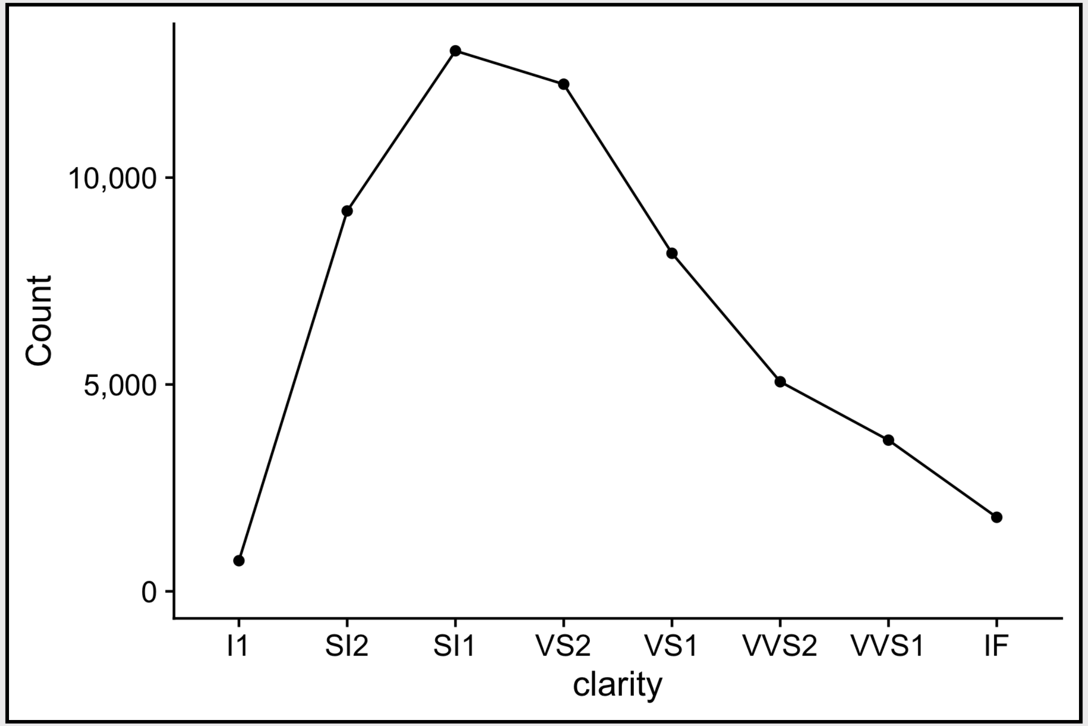
*most of the time

Stacked bars are rarely a good idea

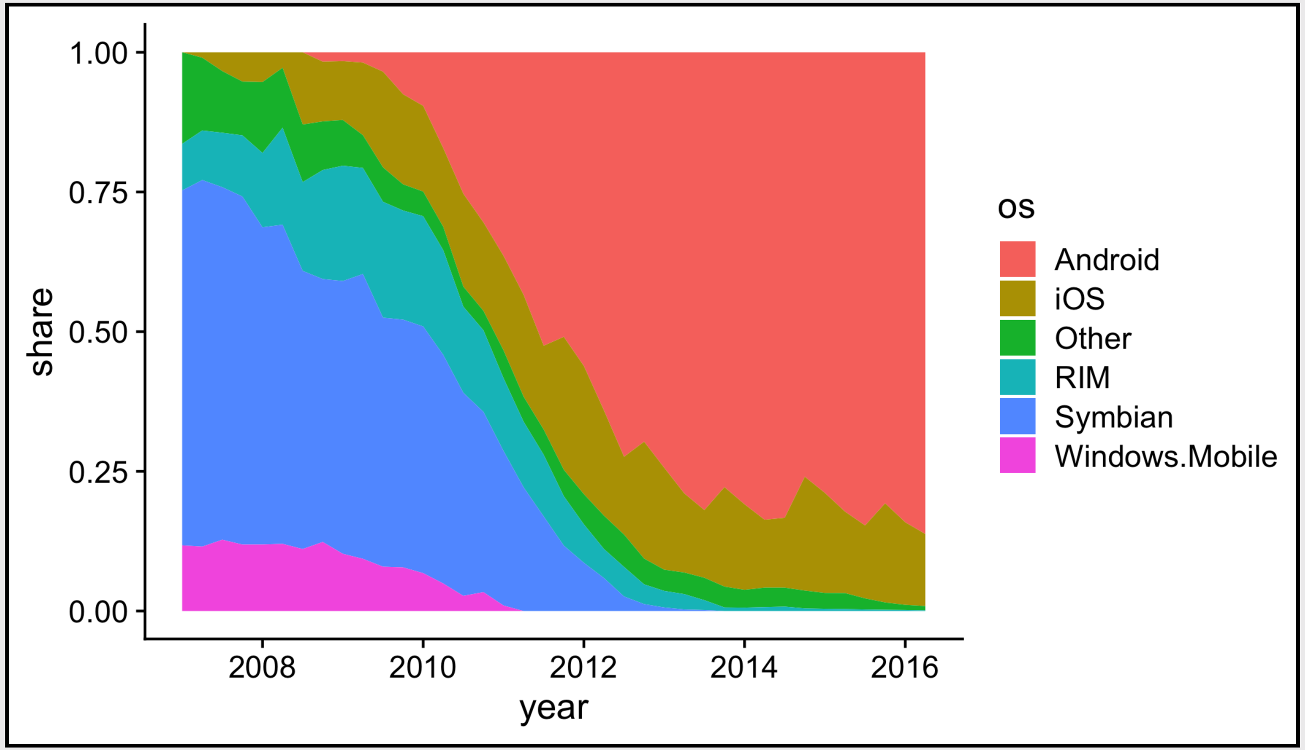


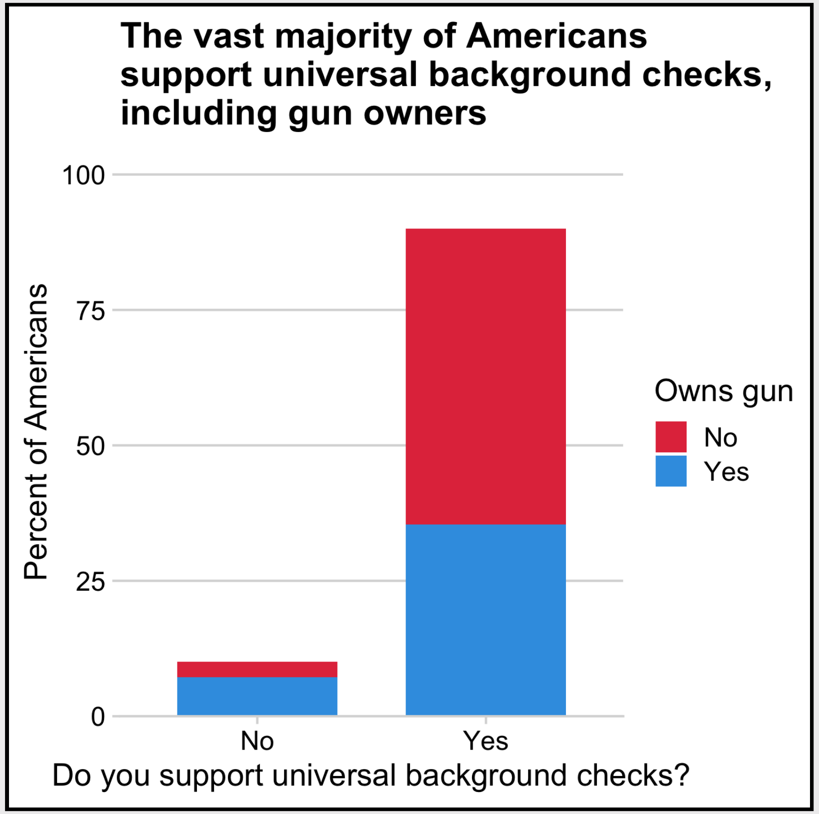
"Parallel coordinates" plot usually works better





Exception:
When you care about the *total* more than the categories



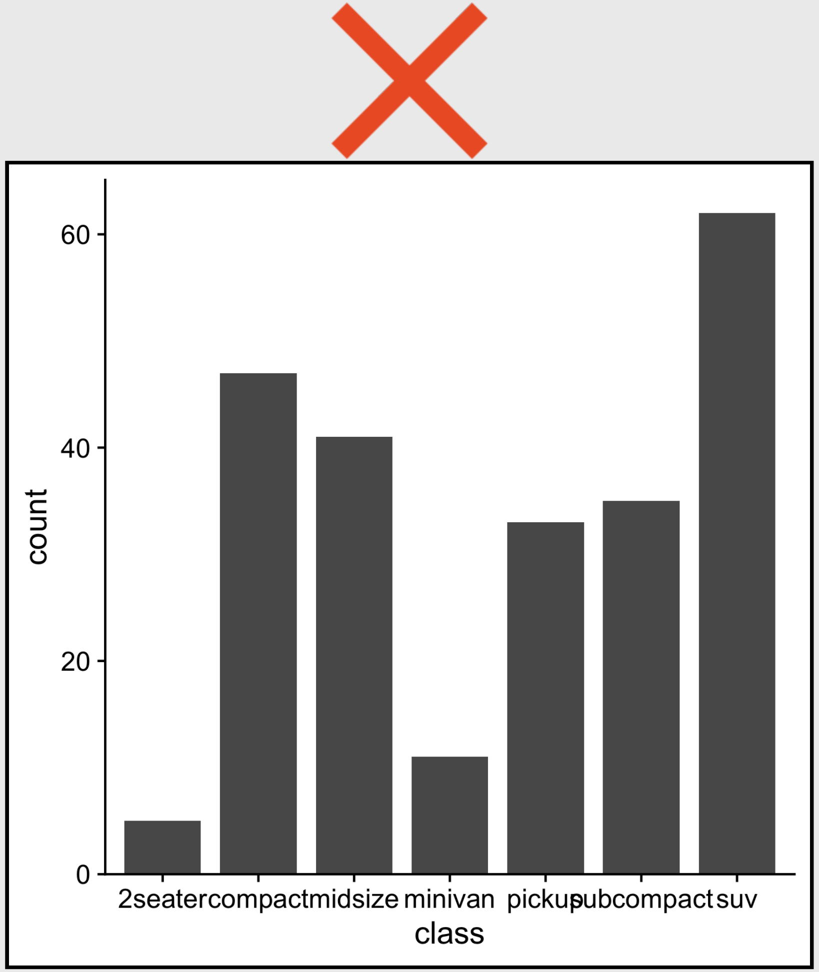


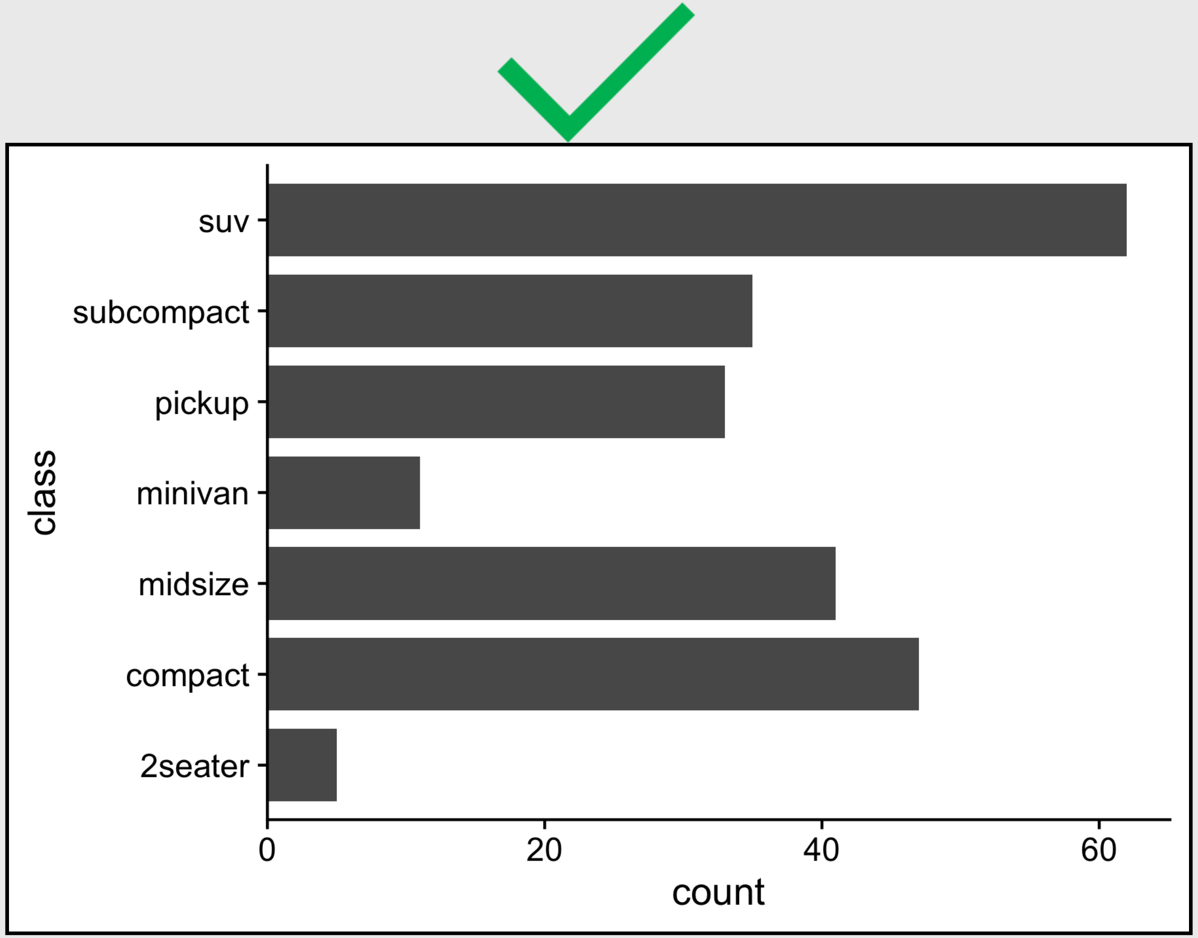
10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

*most of the time

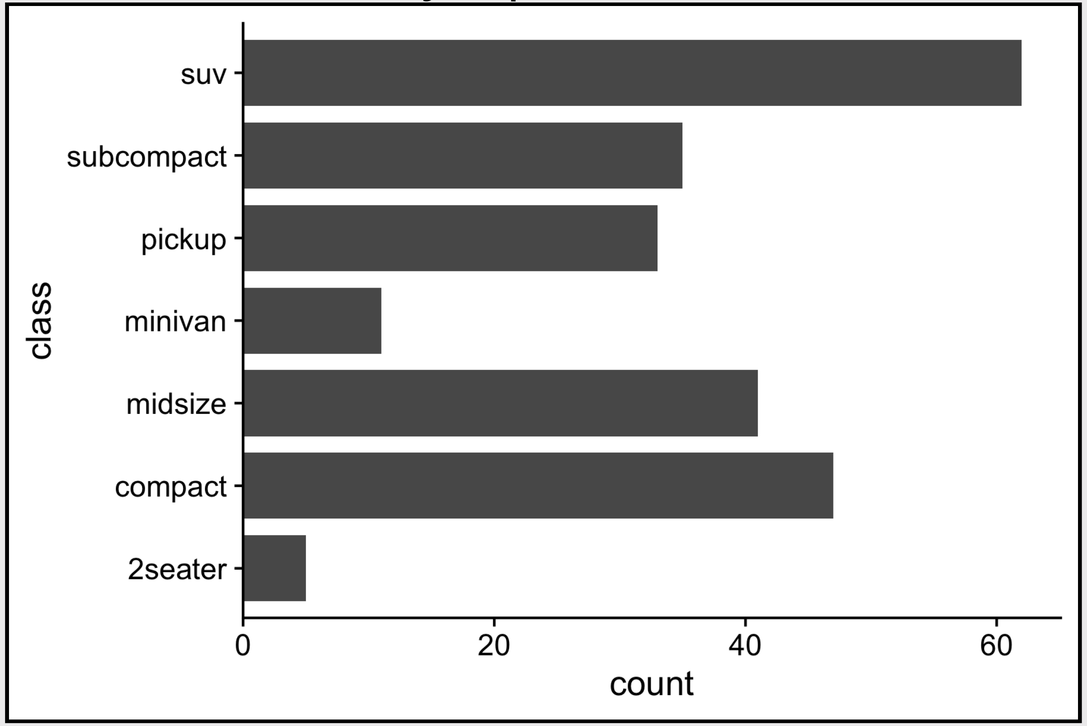
Rotate axes if you can't read them



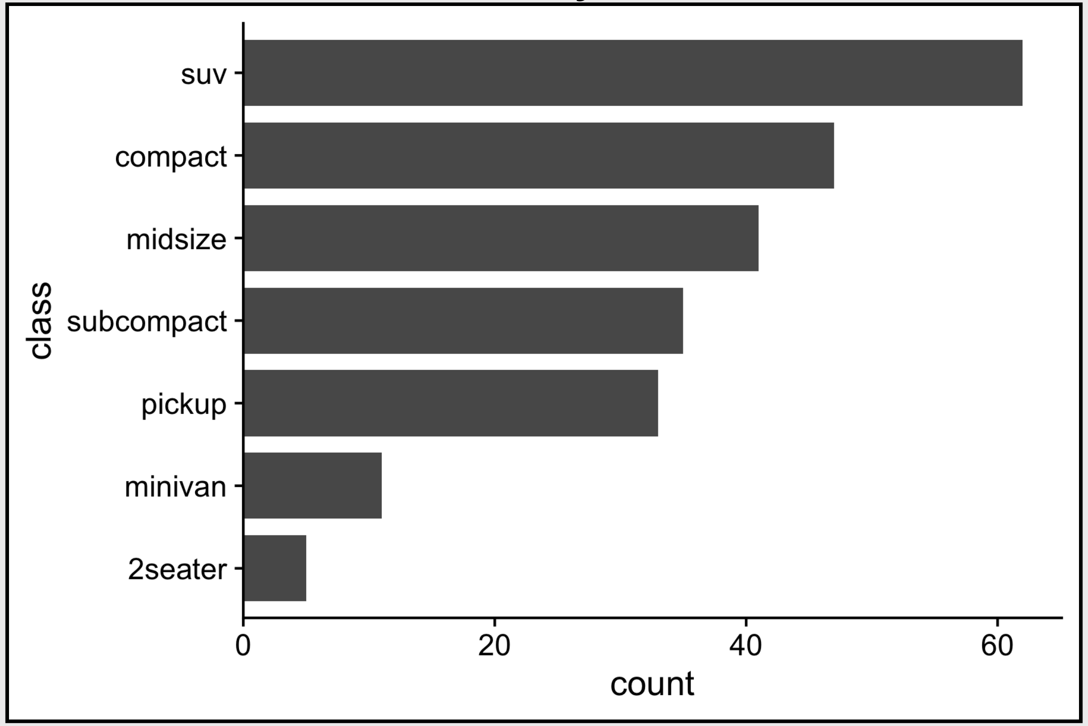


Default order is almost always wrong

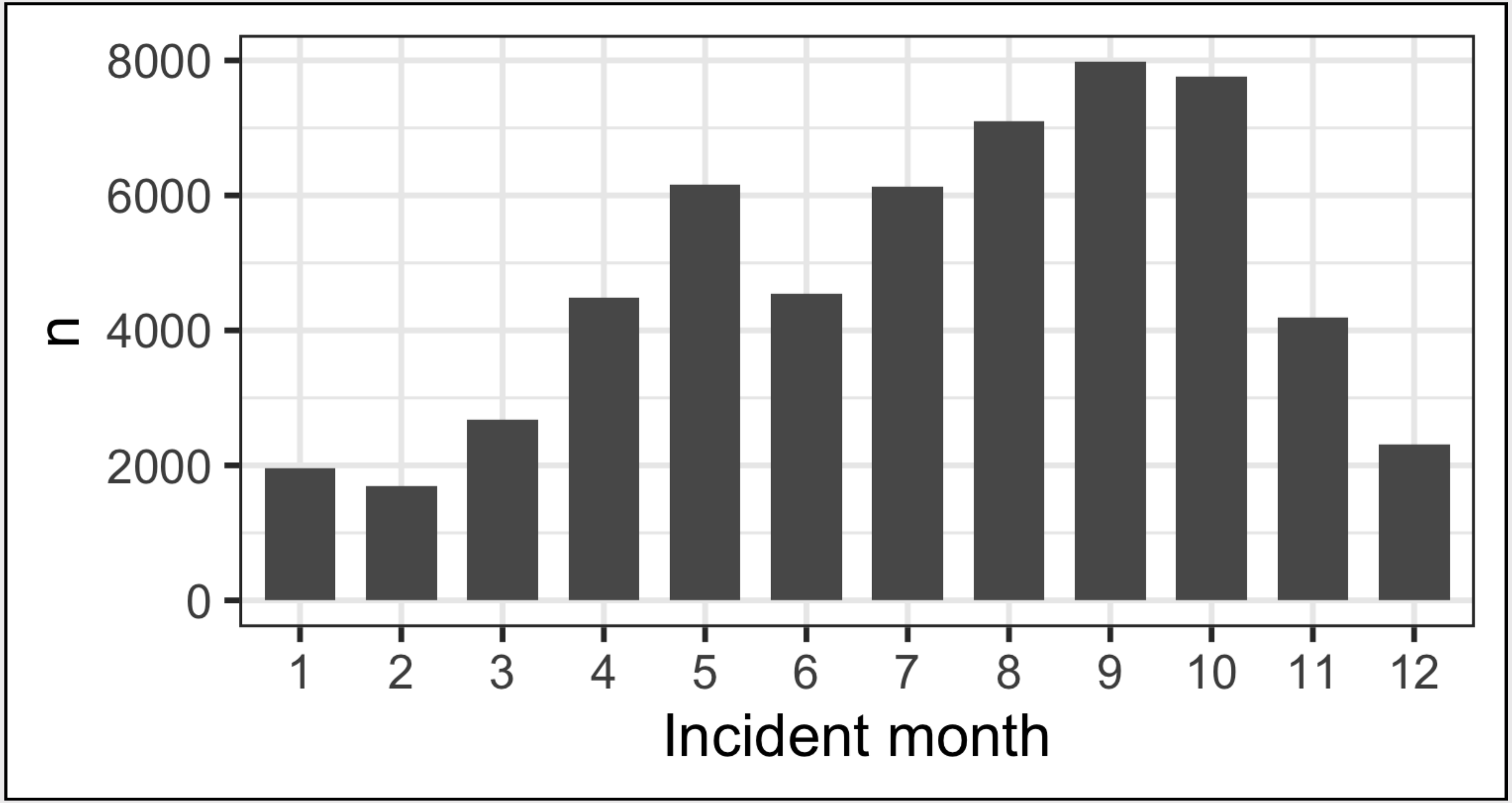
Ordered by alphabet (default)



Ordered by count



Exception: Ordinal variables

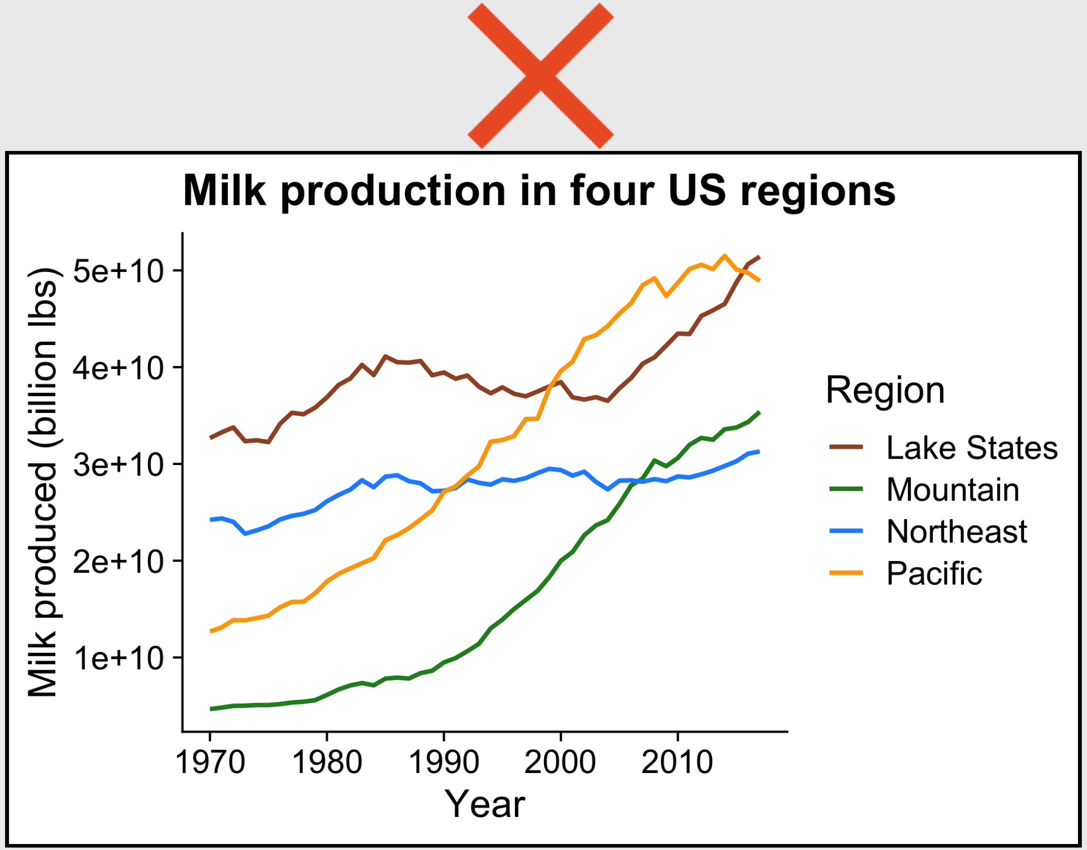


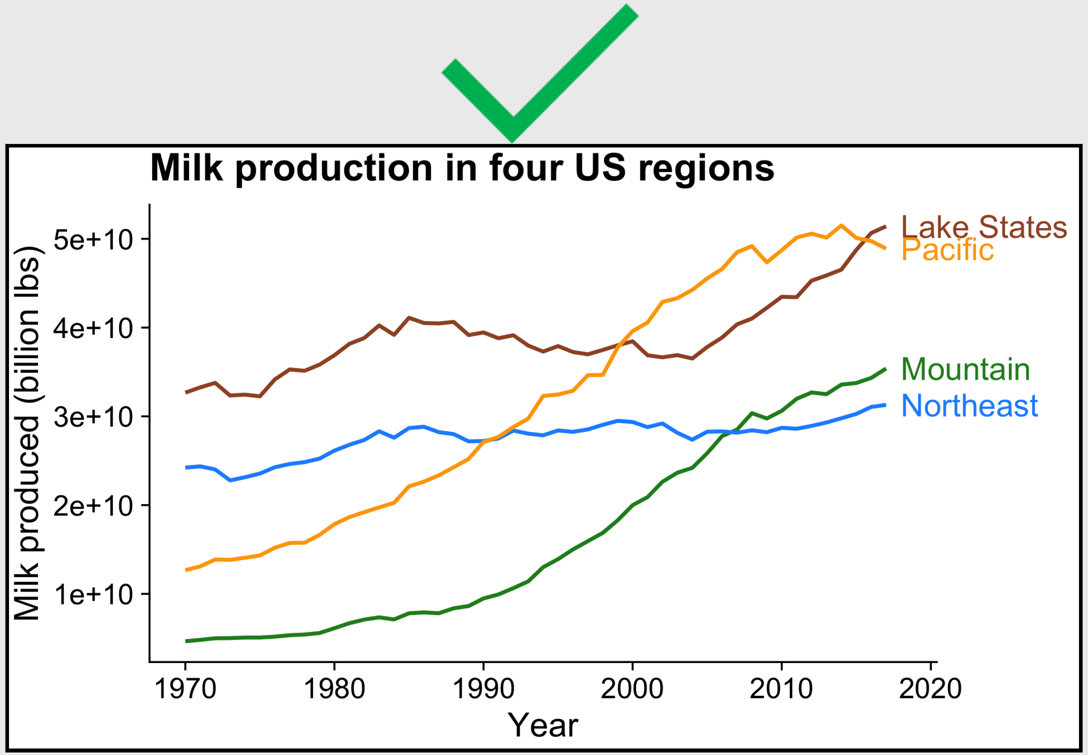
10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

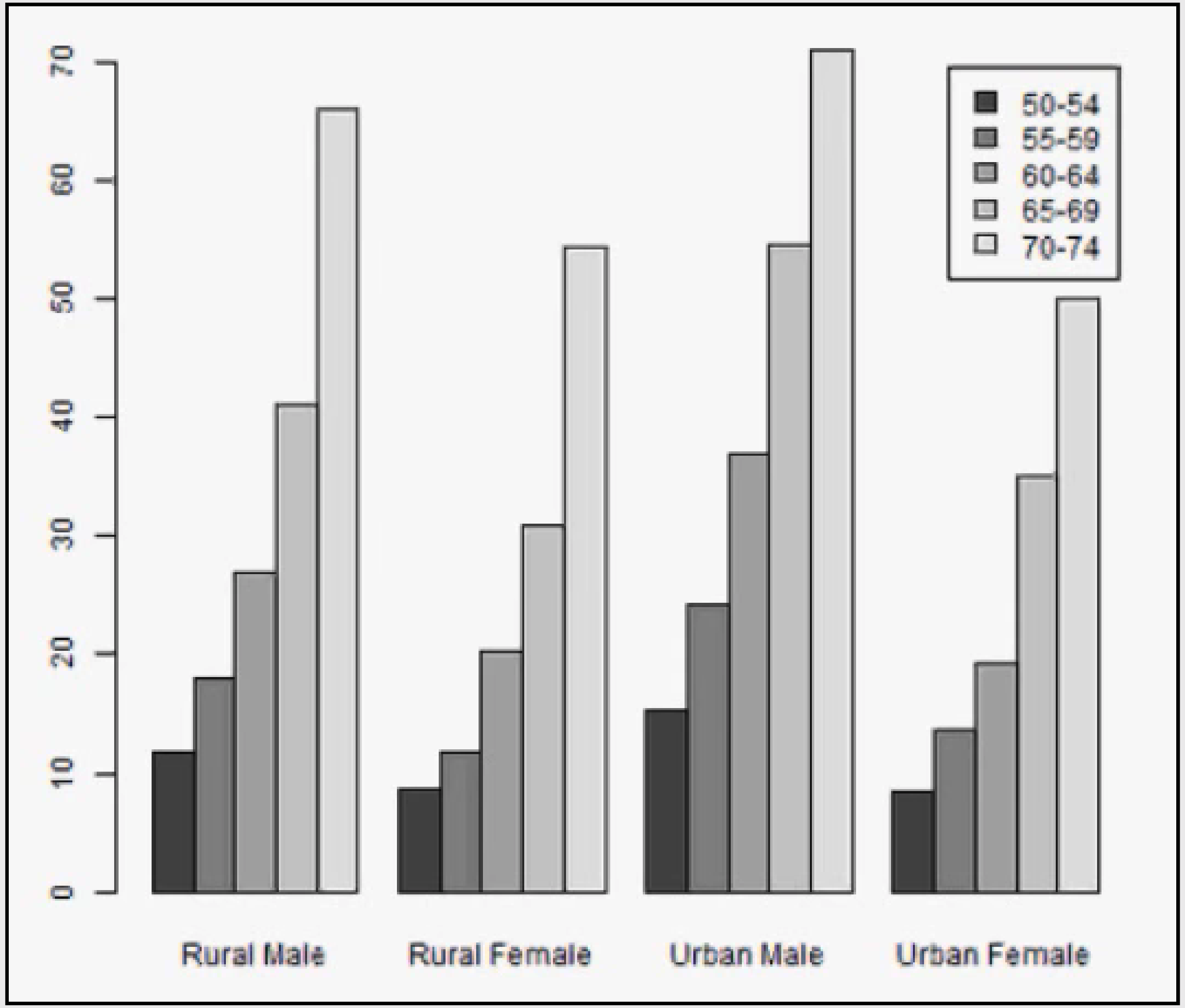
*most of the time

Directly label geoms





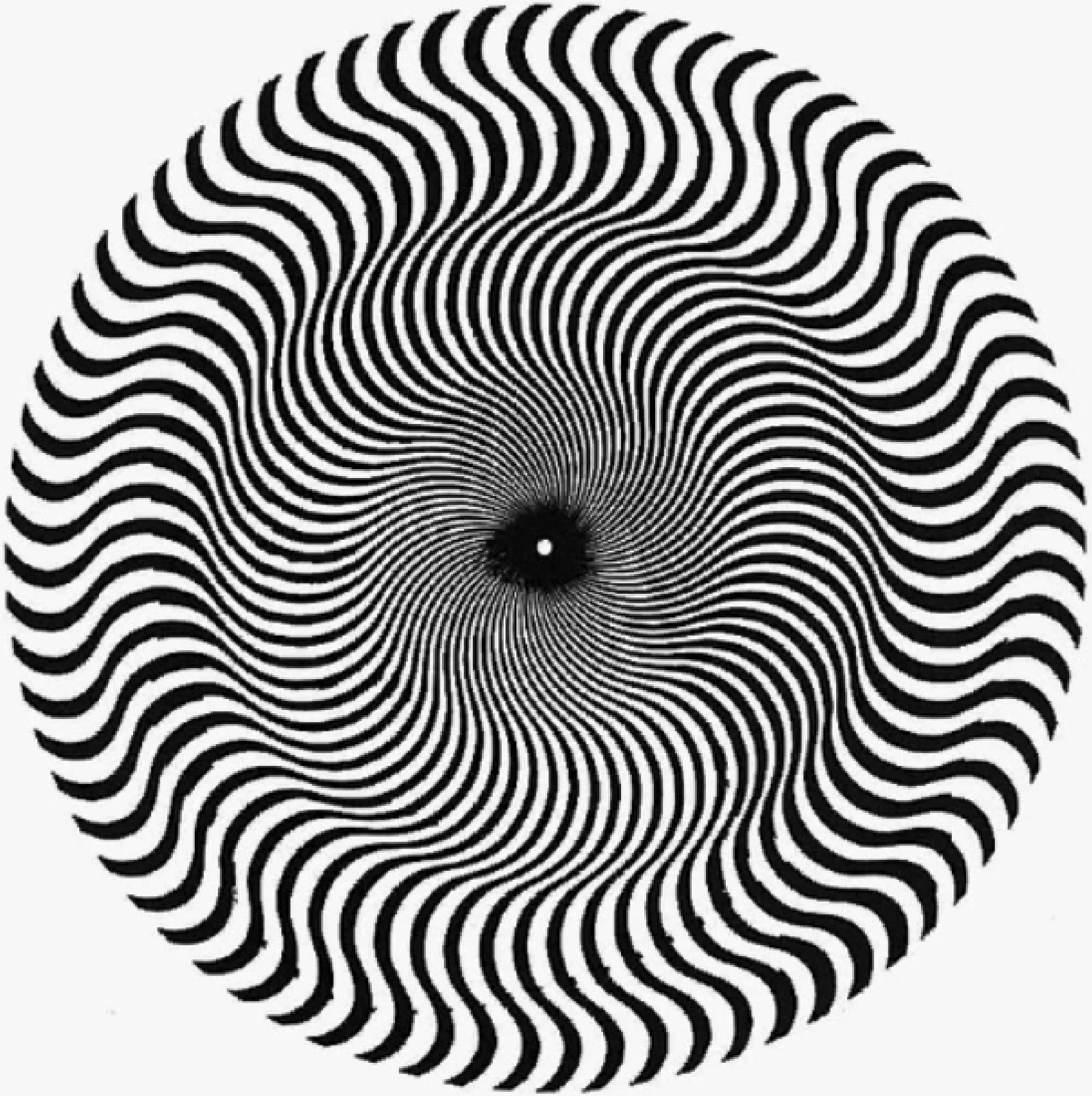
Exception: When you have repeated categories

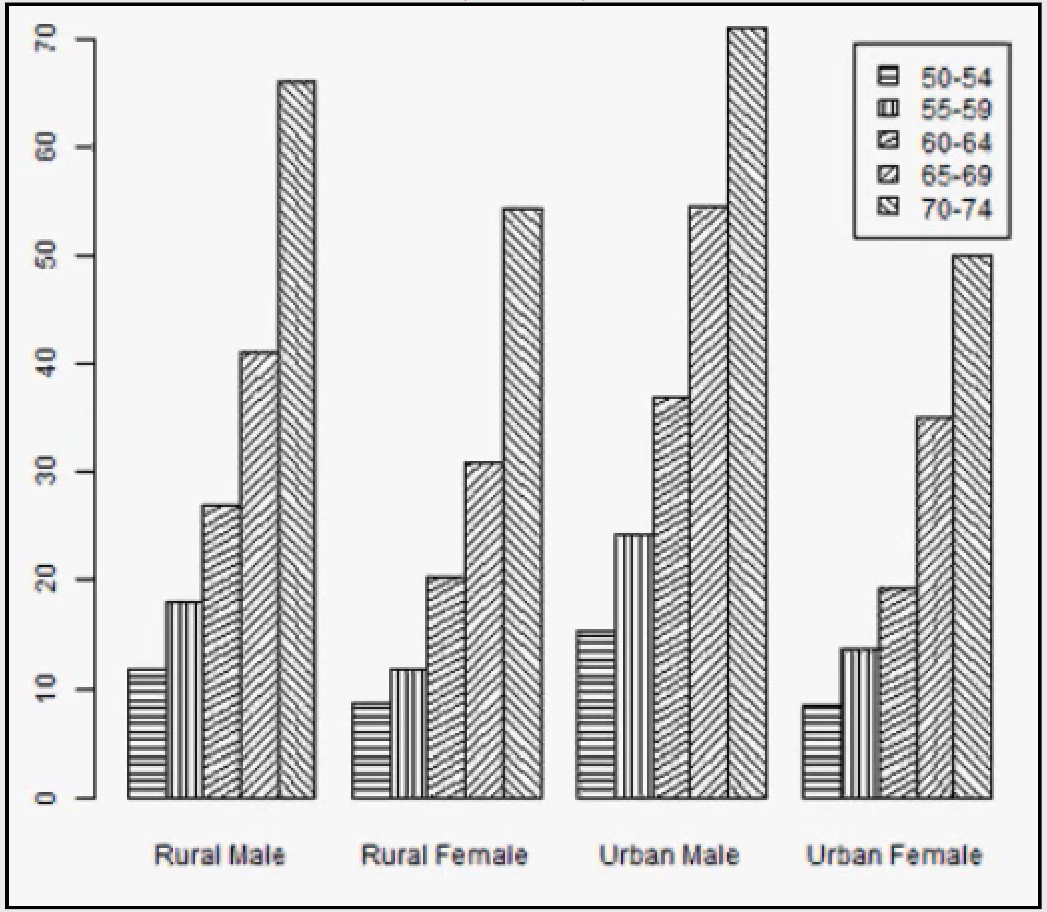


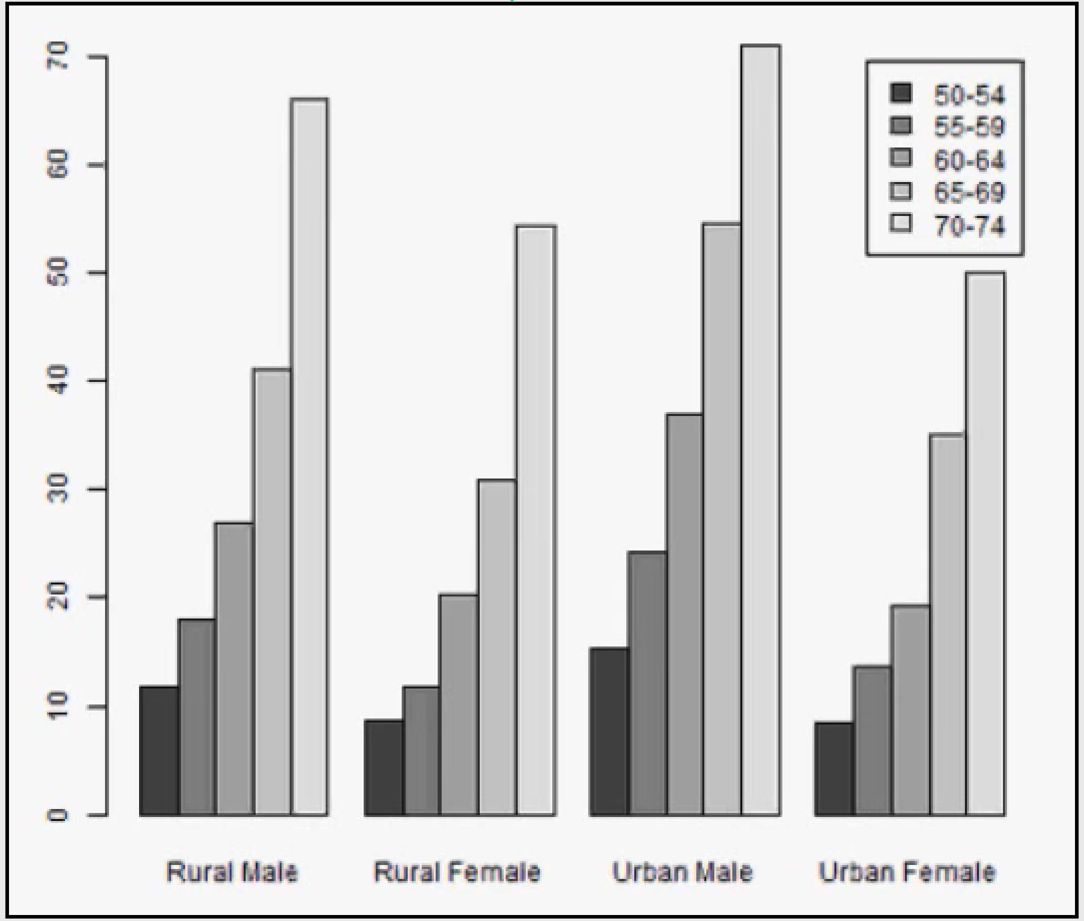
10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

*most of the time




70
60
50
40
30
20
10
0
50-54
55-59
60-64
65-69
70-74
Rural Male
Rural Female
Urban Male
Urban Female

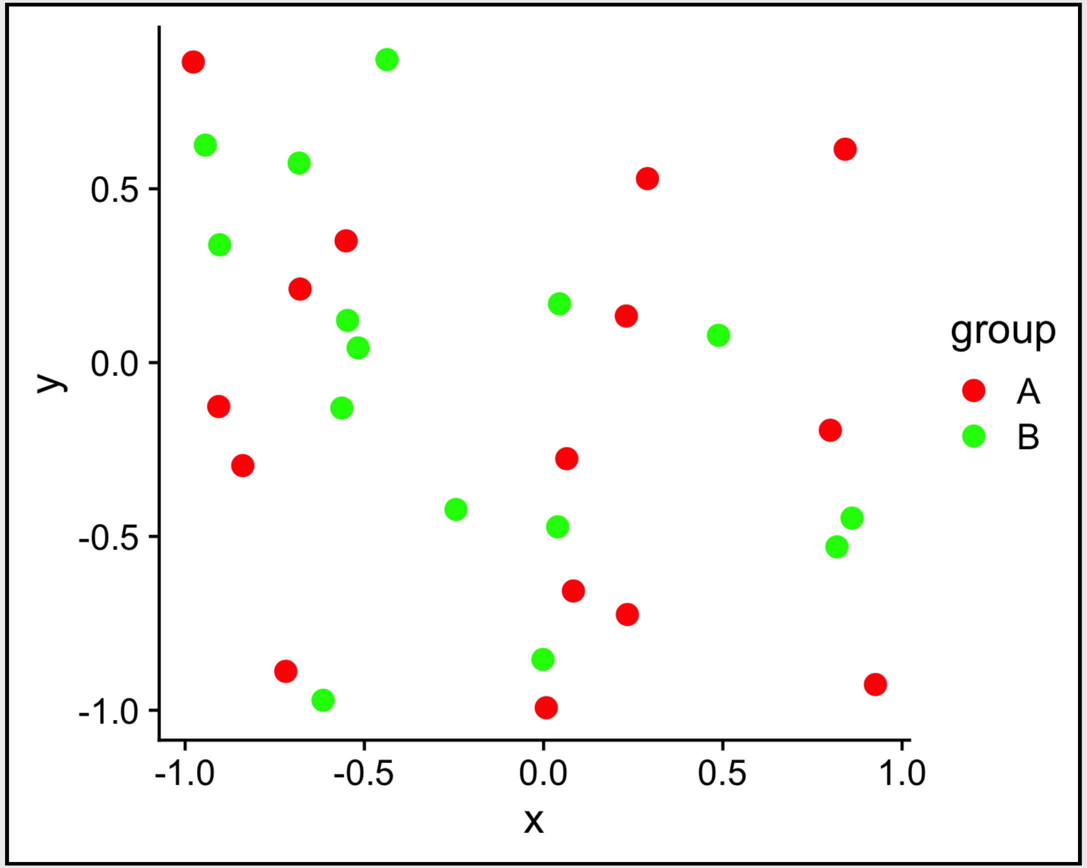

70
60
50
40
30
20
10
0
50-54
55-59
60-64
65-69
70-74
Rural Male
Rural Female
Urban Male
Urban Female

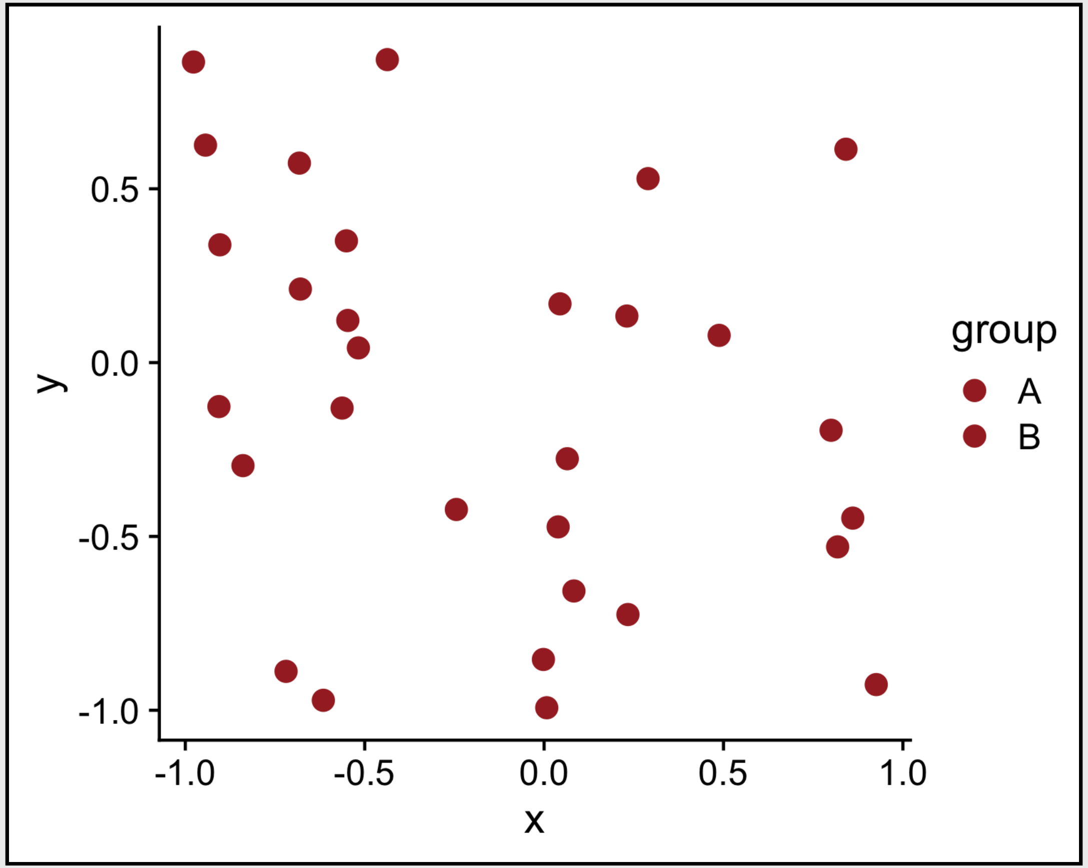
10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

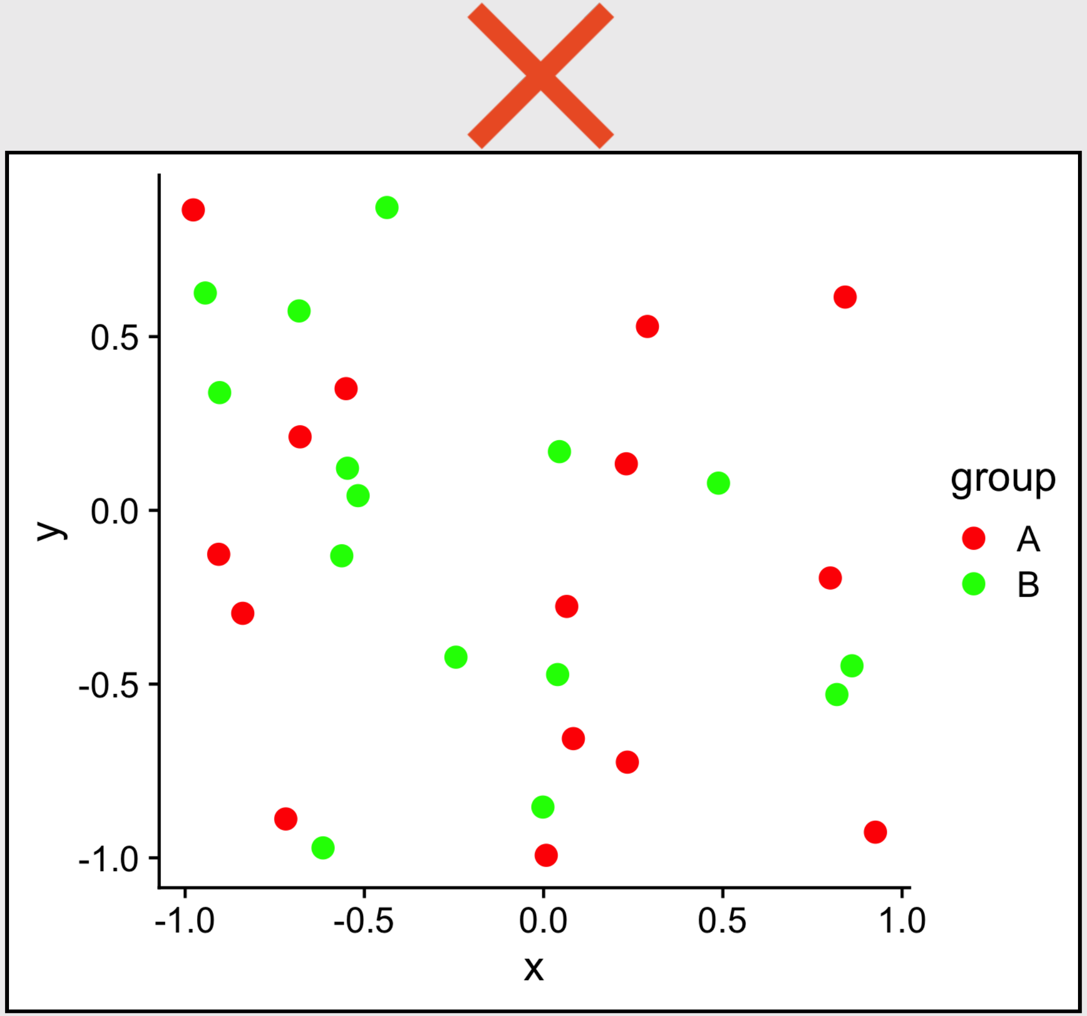
*most of the time

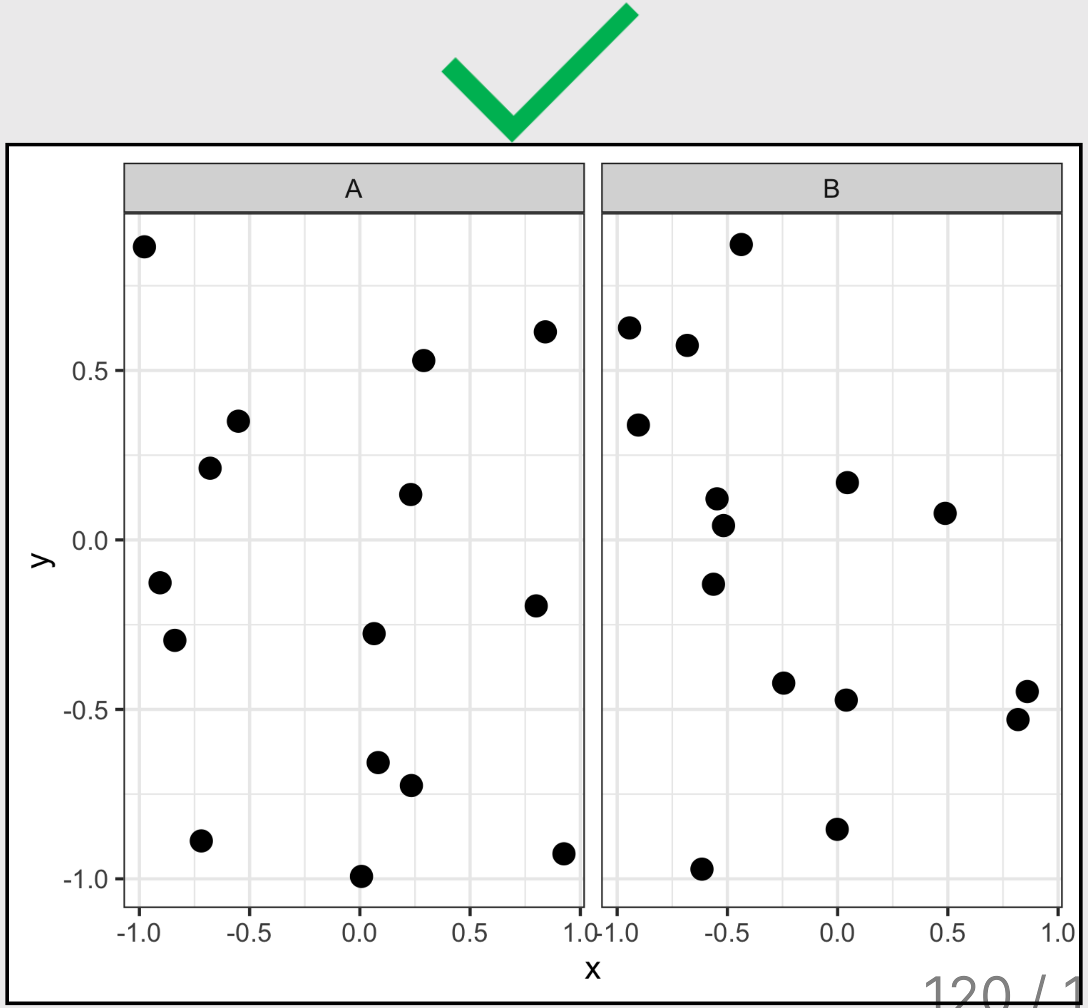
10% of males and 1% of females are color blind





Facets can be used to avoid color altogether

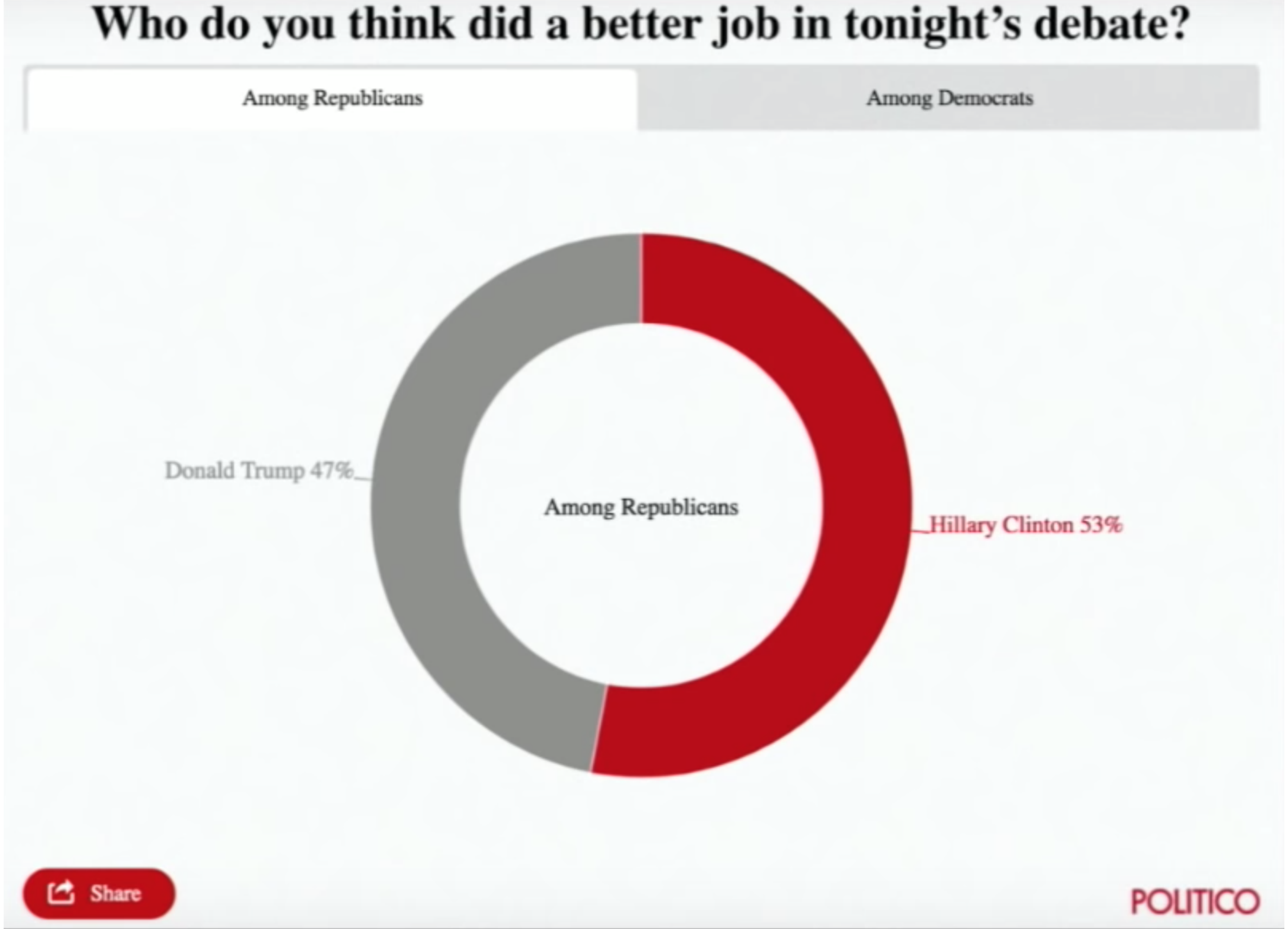




10 Data Viz Best Practices

1. Remove chart chunk
2. Don't make 3D plots*
3. Don't lie
4. Don't use pie charts for proportions*
5. Don't stack bars*
6. Rotate and sort categorical axes*
7. Eliminate legends & directly label geoms*
8. Don't use pattern fills
9. Don't use red & green together
10. Consider tables for small data sets

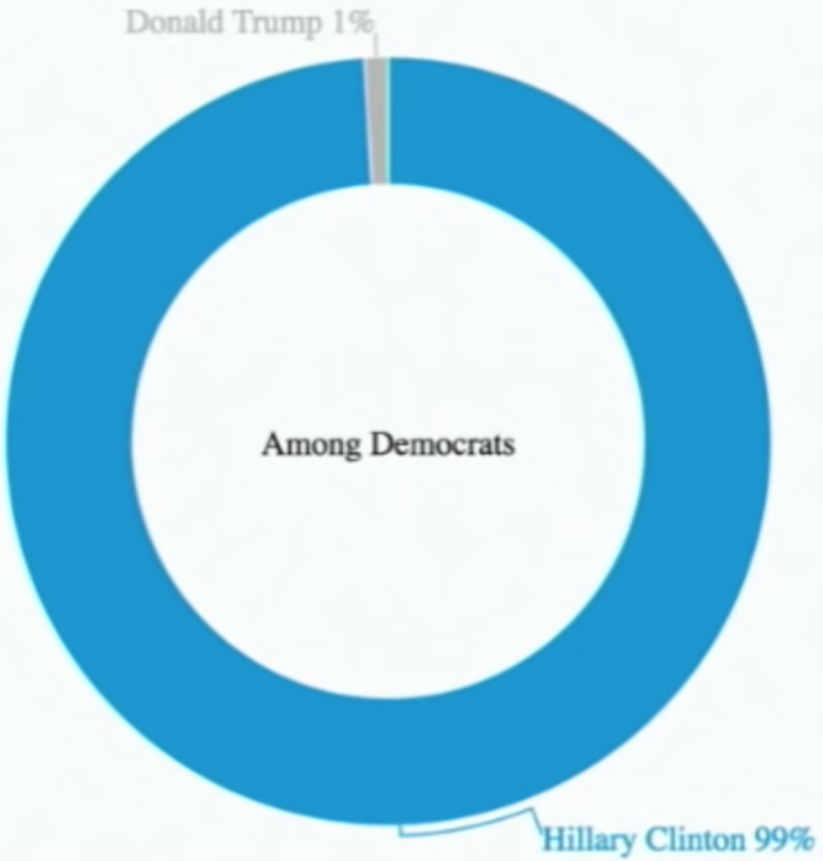
*most of the time



Who do you think did a better job in tonight's debate?
Among Republicans
Among Democrats
Donald Trump 47%
Among Republicans
Hillary Clinton 53%
Share
POLITICO

Who do you think did a better job in tonight's debate?

Among Republicans | Among Democrats



Share

POLITICO

Who do you think did a better job in tonight's debate?

	Clinton	Trump
Among Democrats	99%	1%
Among Republicans	53%	47%

References:

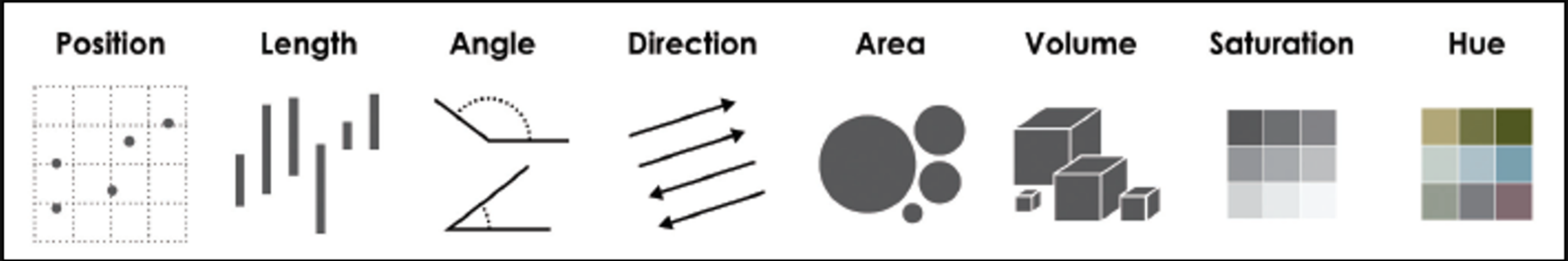
- Data Viz "Cheat Sheet"
- Data Viz Reference Page

Your turn - go here

10:00

For your "bad" visualization:

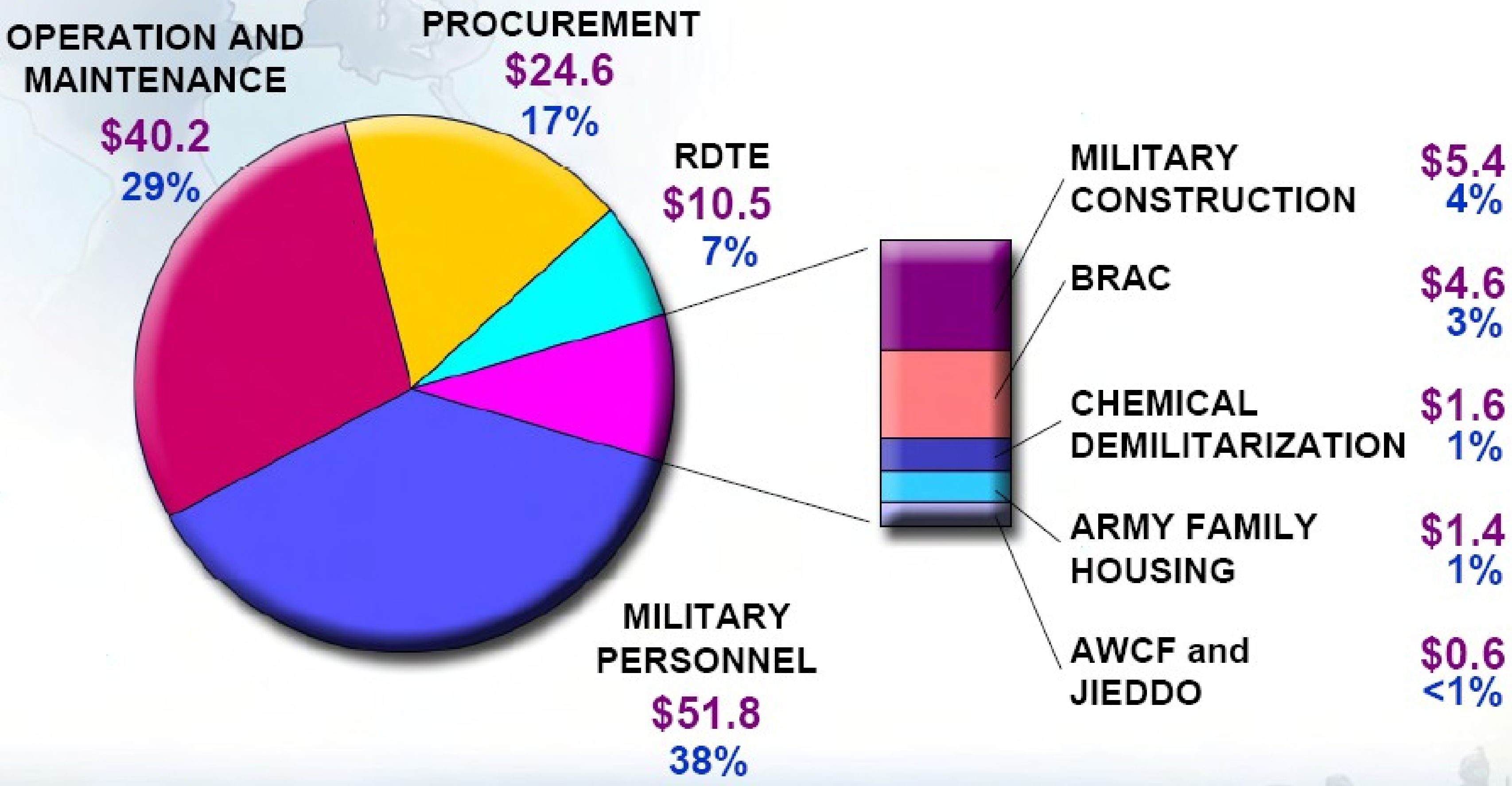
1) Identify where the graphic falls on Cleveland's pattern recognition hierarchy

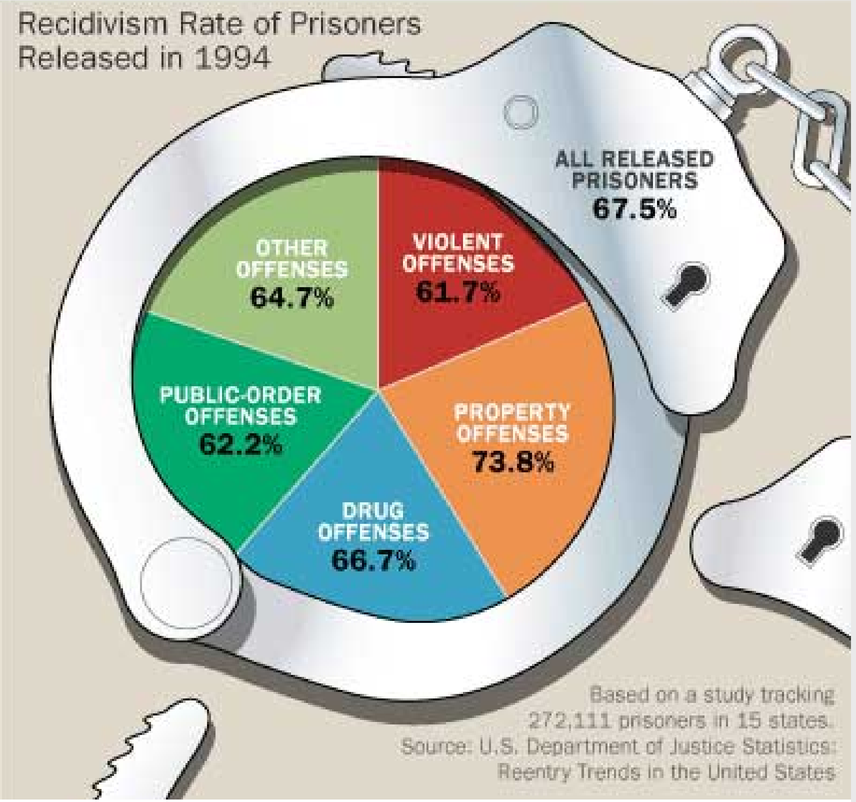


2) Any design rules that are broken

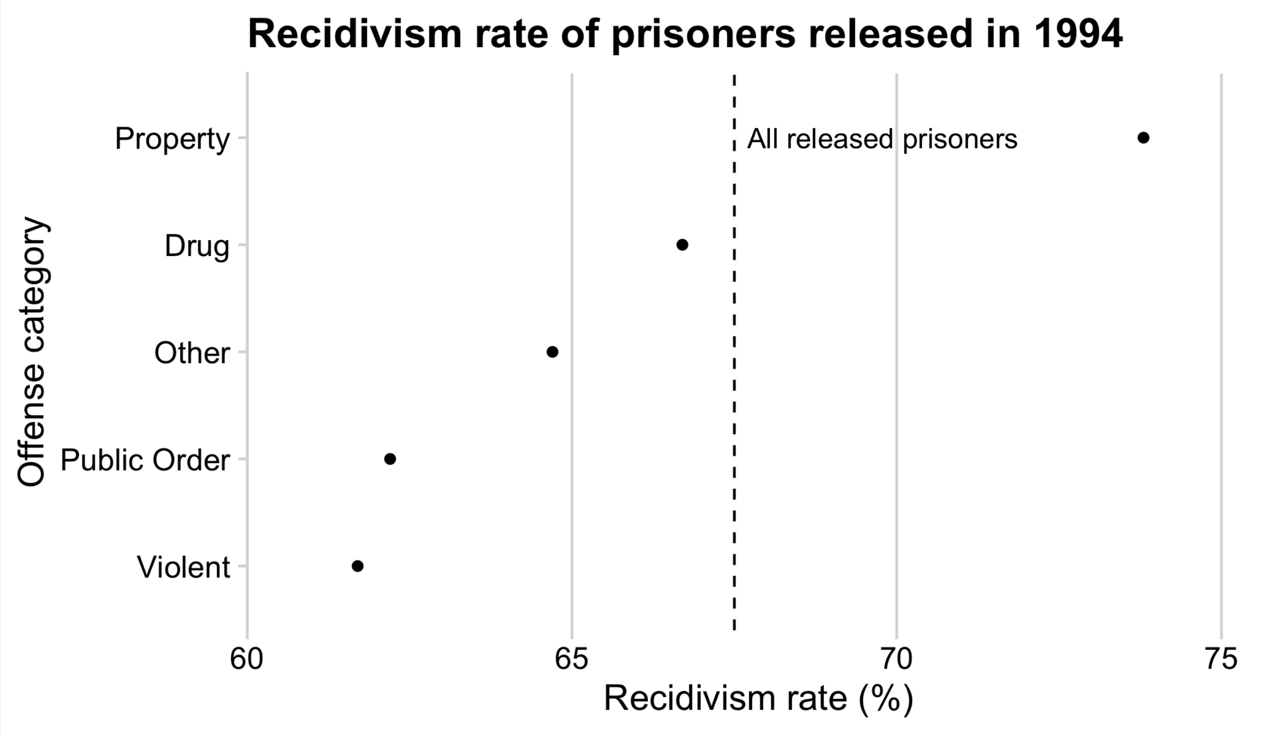
3) Suggest at least two improvements

FY09 Obligation Authority



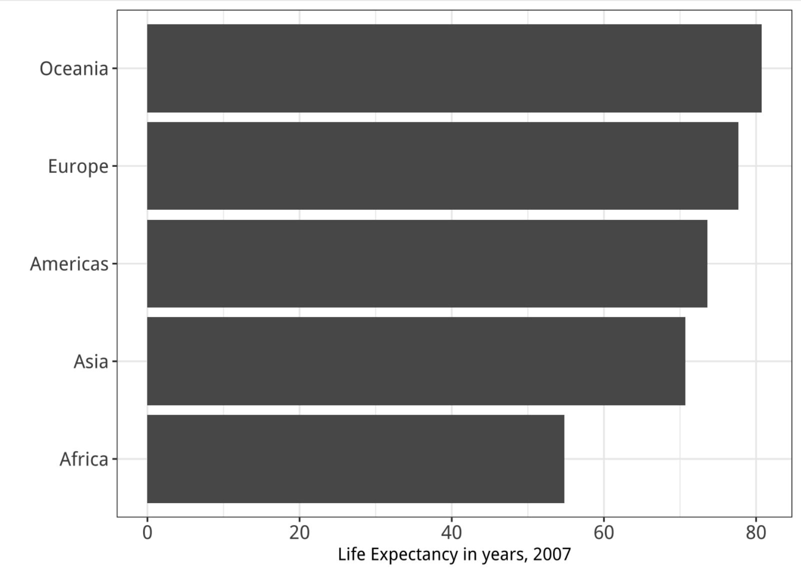


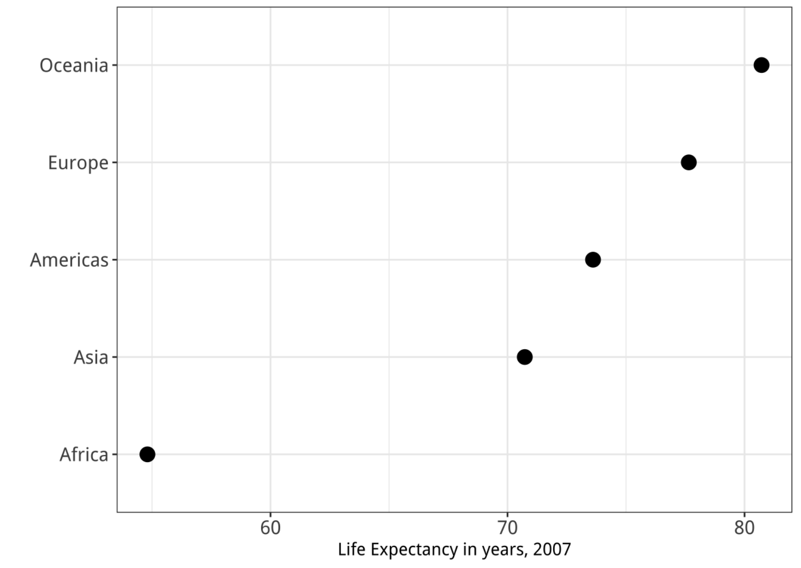
Recidivism Rate of Prisoners Released in 1994
ALL RELEASED PRISONERS 67.5%
OTHER OFFENSES 64.7%
VIOLENT OFFENSES 61.7%
PUBLIC-ORDER OFFENSES 62.2%
PROPERTY OFFENSES 73.8%
DRUG OFFENSES 66.7%
Based on a study tracking 272,111 prisoners in 15 states.
Source: U.S. Department of Justice Statistics: Reentry Trends in the United States

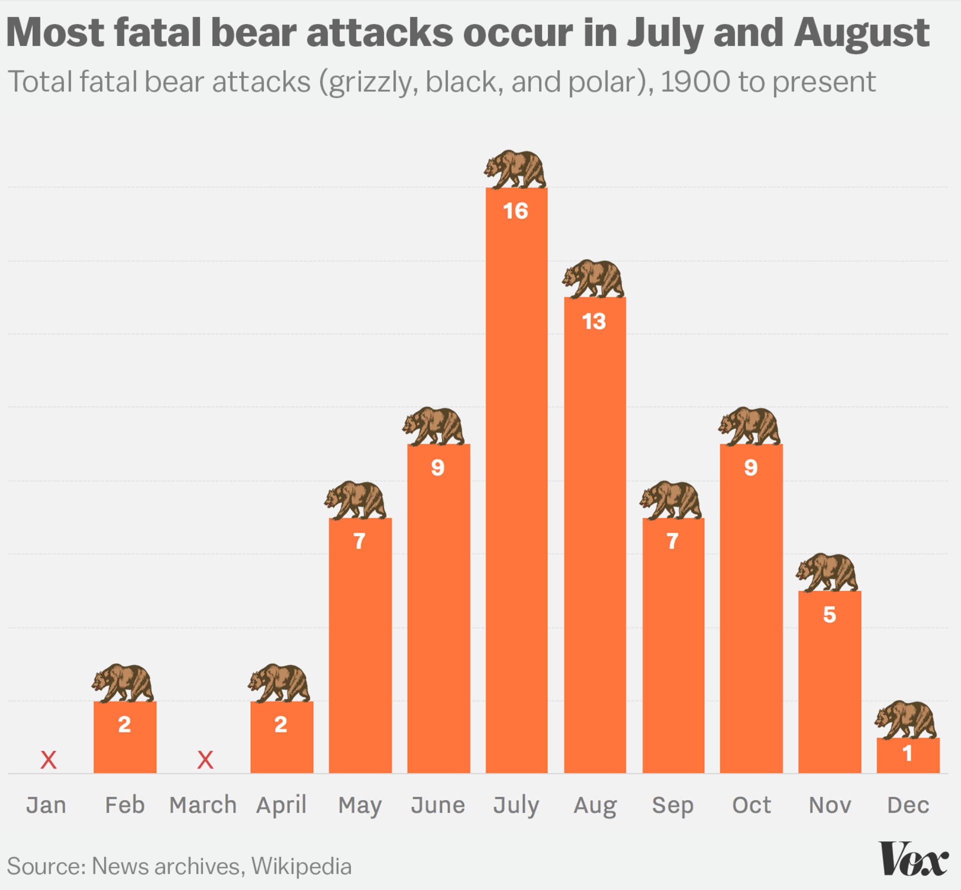


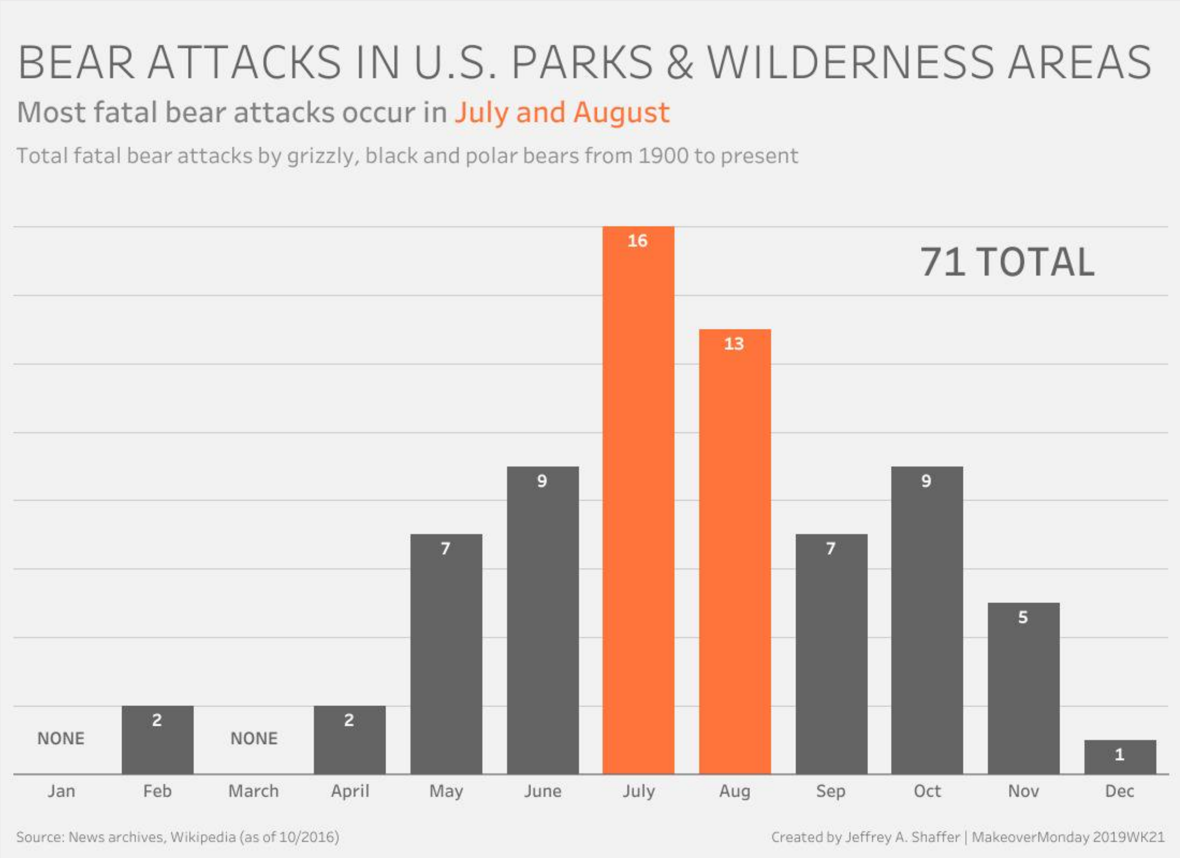
Recidivism rate of prisoners released in 1994
Property
Drug
Other
Public Order
Violent
All released prisoners
Offense category
60
65
70
75
Recidivism rate (%)

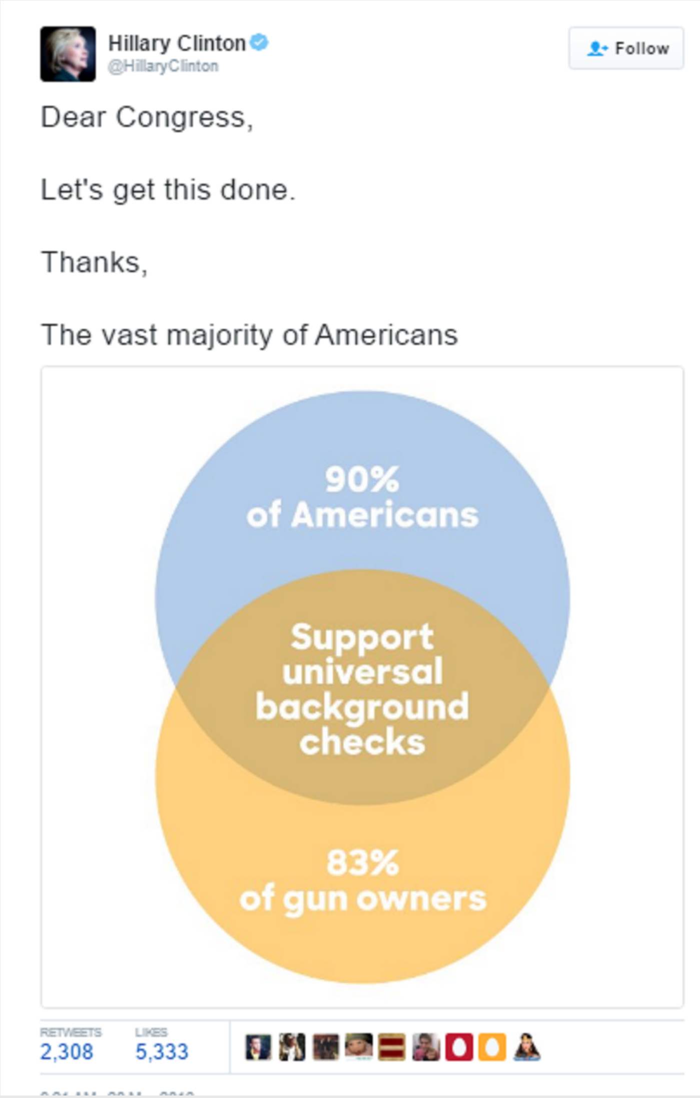

Africa
Americas
Asia
Europe
Oceania
0
22.5
45
67.5
90
Years
CONTINENT
Life Expectancy: 2007

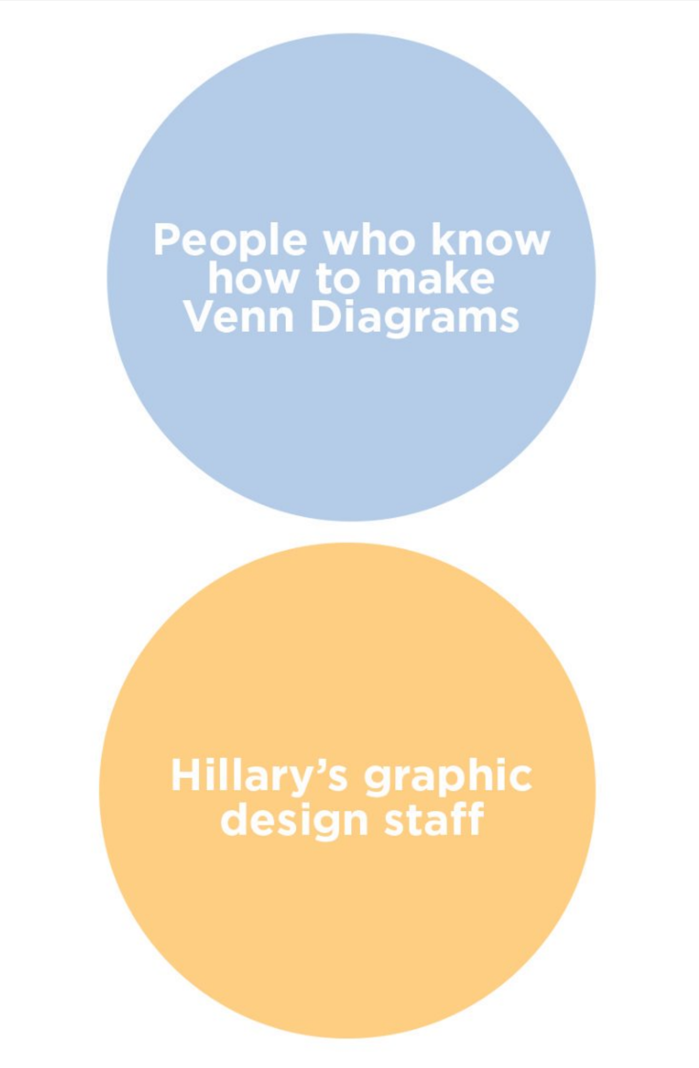

Oceania
Europe
Americas
Asia
Africa
0
20
40
60
80
Life Expectancy in years, 2007

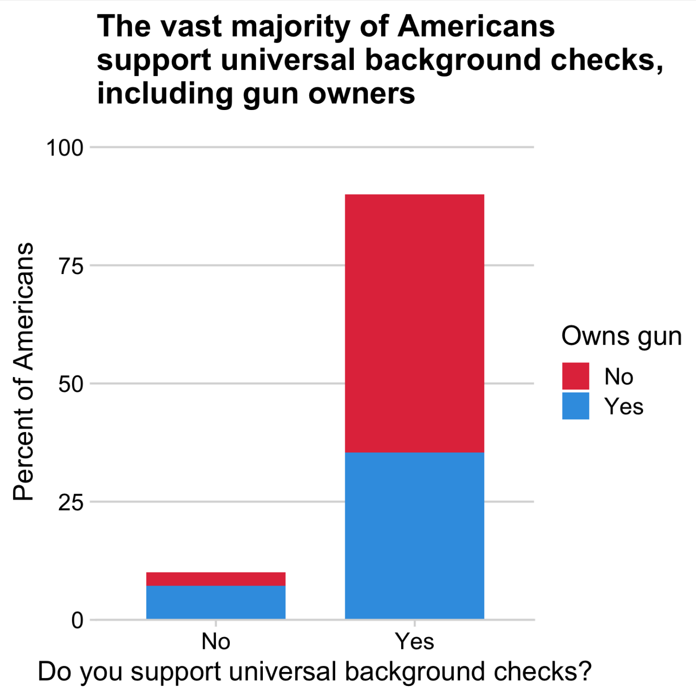

Oceania
Europe
Americas
Asia
Africa
60
70
80
Life Expectancy in years, 2007


Most fatal bear attacks occur in July and August
Total fatal bear attacks (grizzly, black, and polar), 1900 to present
16
13
9
7
9
7
5
2
2
1
X
X
Jan
Feb
March
April
May
June
July
Aug
Sep
Oct
Nov
Dec
Source: News archives, Wikipedia
Vox


BEAR ATTACKS IN U.S. PARKS & WILDERNESS AREAS
Most fatal bear attacks occur in July and August
Total fatal bear attacks by grizzly, black and polar bears from 1900 to present
71 TOTAL
16
13
9
9
7
7
5
2
2
1
NONE
NONE
Jan
Feb
March
April
May
June
July
Aug
Sep
Oct
Nov
Dec
Source: News archives, Wikipedia (as of 10/2016)
Created by Jeffrey A. Shaffer | MakeoverMonday 2019WK21


Hillary Clinton
@HillaryClinton
Follow
Dear Congress,
Let's get this done.
Thanks,
The vast majority of Americans
90%
of Americans
Support
universal
background
checks
83%
of gun owners
RETWEETS
2,308
LIKES
5,333


People who know
how to make
Venn Diagrams
Hillary's graphic
design staff


The vast majority of Americans
support universal background checks,
including gun owners
Percent of Americans
100
75
50
25
0
No
Yes
Do you support universal background checks?
Owns gun
No
Yes

Week 5: *Visualizing Information*

1. The Human Visual-Memory System

2. The Psychology of Data Viz

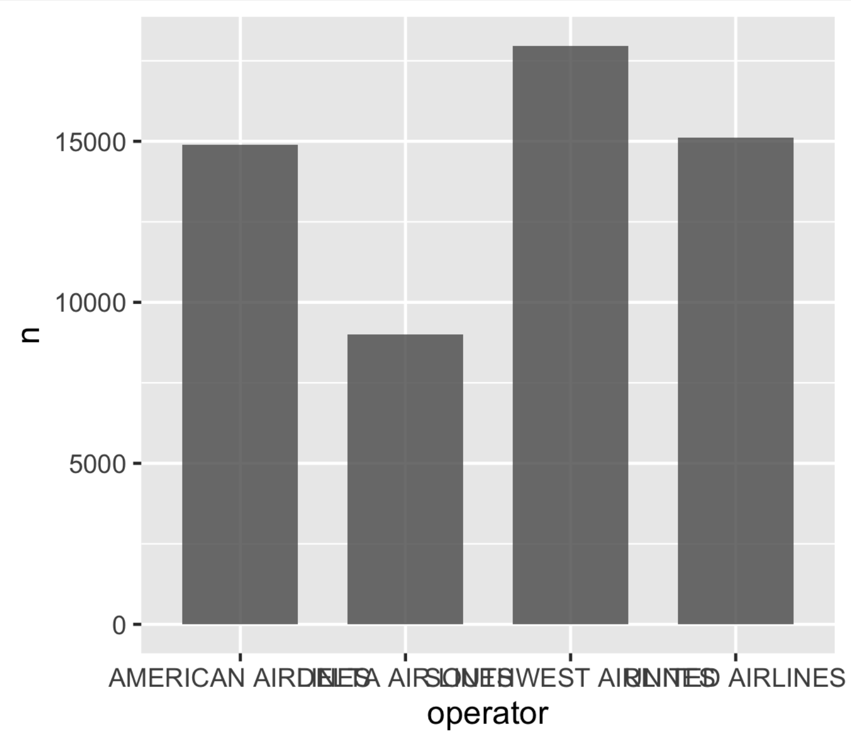
BREAK

3. 10 Data Viz Best Practices

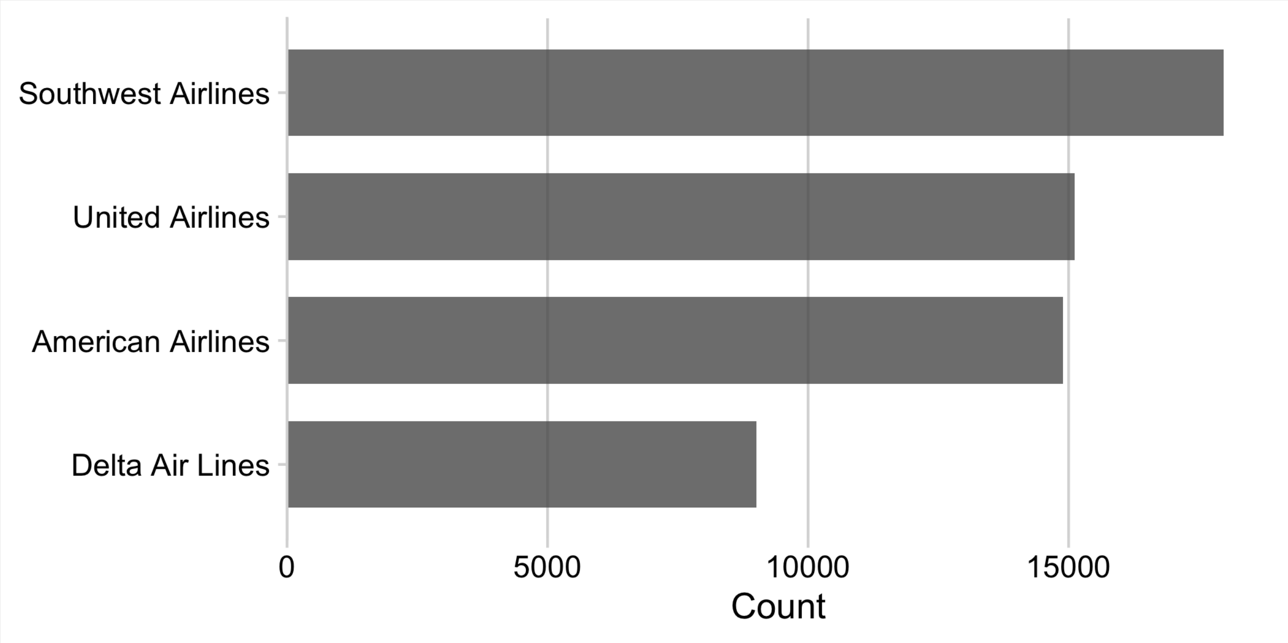
4. Making a (good) ggplot

Making a (good) ggplot

Before:



After:



Making a (good) ggplot

1. Format data frame
2. Add geoms
3. Flip coordinates?
4. Reorder factors?
5. Adjust scales
6. Adjust theme
7. Annotate

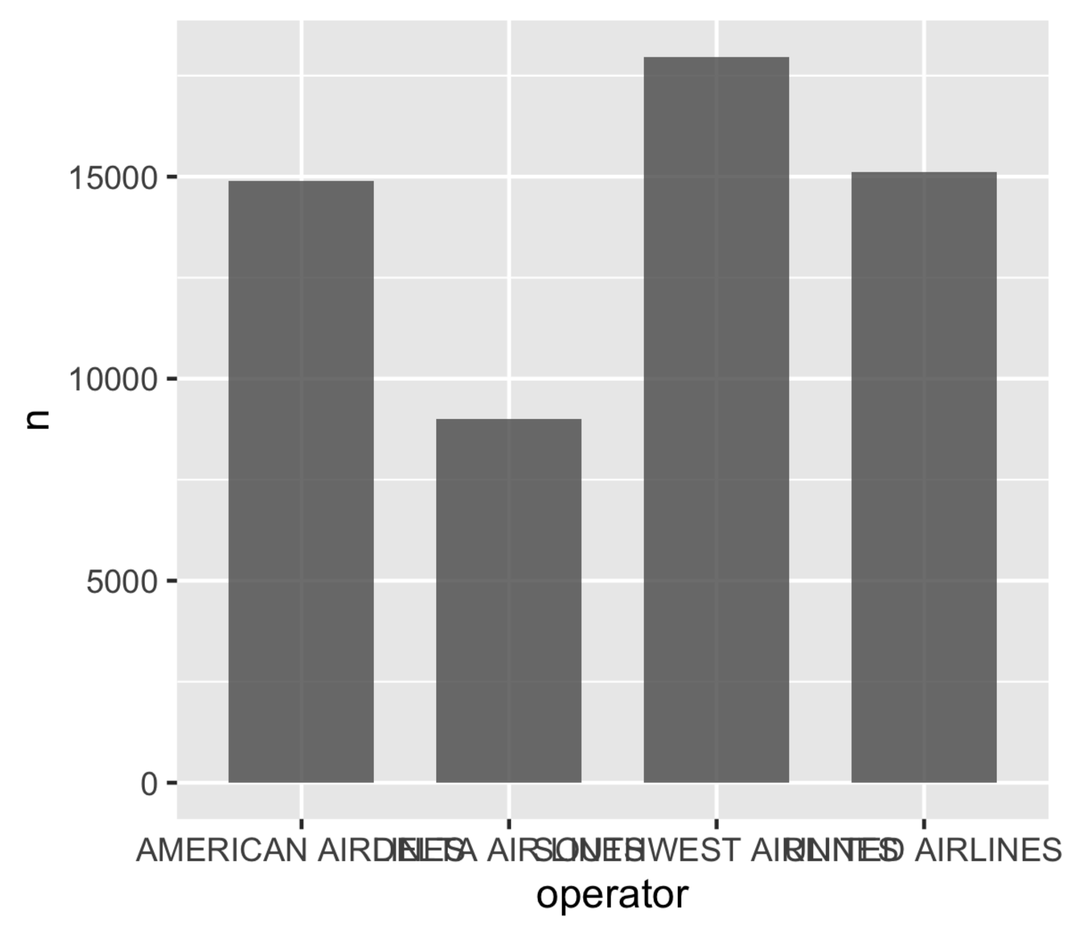
1) Format data frame

```
# Format the data frame
wildlife_impacts %>%
  count(operator)
```

```
#> # A tibble: 4 x 2
#>   operator               n
#>   <chr>              <int>
#> 1 AMERICAN AIRLINES  14887
#> 2 DELTA AIR LINES     9005
#> 3 SOUTHWEST AIRLINES 17970
#> 4 UNITED AIRLINES    15116
```

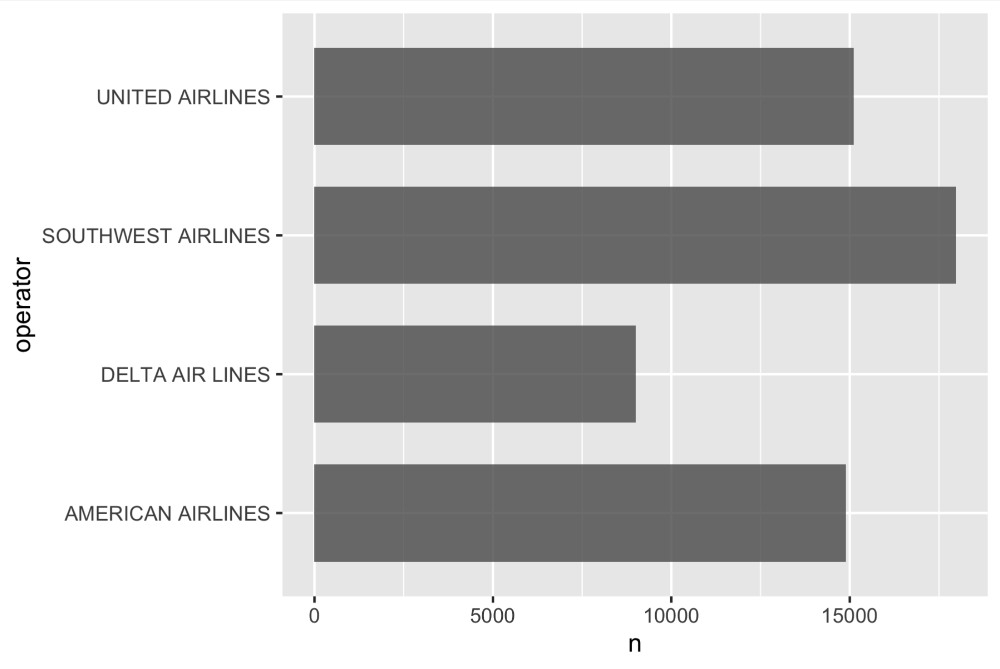
2) Add geoms

```r
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = operator, y = n),
           width = 0.7, alpha = 0.8)
```



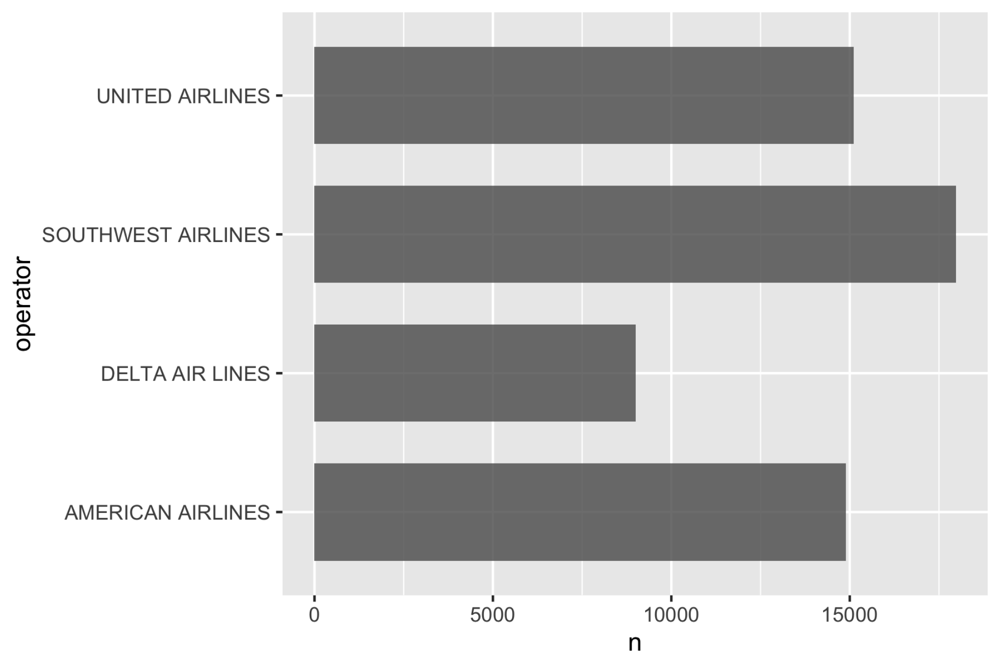
3) Flip coordinates - can you read the labels?

```
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = operator, y = n),
           width = 0.7, alpha = 0.8) +

# Flip coordinates
  coord_flip()
```



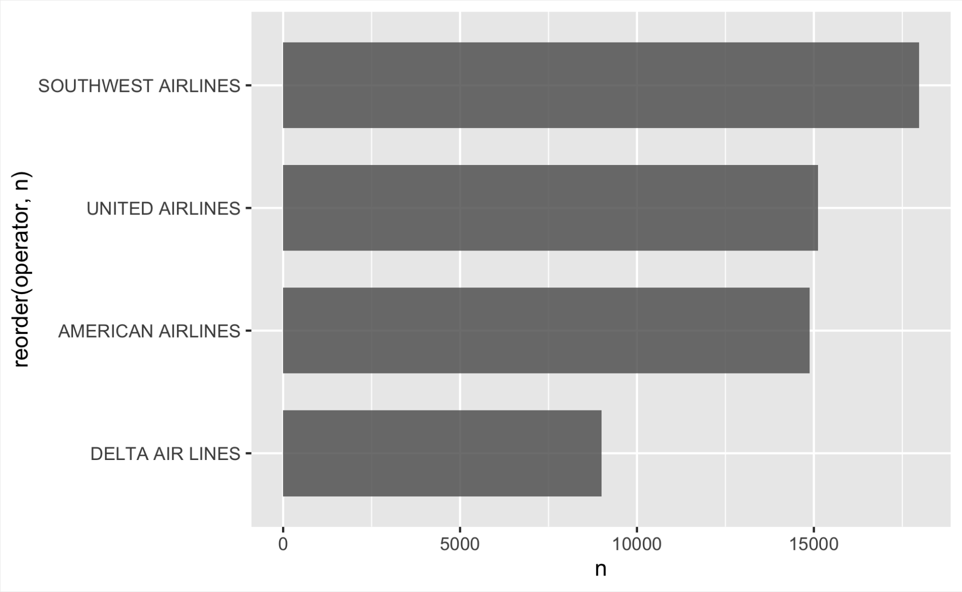
3) Flip coordinates - can you read the labels?

```
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = n, y = operator),
           width = 0.7, alpha = 0.8)
```



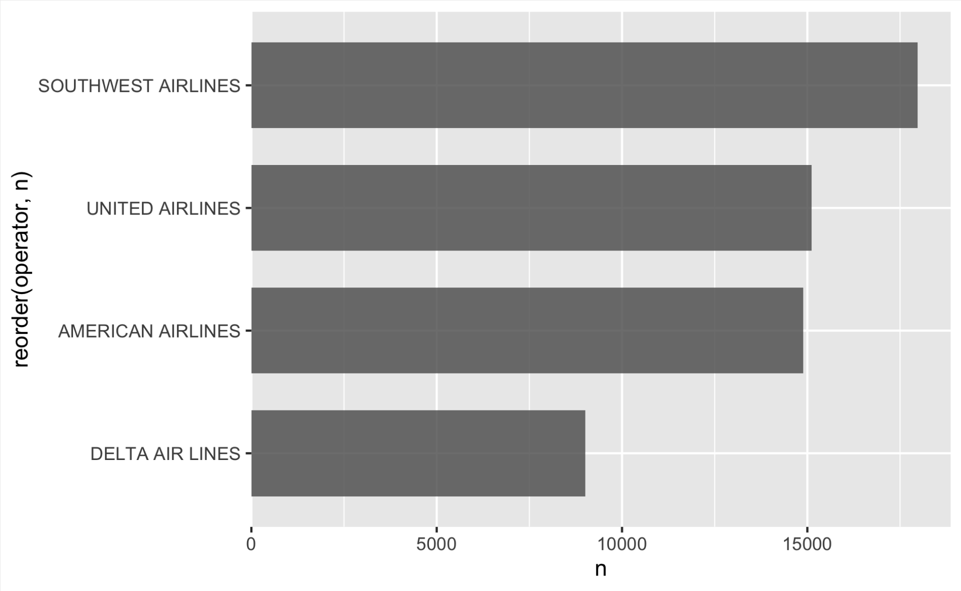
4) Reorder factors with `reorder()`

```
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = n, y = reorder(operator, n)),
           width = 0.7, alpha = 0.8)
```



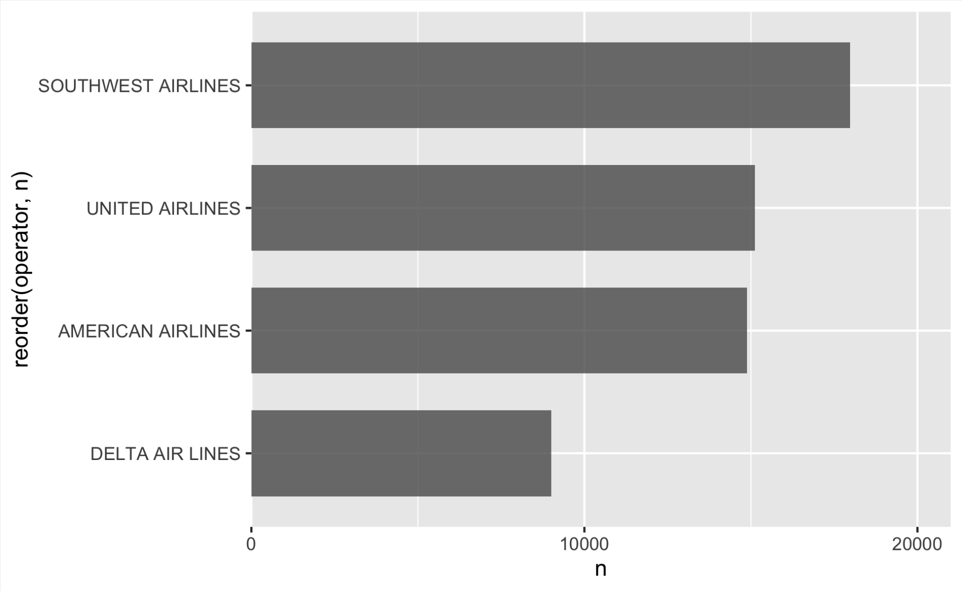
5) Adjust scales

```r
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = n, y = reorder(operator, n)),
           width = 0.7, alpha = 0.8) +

# Adjust x axis scale
  scale_x_continuous(
    expand = expansion(mult = c(0, 0.05)))
```



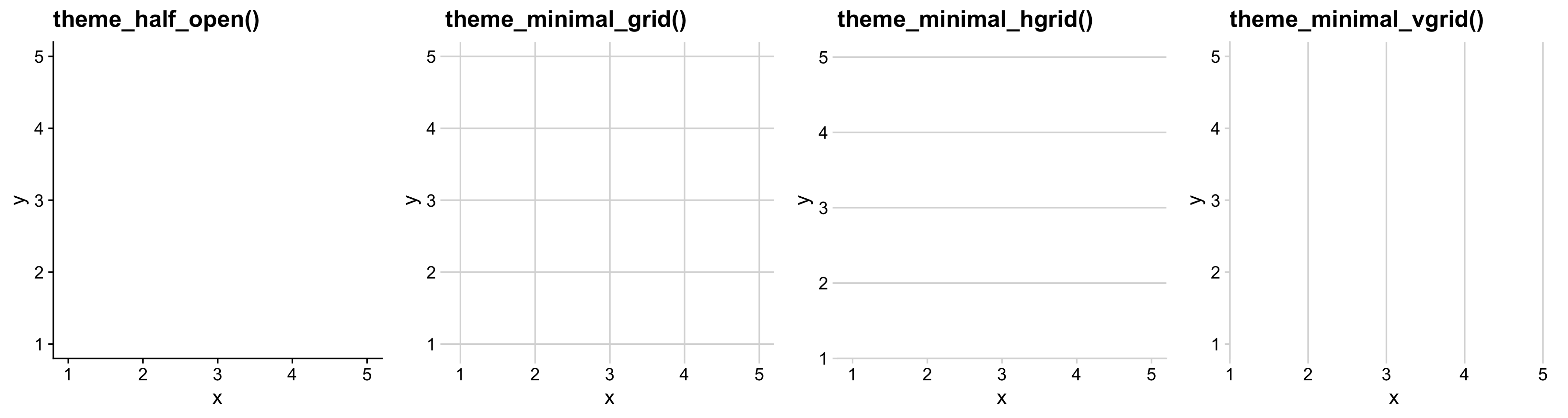
5) Adjust scales - customize break points (if you want)

```r
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = n, y = reorder(operator, n)),
           width = 0.7, alpha = 0.8) +

# Adjust x axis scale
  scale_x_continuous(
    expand = expansion(mult = c(0, 0.05)),
    breaks = c(0, 10000, 20000),
    limits = c(0, 20000))
```



6) Adjust theme

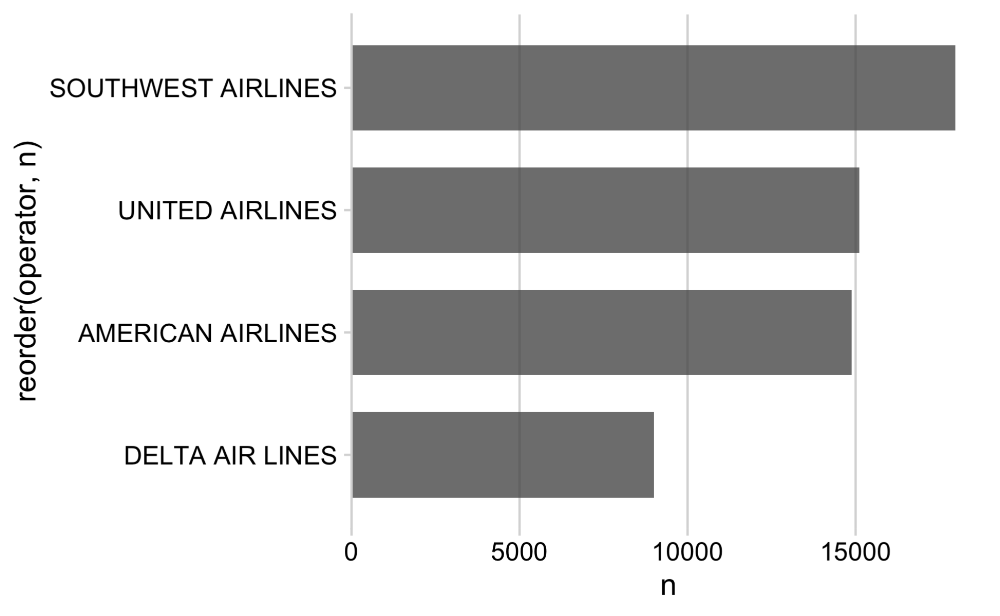
Four `cowplot` themes you should know



6) Adjust theme

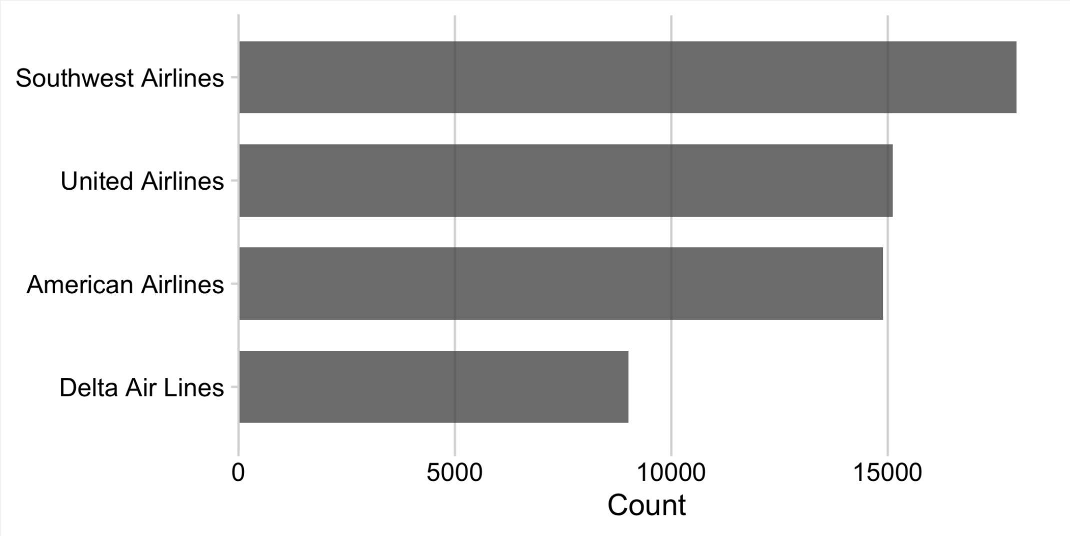
For horizontal bars, add only vertical grid

```
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = n, y = reorder(operator, n)),
           width = 0.7, alpha = 0.8) +

# Adjust x axis scale
  scale_x_continuous(
    expand = expansion(mult = c(0, 0.05))) +

# Adjust theme
  theme_minimal_vgrid()
```



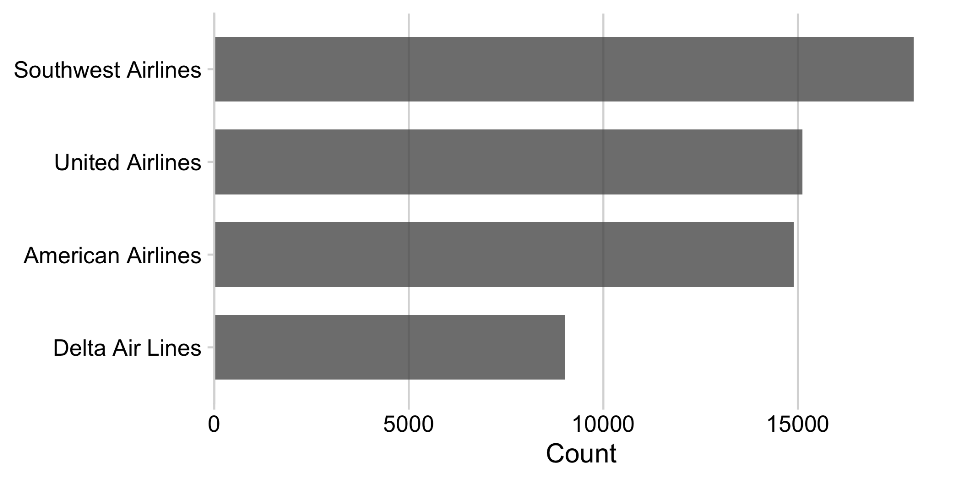
7) Annotate

```r
# Format the data frame
wildlife_impacts %>%
  count(operator) %>%
  mutate(operator = str_to_title(operator)) %>%

# Add geoms
  ggplot() +
  geom_col(aes(x = n, y = reorder(operator, n)),
           width = 0.7, alpha = 0.8) +

# Adjust x axis scale
  scale_x_continuous(
    expand = expansion(mult = c(0, 0.05))) +

# Adjust theme
  theme_minimal_vgrid() +

# Annotate
  labs(
    x = 'Count',
    y = NULL)
```



Finished product

```
wildlife_impacts %>%
  count(operator) %>%
  mutate(operator = str_to_title(operator)) %>%
  ggplot() +
  geom_col(aes(x = n, y = reorder(operator, n)),
           width = 0.7, alpha = 0.8) +
  scale_x_continuous(
    expand = expansion(mult = c(0, 0.05))) +
  theme_minimal_vgrid() +
  labs(
    x = 'Count',
    y = NULL)
```



Your turn

15:00

Use the `gapminder.csv` data to create the following plot, following these steps:

1. Format data frame
2. Add geoms
3. Flip coordinates?
4. Reorder factors?
5. Adjust scales
6. Adjust theme
7. Annotate