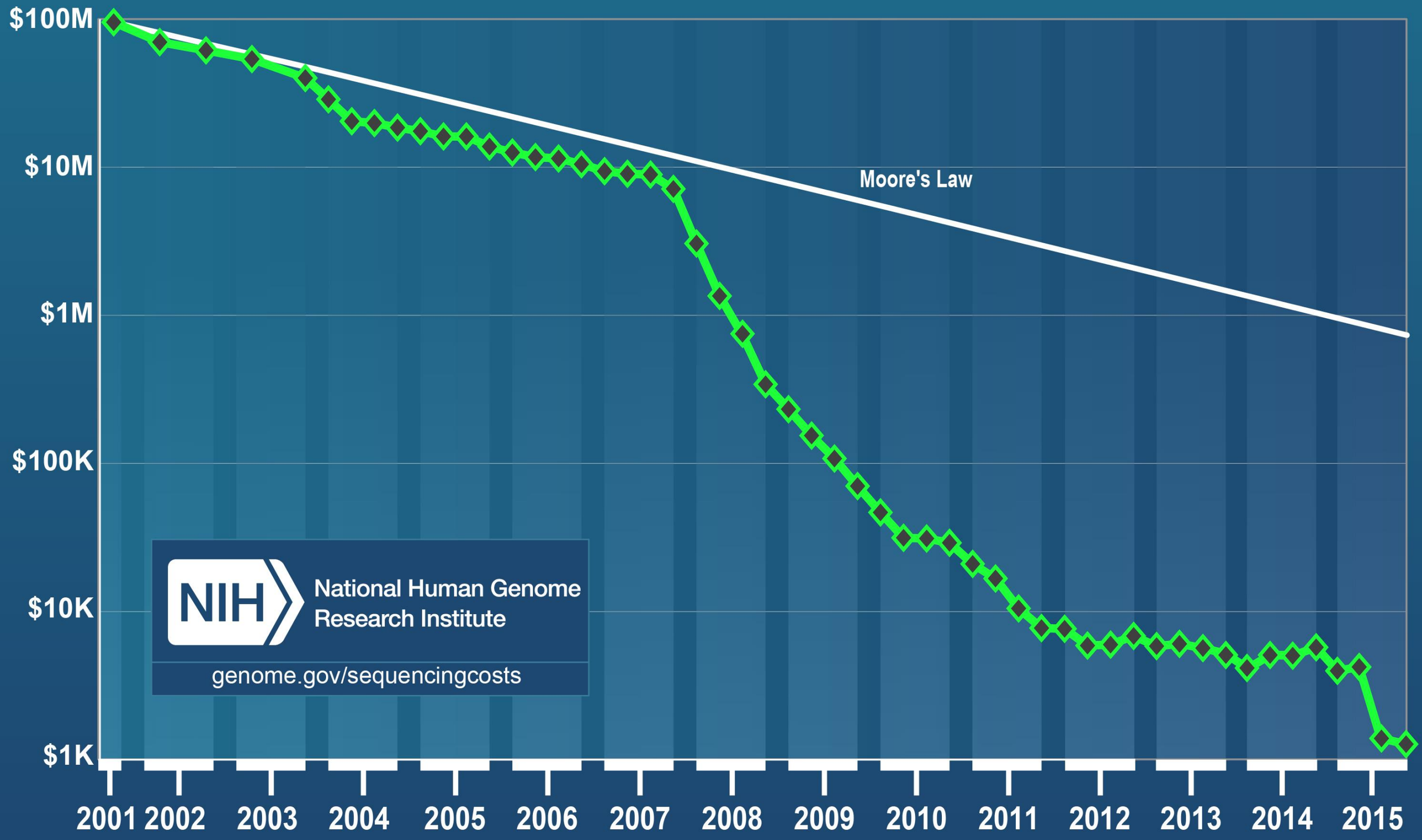
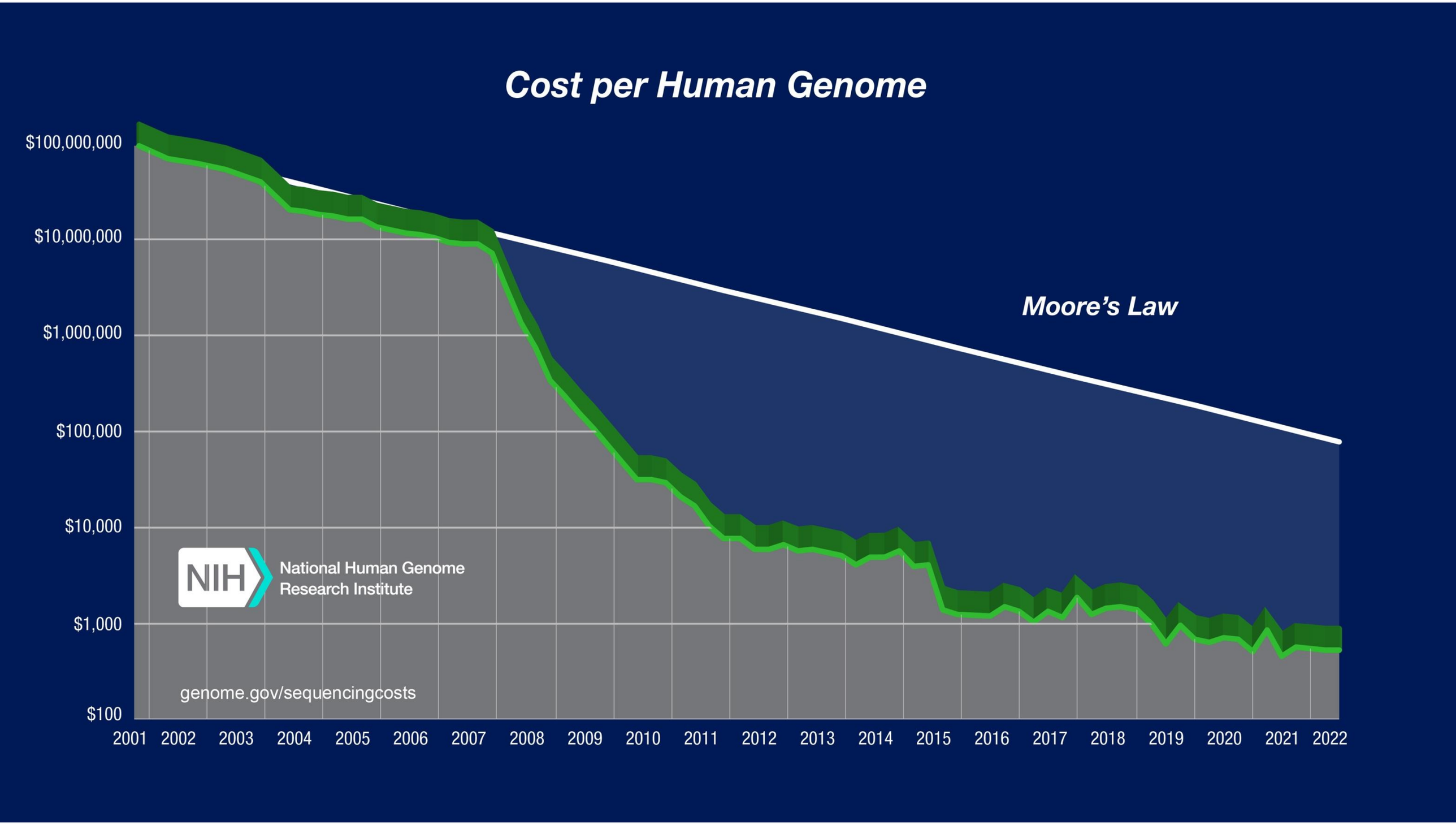
Plotting and Visualization with R

https://r4ds.had.co.nz/data-visualisation.html
https://is.gd/ggplot

Cost per Genome



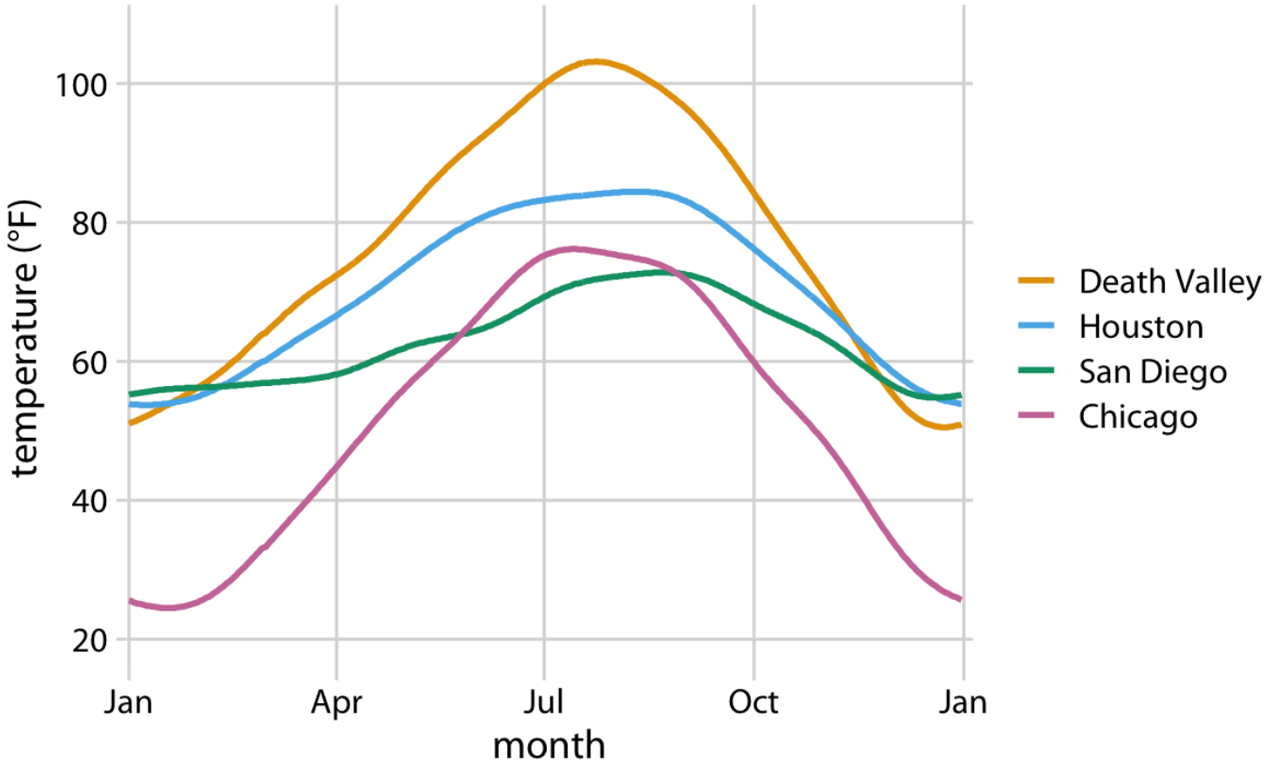

Cost per Human Genome
$100,000,000
$10,000,000
$1,000,000
$100,000
$10,000
$1,000
$100
Moore's Law
NIH
National Human Genome Research Institute
genome.gov/sequencingcosts
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

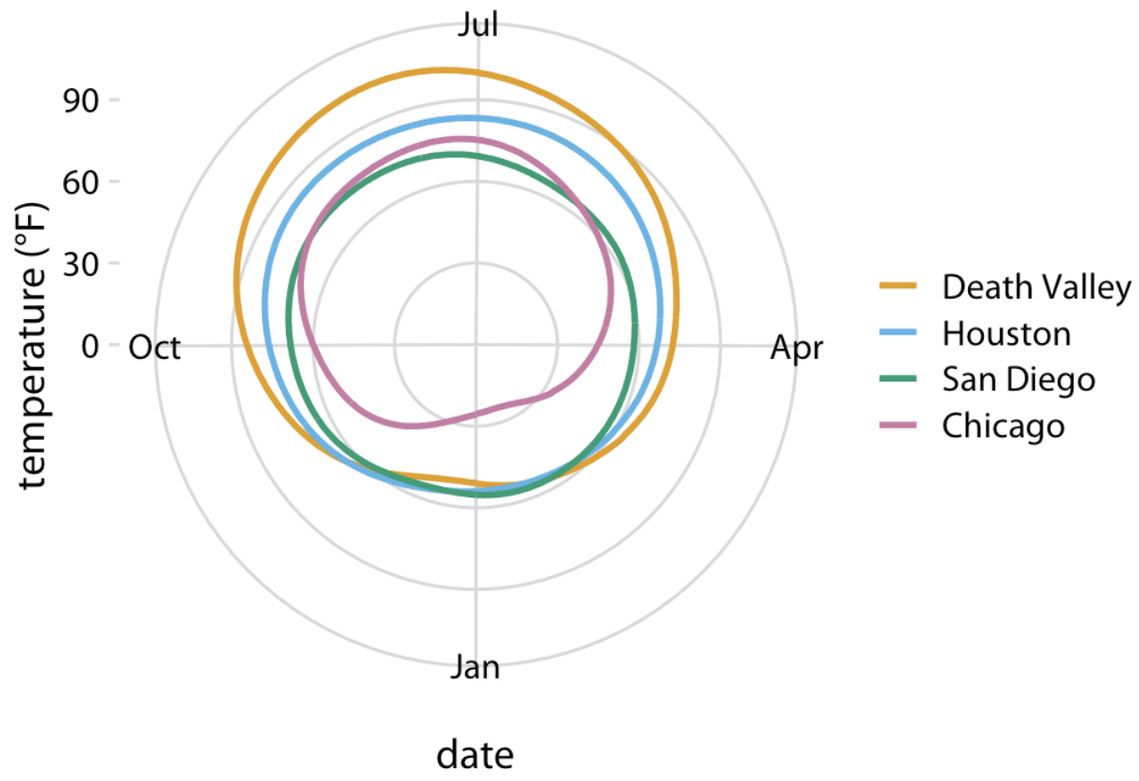

$100,000,000
$10,000,000
$1,000,000
$100,000
$10,000
$1,000
$100
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Graphic Design
IS
MY
PASSION

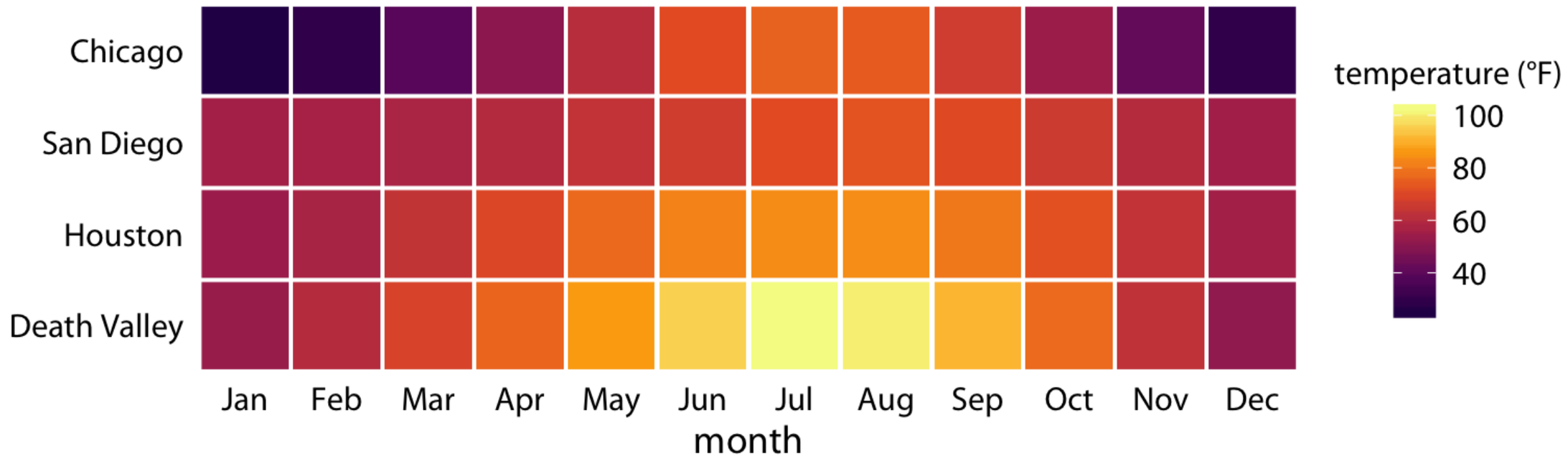
Fundamentals of Data visualization

Adapted from https://clauswilke.com/dataviz/

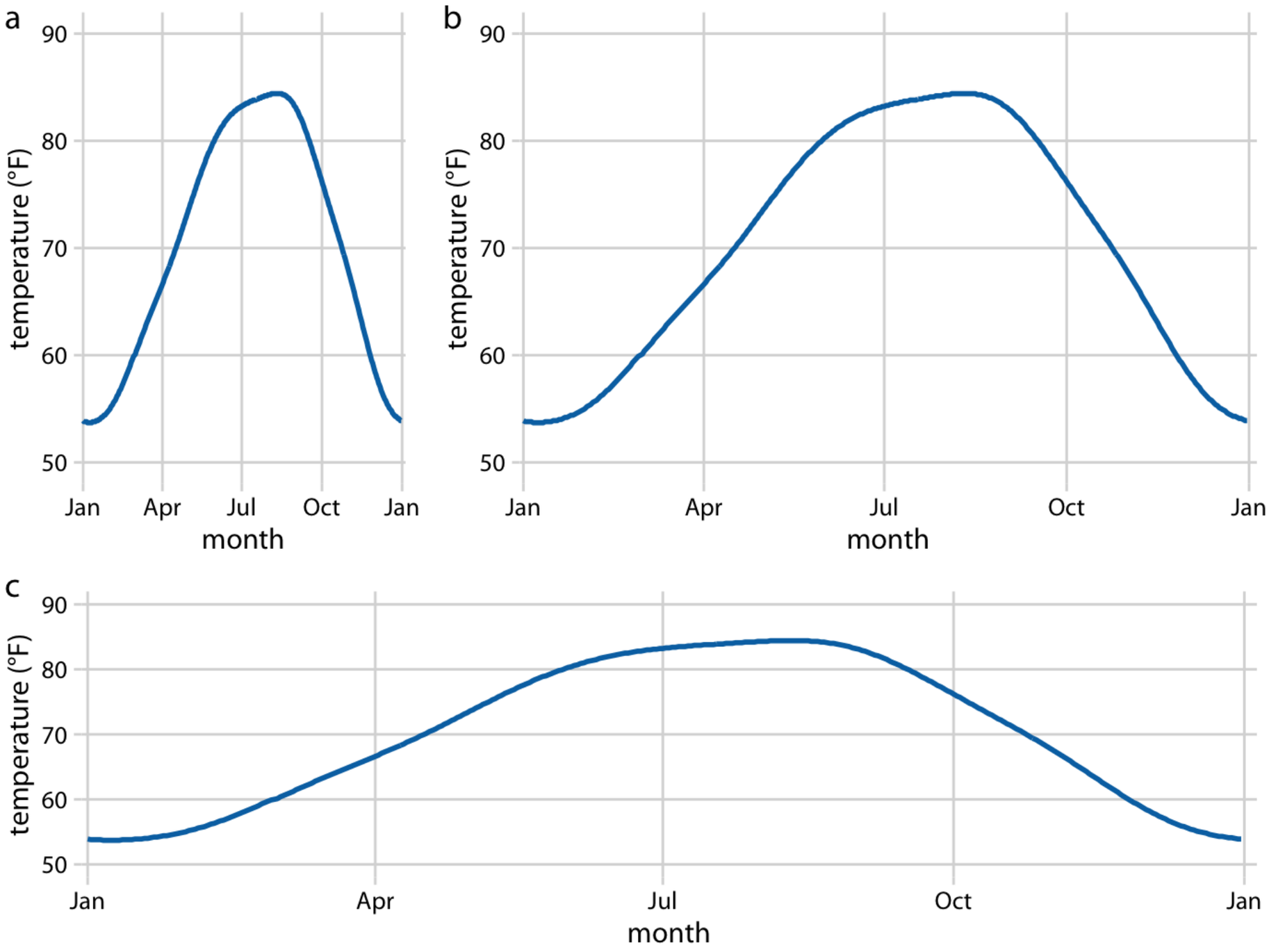
Alternative representations



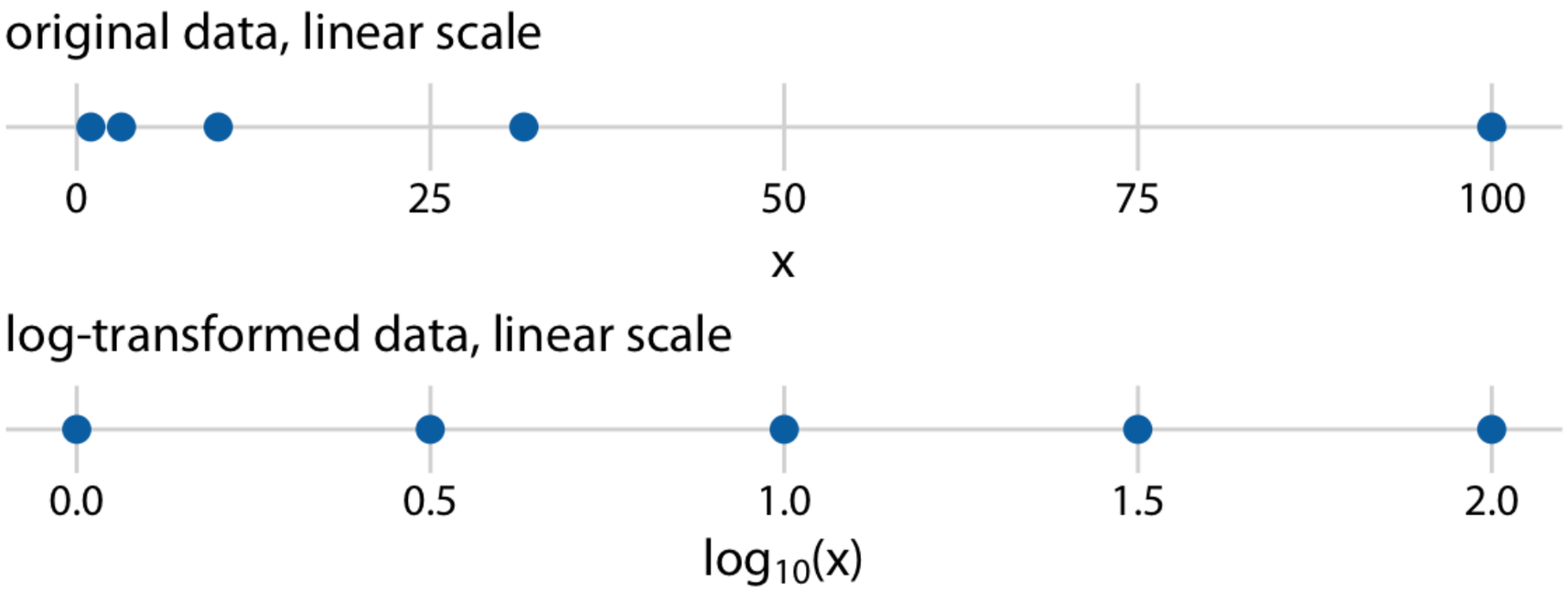




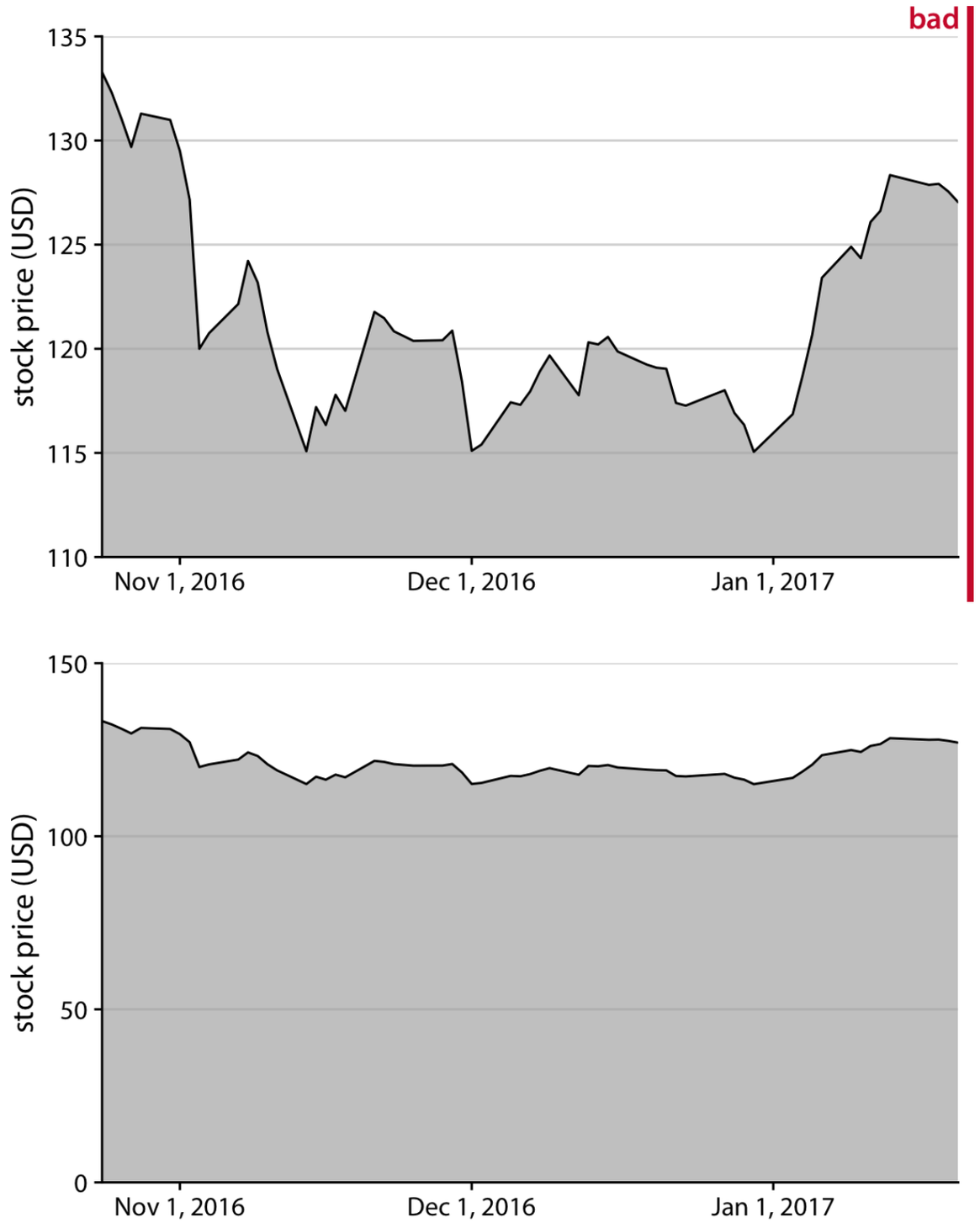
Aspect ratio



Scaling axes



Scaling axes



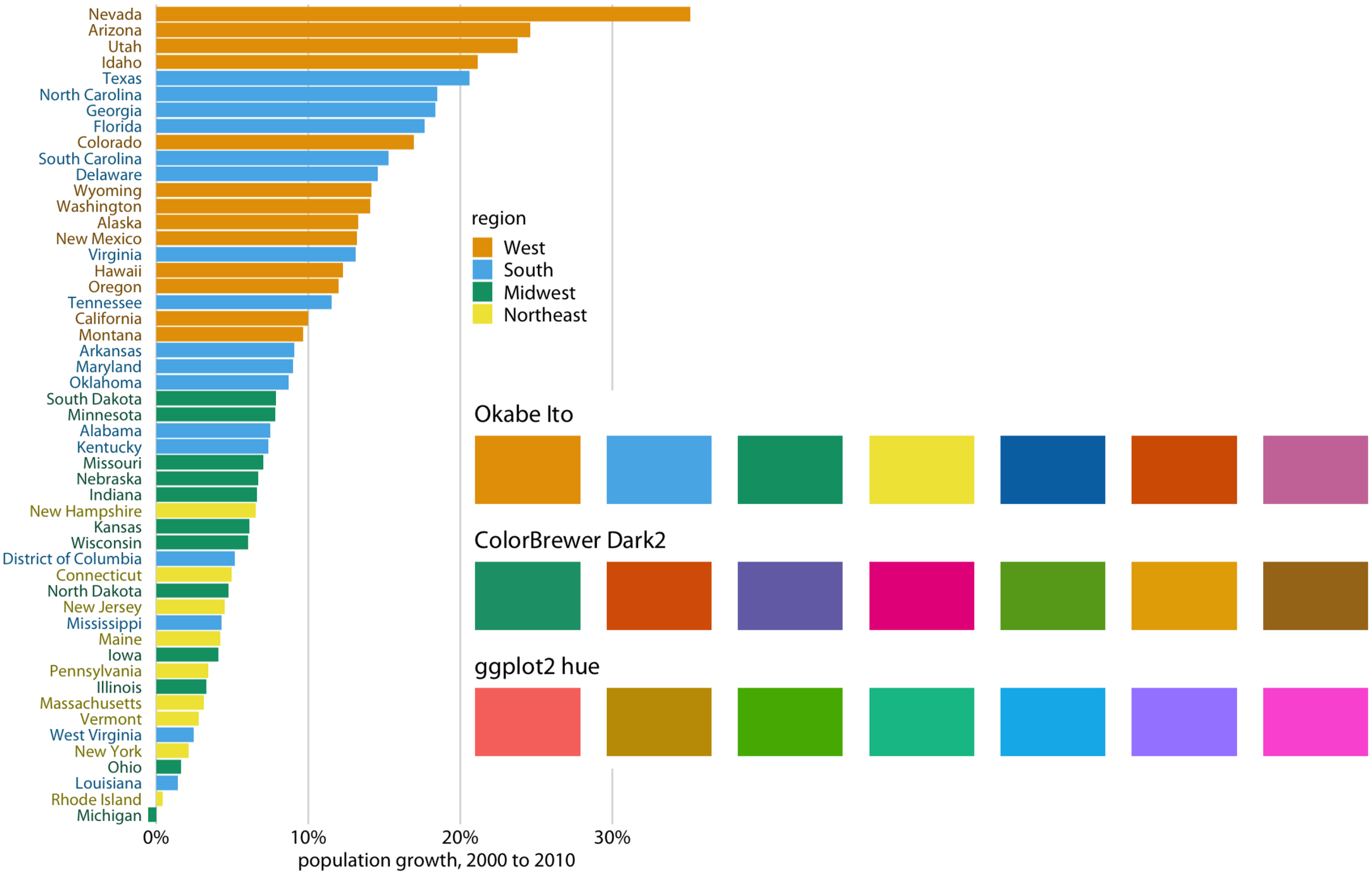
Color scales – three uses

1) To distinguish between groups

2) To represent data values

3) To highlight

Color to distinguish

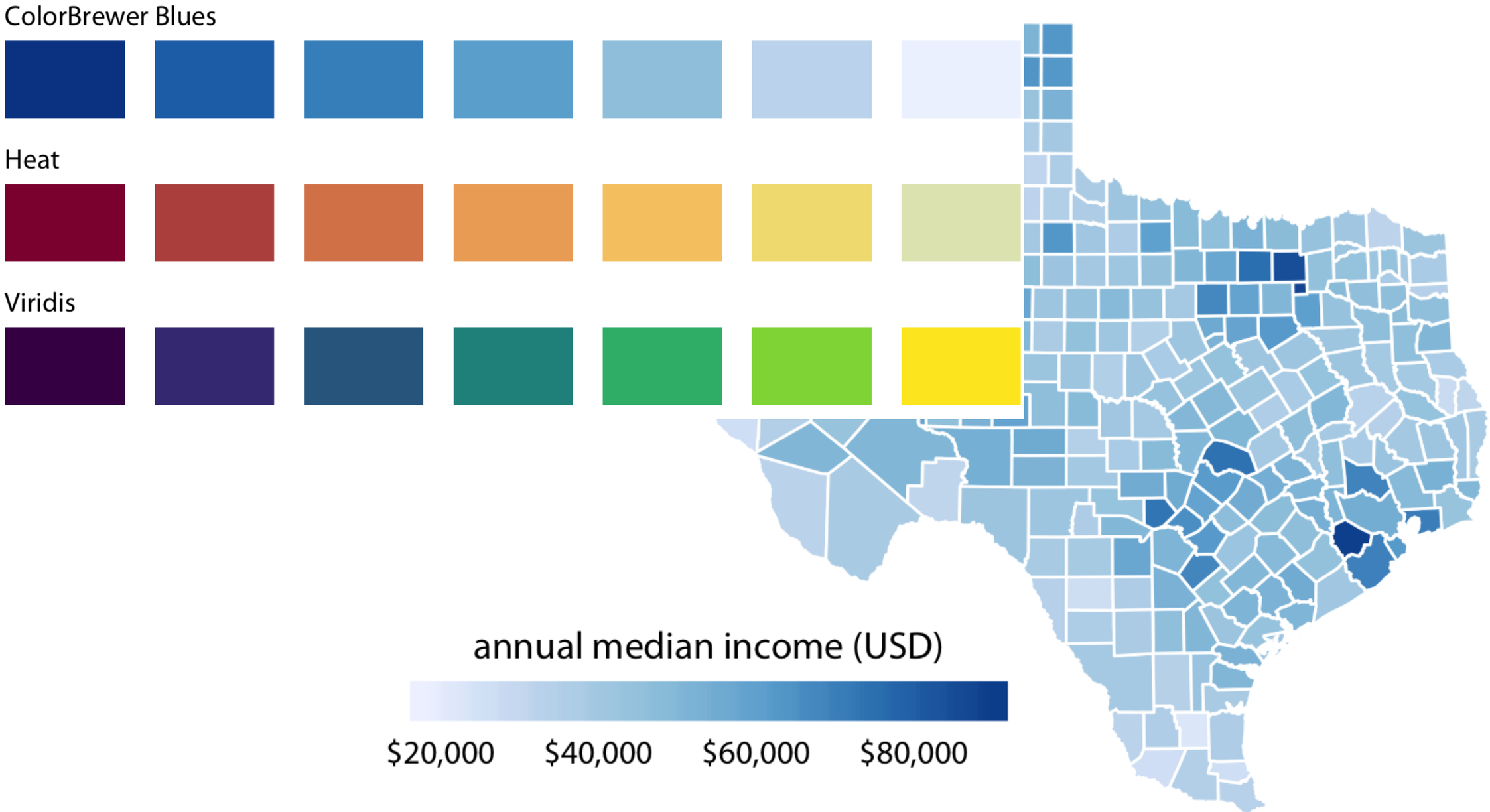


Color to represent data values

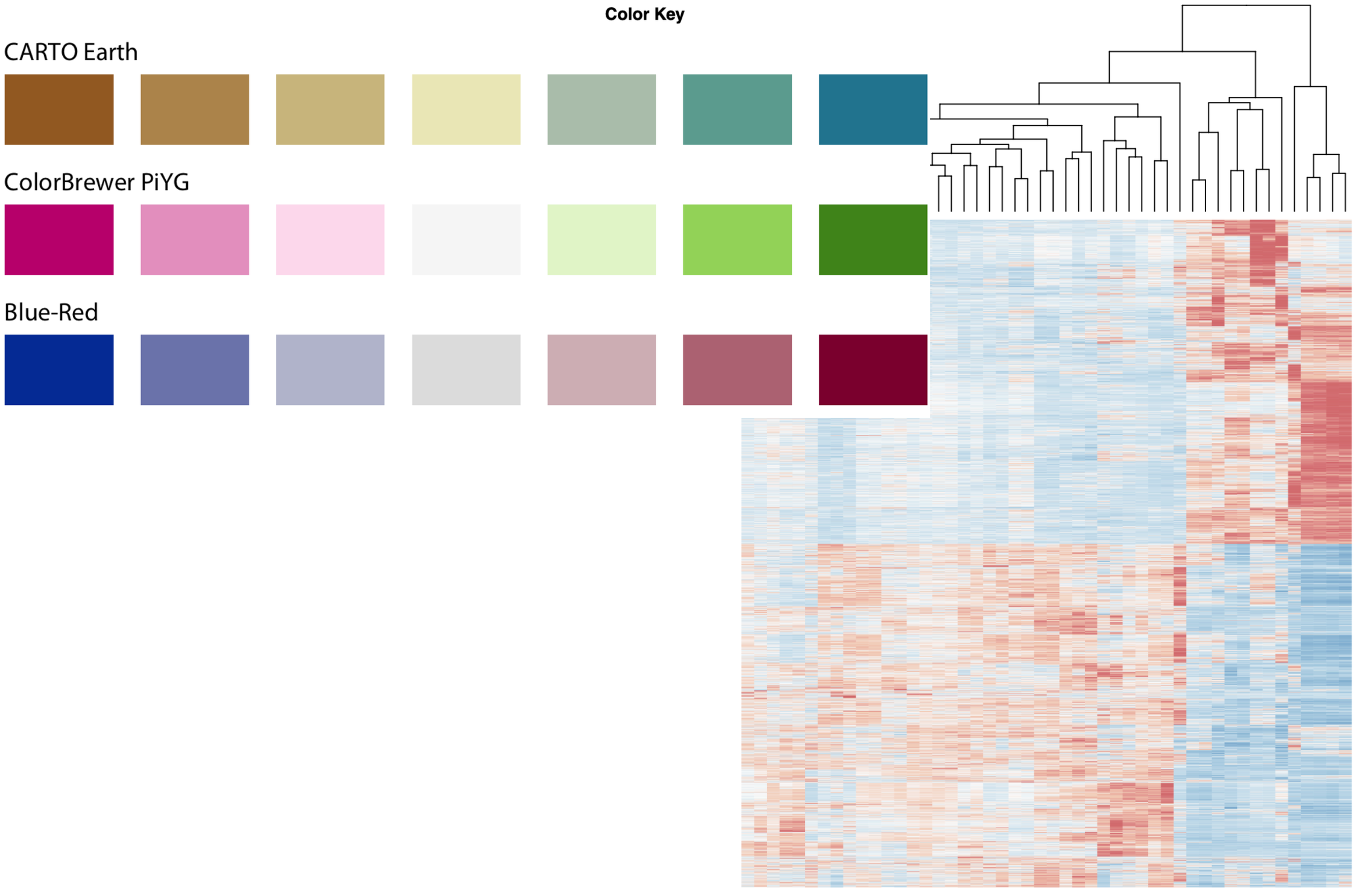
ColorBrewer Blues

Heat

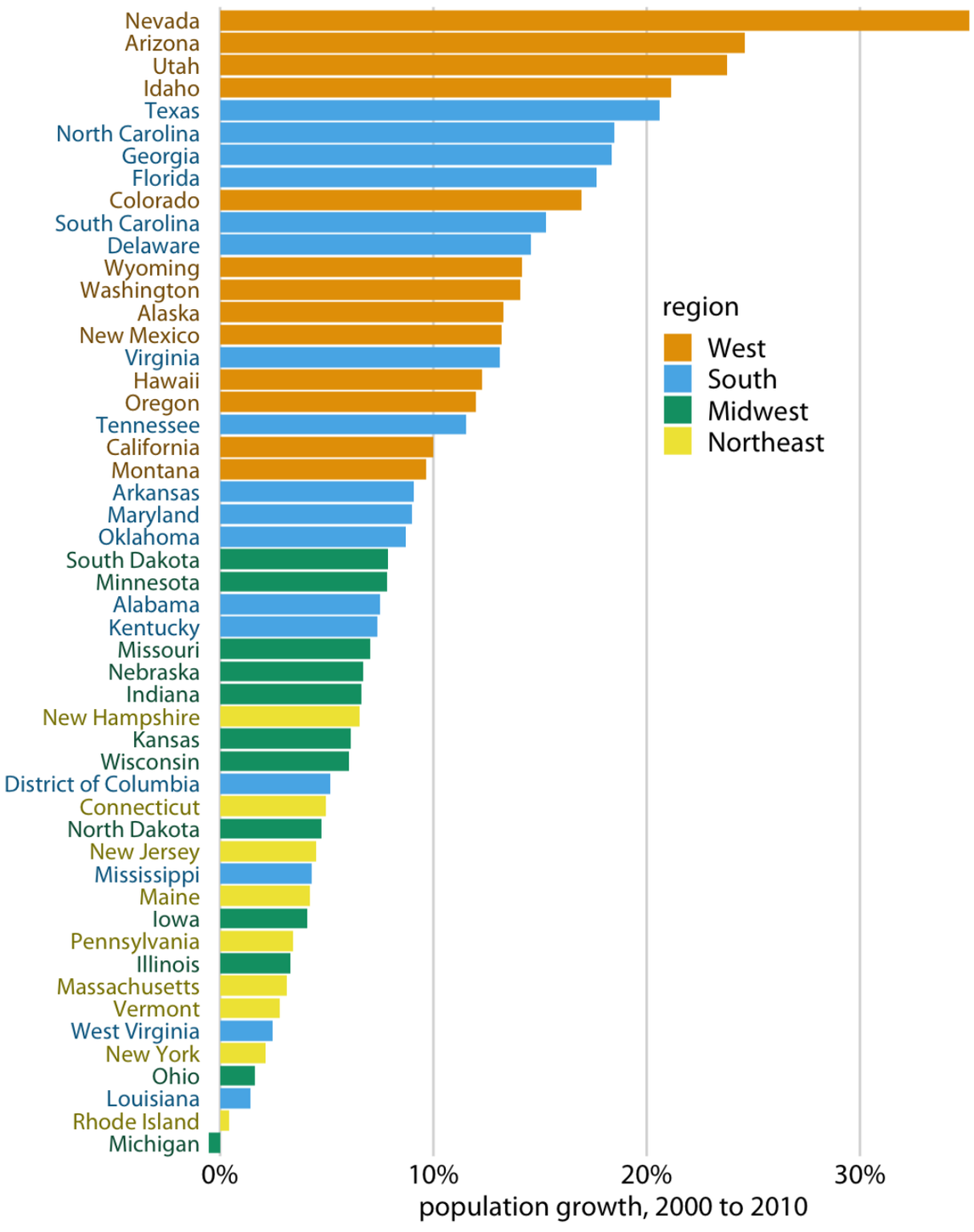
Viridis



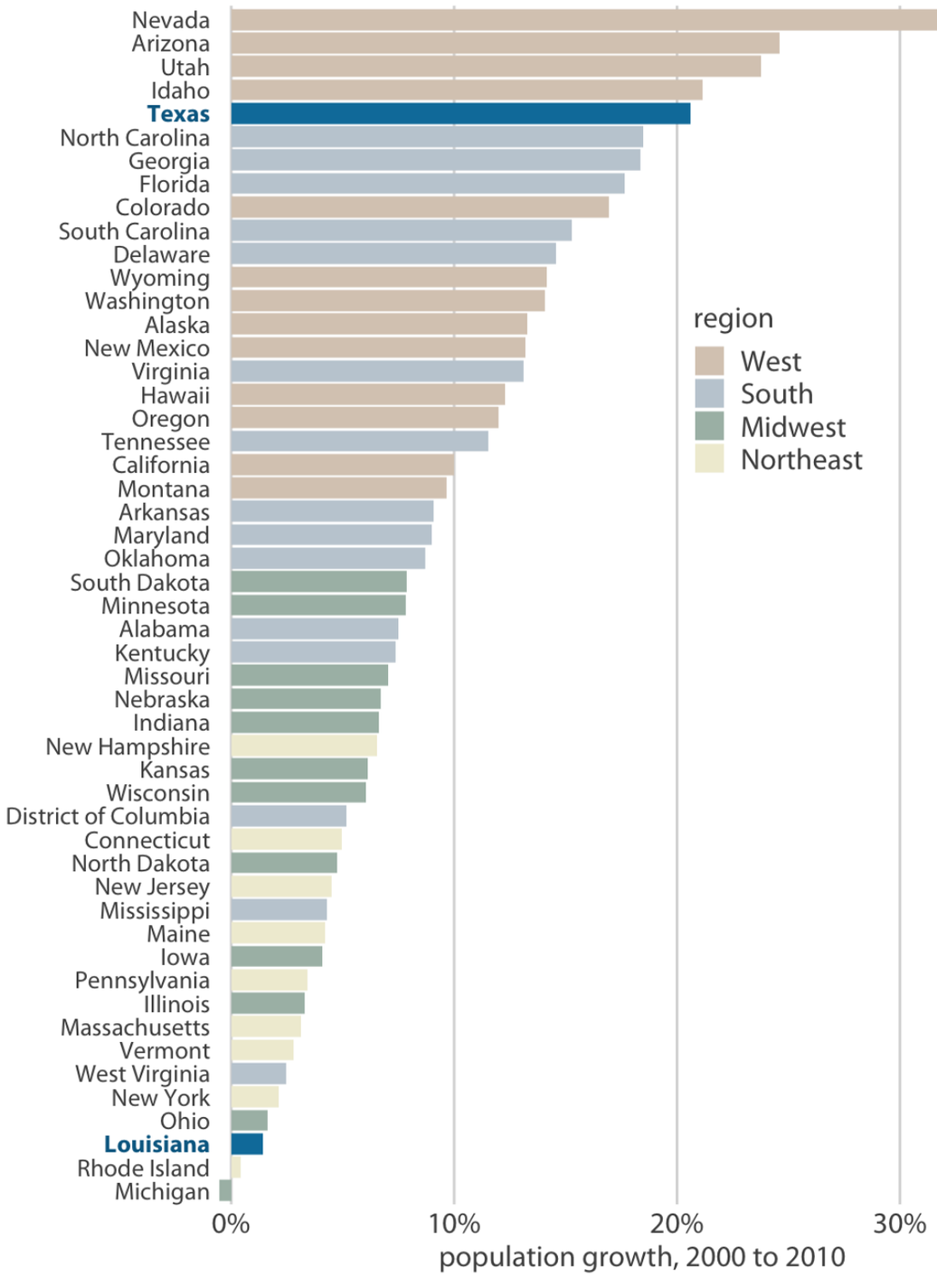
Color to represent data values

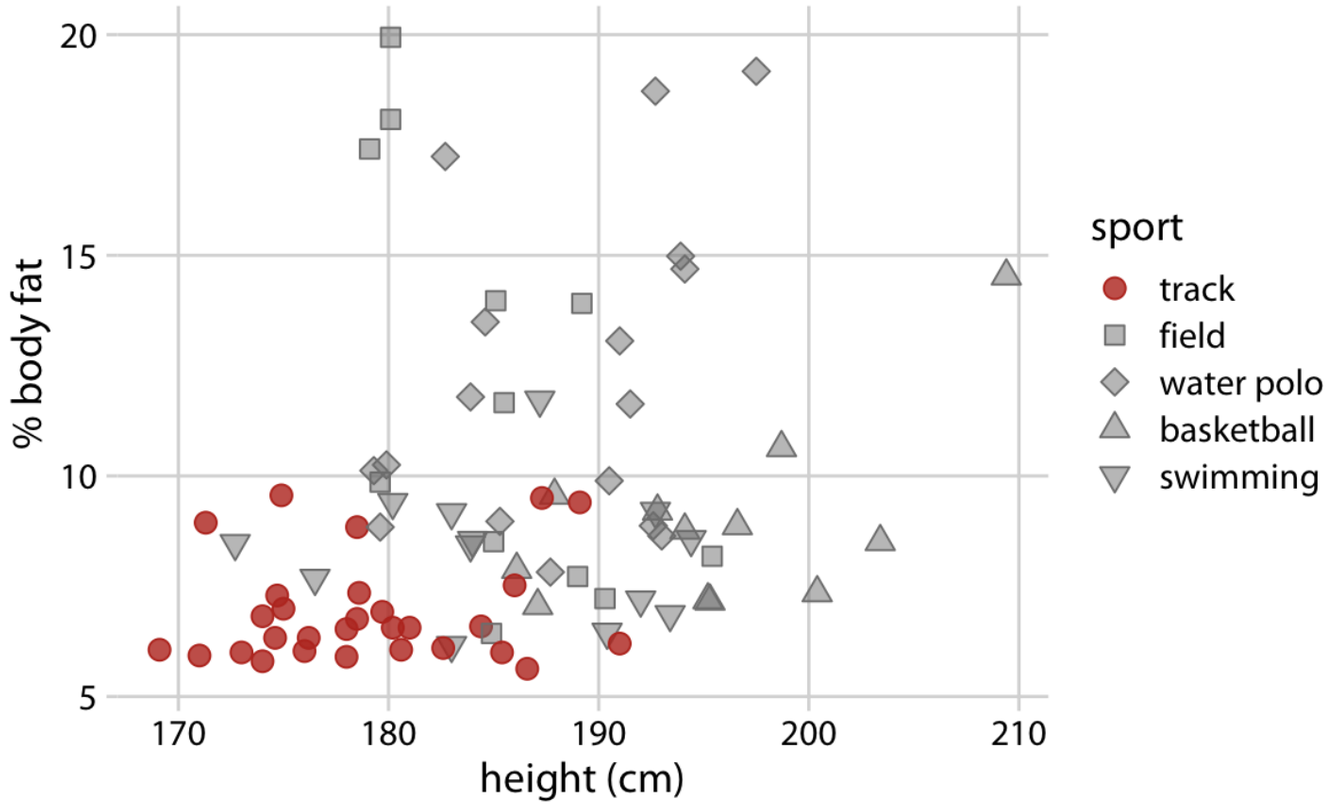


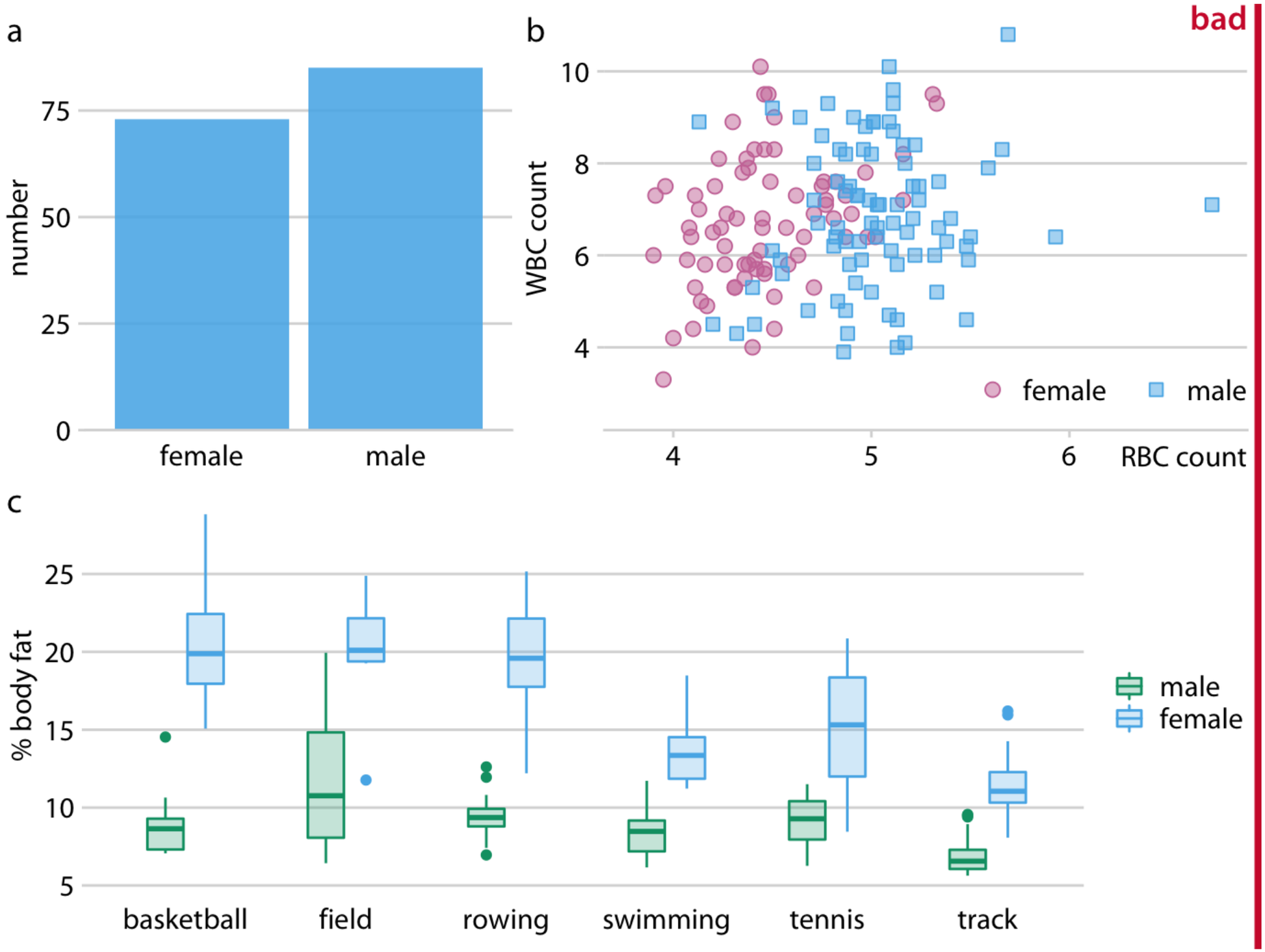
Color to highlight



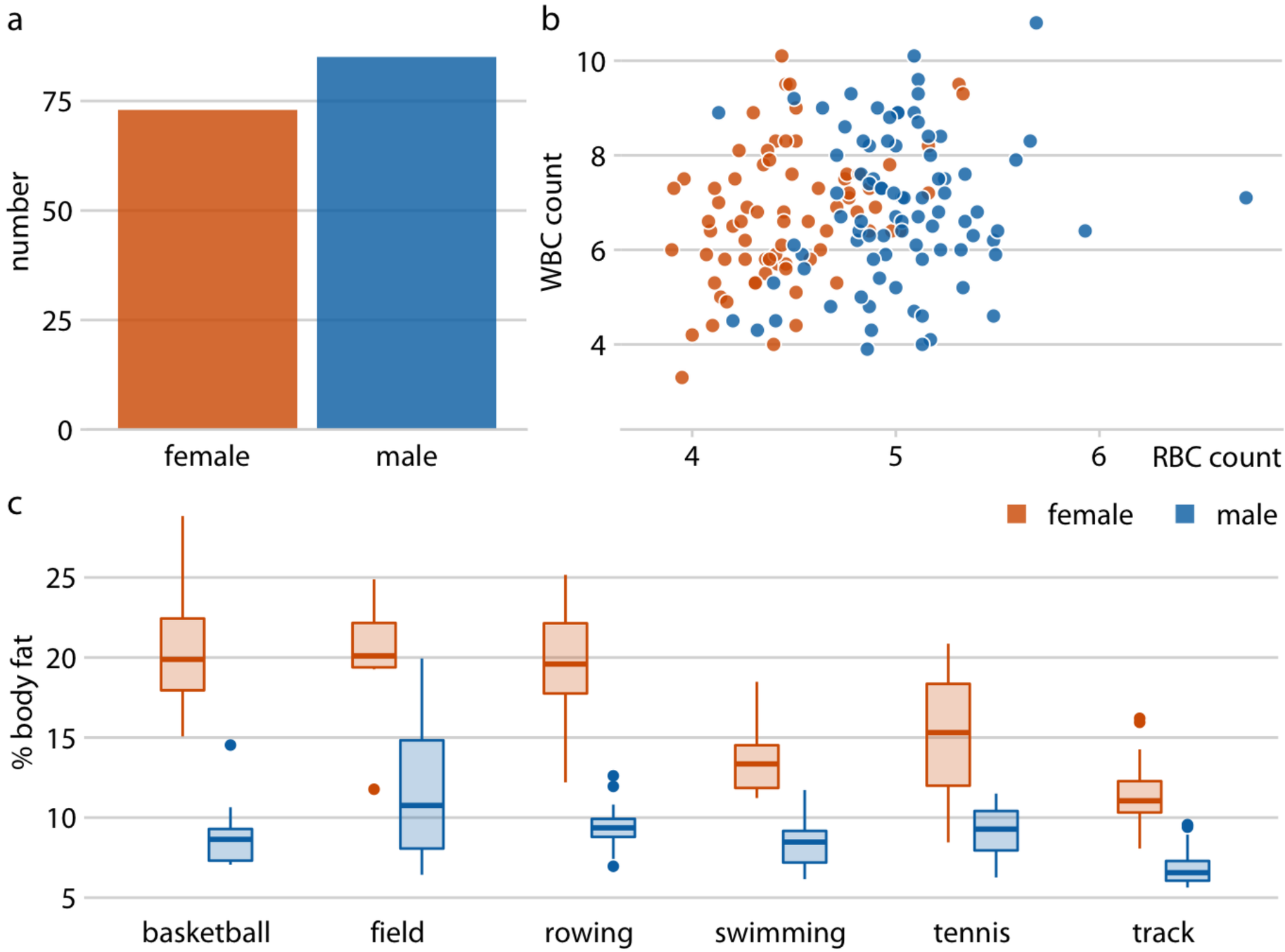
Color to highlight



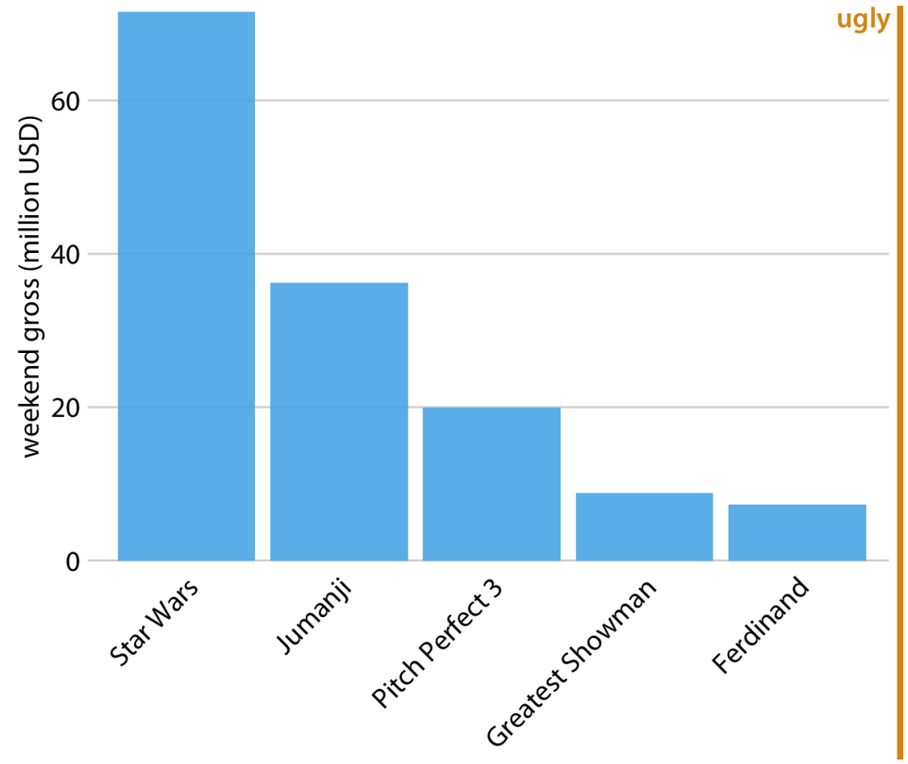


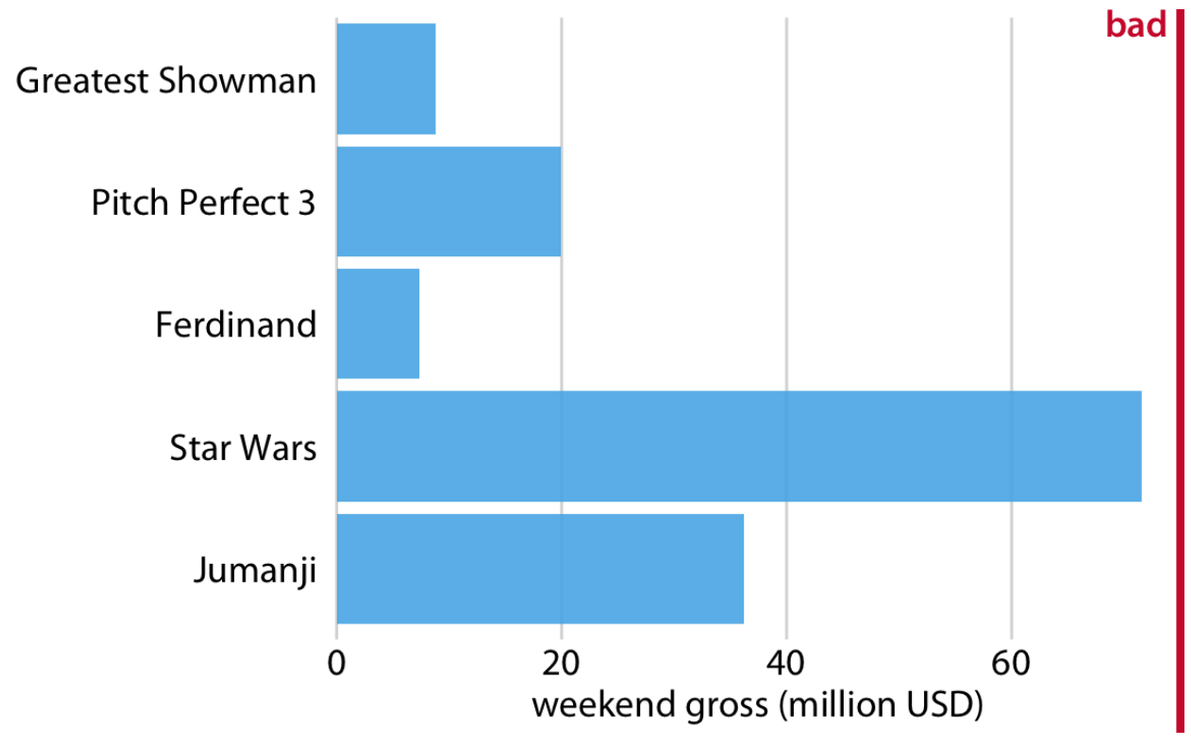

a
number
75
50
25
0
female
male
b
bad
WBC count
10
8
6
4
female
male
4
5
6
RBC count
c
% body fat
25
20
15
10
5
basketball
field
rowing
swimming
tennis
track
male
female

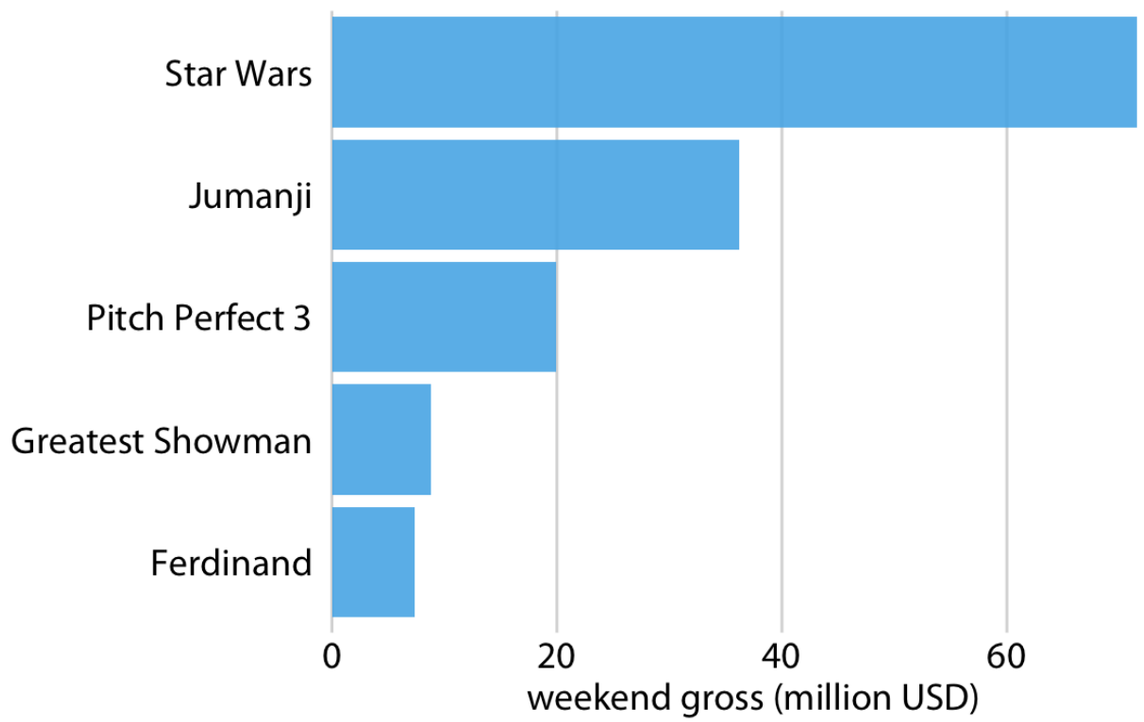
Be consistent across panels and figures



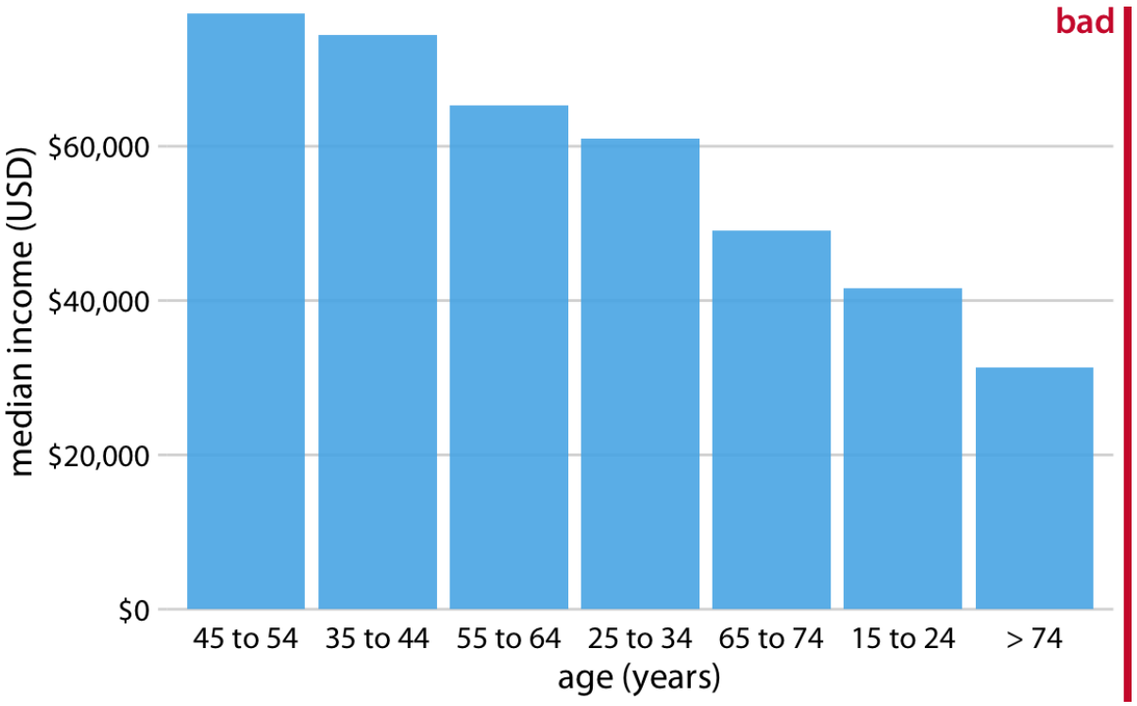
Bar plots

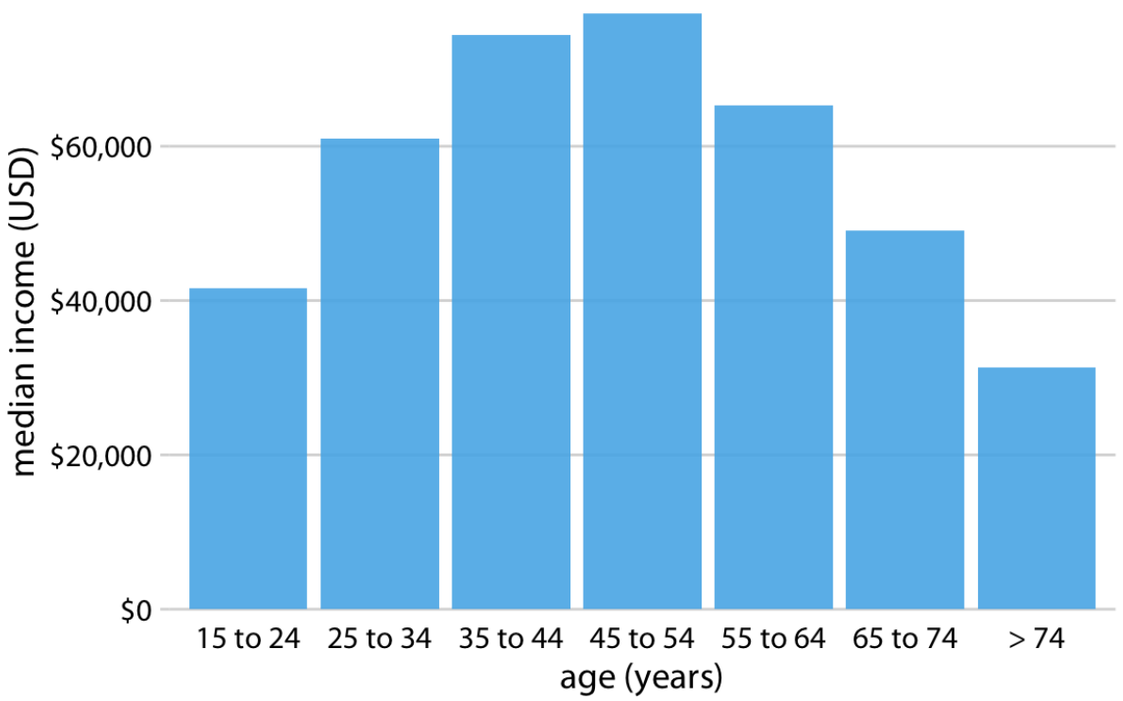

ugly
weekend gross (million USD)
60
40
20
0
Star Wars
Jumanji
Pitch Perfect 3
Greatest Showman
Ferdinand


bad
Greatest Showman
Pitch Perfect 3
Ferdinand
Star Wars
Jumanji
0
20
40
60
weekend gross (million USD)

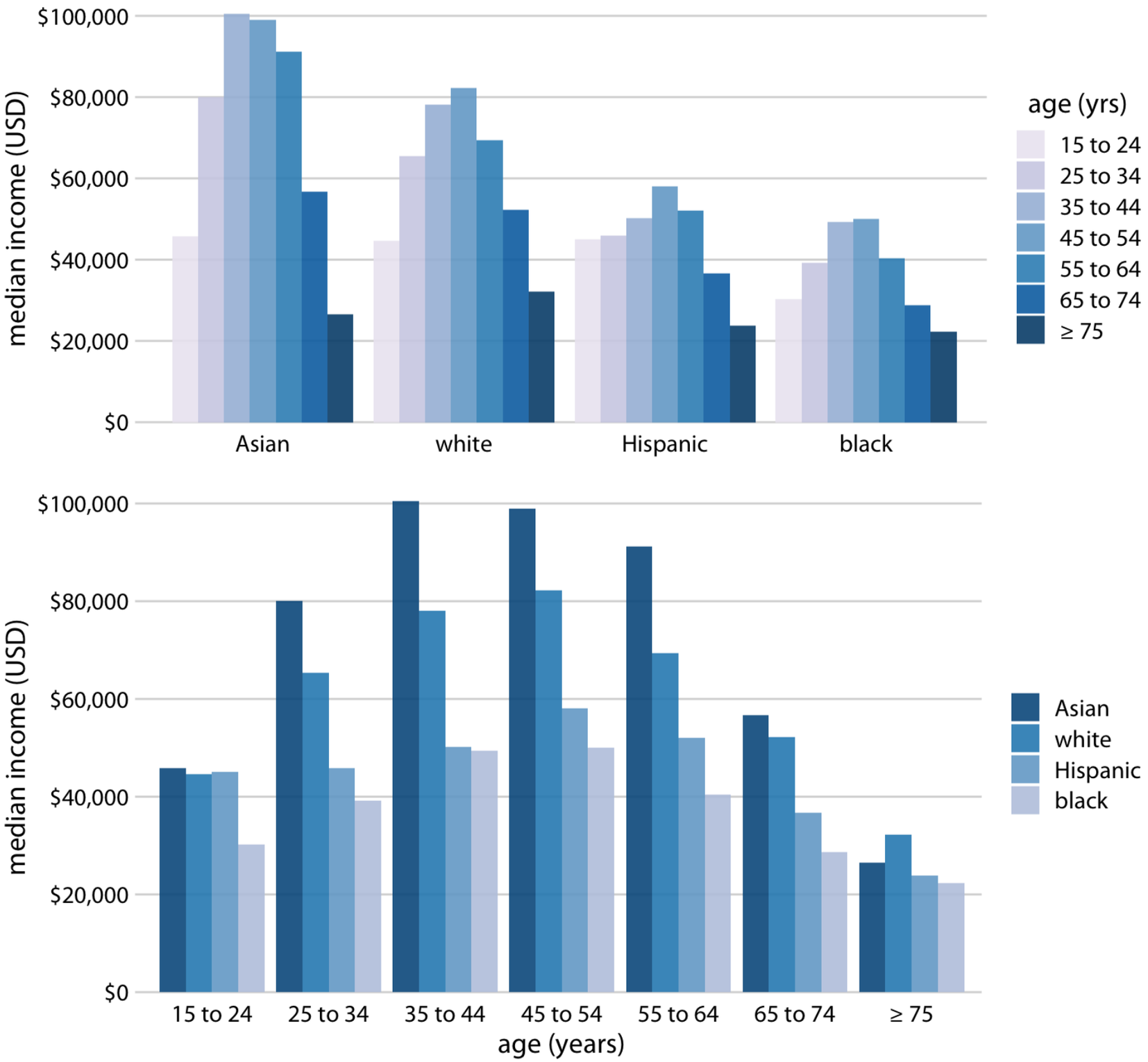

Star Wars
Jumanji
Pitch Perfect 3
Greatest Showman
Ferdinand
0
20
40
60
weekend gross (million USD)

Bar plots

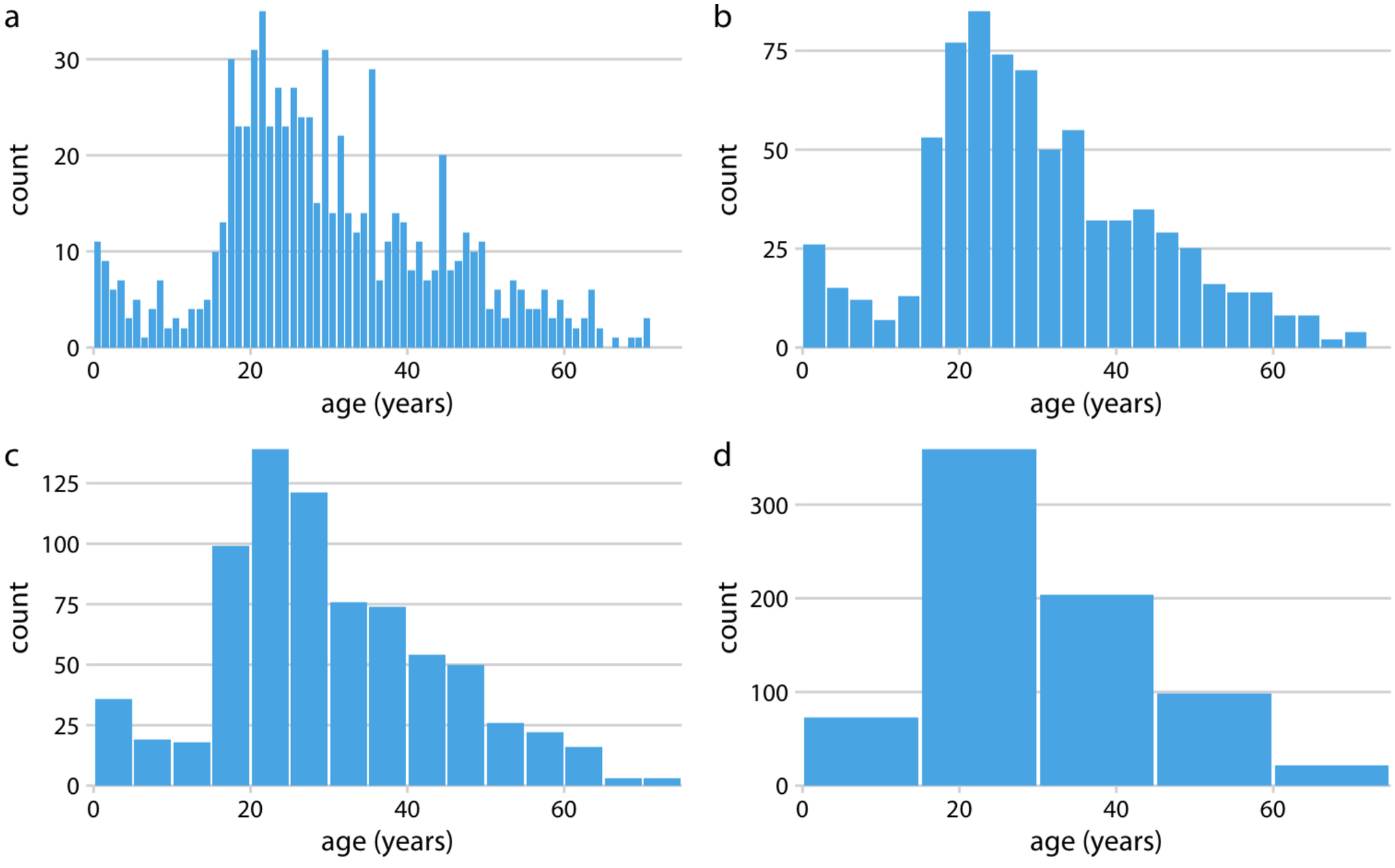

bad
$60,000
$40,000
$20,000
$0
median income (USD)
45 to 54
35 to 44
55 to 64
25 to 34
65 to 74
15 to 24
> 74
age (years)


$60,000
$40,000
$20,000
$0
median income (USD)
15 to 24
25 to 34
35 to 44
45 to 54
55 to 64
65 to 74
> 74
age (years)

Grouped Bar Plots



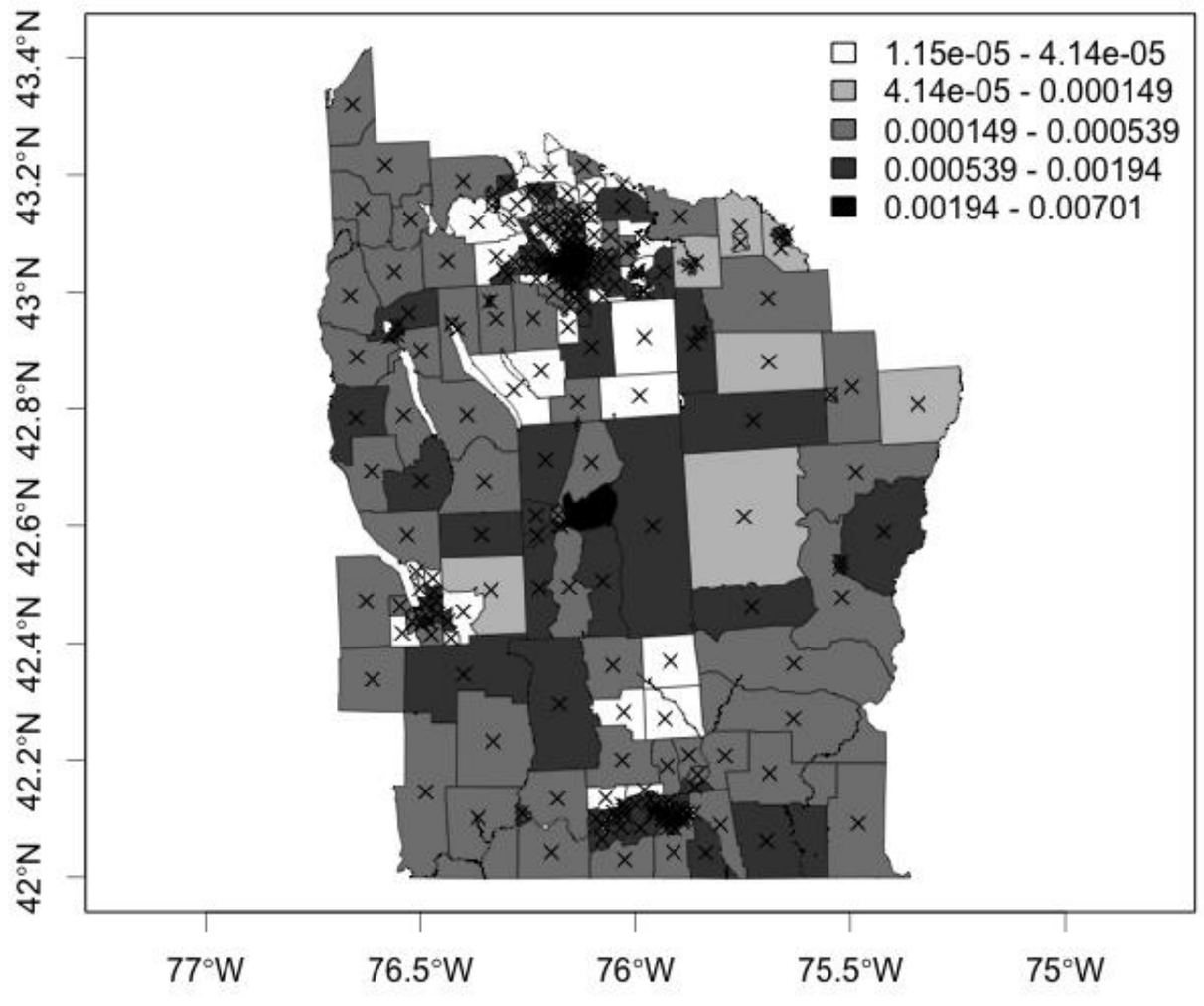
Granularity of your data



Advanced and Other Applications in R

Spatial Epidemiology and Maps

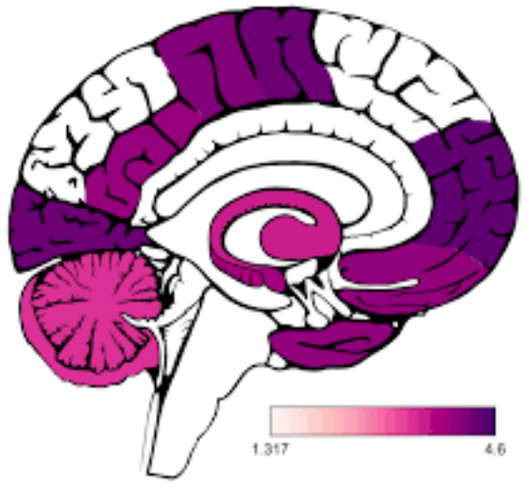
```
library(SpatialEpi)

data(NYleukemia)
sp.obj <- NYleukemia$spatial.polygon
centroids <- latlong2grid(NYleukemia$geo[, 2:3])
population <- NYleukemia$data$population
cases <- NYleukemia$data$cases

plotmap(cases/population, sp.obj, log=TRUE, nclr=5)
points(grid2latlong(centroids), pch=4)
```



Anatomical Mapping

CerebroViz

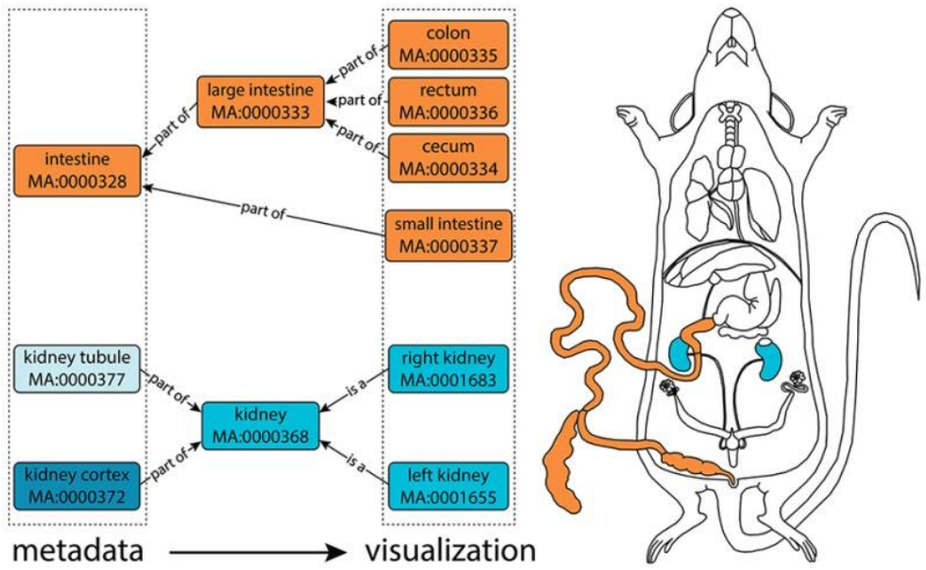
https://github.com/ethanbahl/cerebroViz




```
library("cerebroViz")
data(cerebroEx)
head(cerebroEx)[, c(1:7)]
```

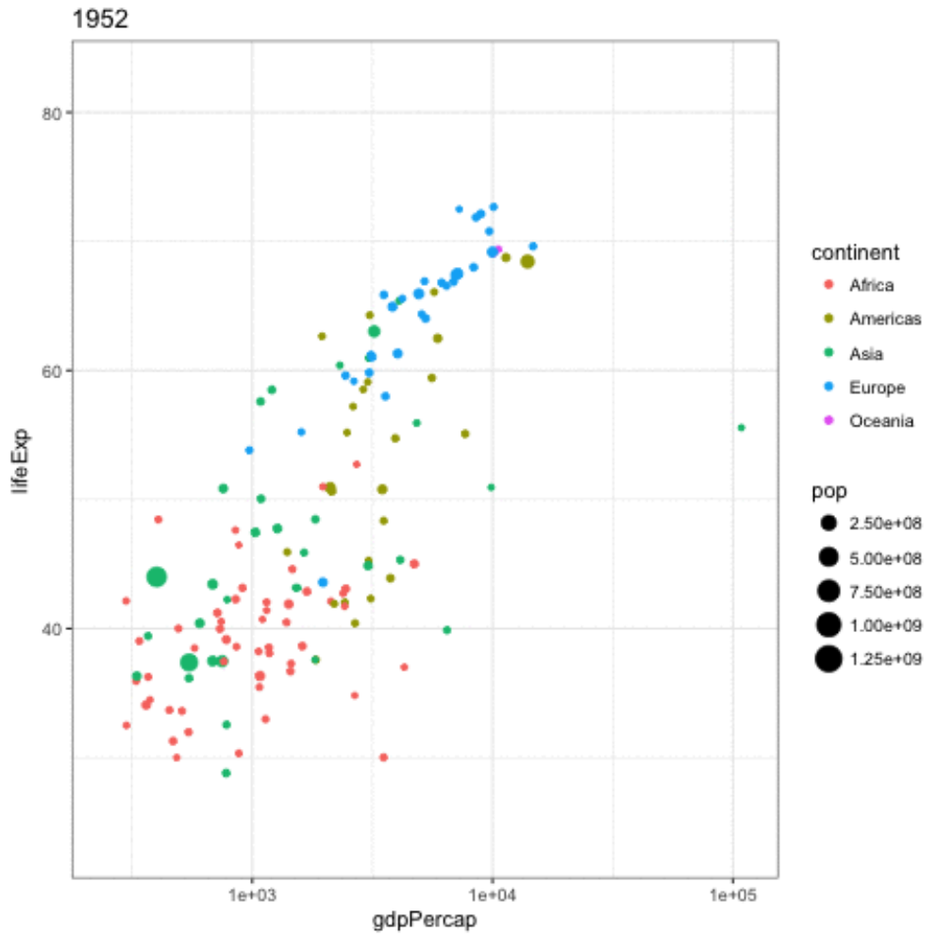
COMICS

https://github.com/y-popov/COMICS



Animation / Interactivity

Indicator Name	2011	2012	2013	2014	2015	2016	Average	Improvement
Prevalence of Obesity	19.1	23.6	23.3	20.5	24.0	23.2	22.28	-21.47
Prevalence of Tobacco Use	17.4	15.0	15.3	12.2	16.6	16.7	15.53	4.02
Prevalence of Cardiovascular Disease	5.0	4.9	1.5	4.4	4.9	6.2	4.48	-24.00
Prevalence of Diabetes	8.0	7.2	9.3	7.2	7.5	10.4	8.27	-30.00



Expose yourself to lots of ideas!