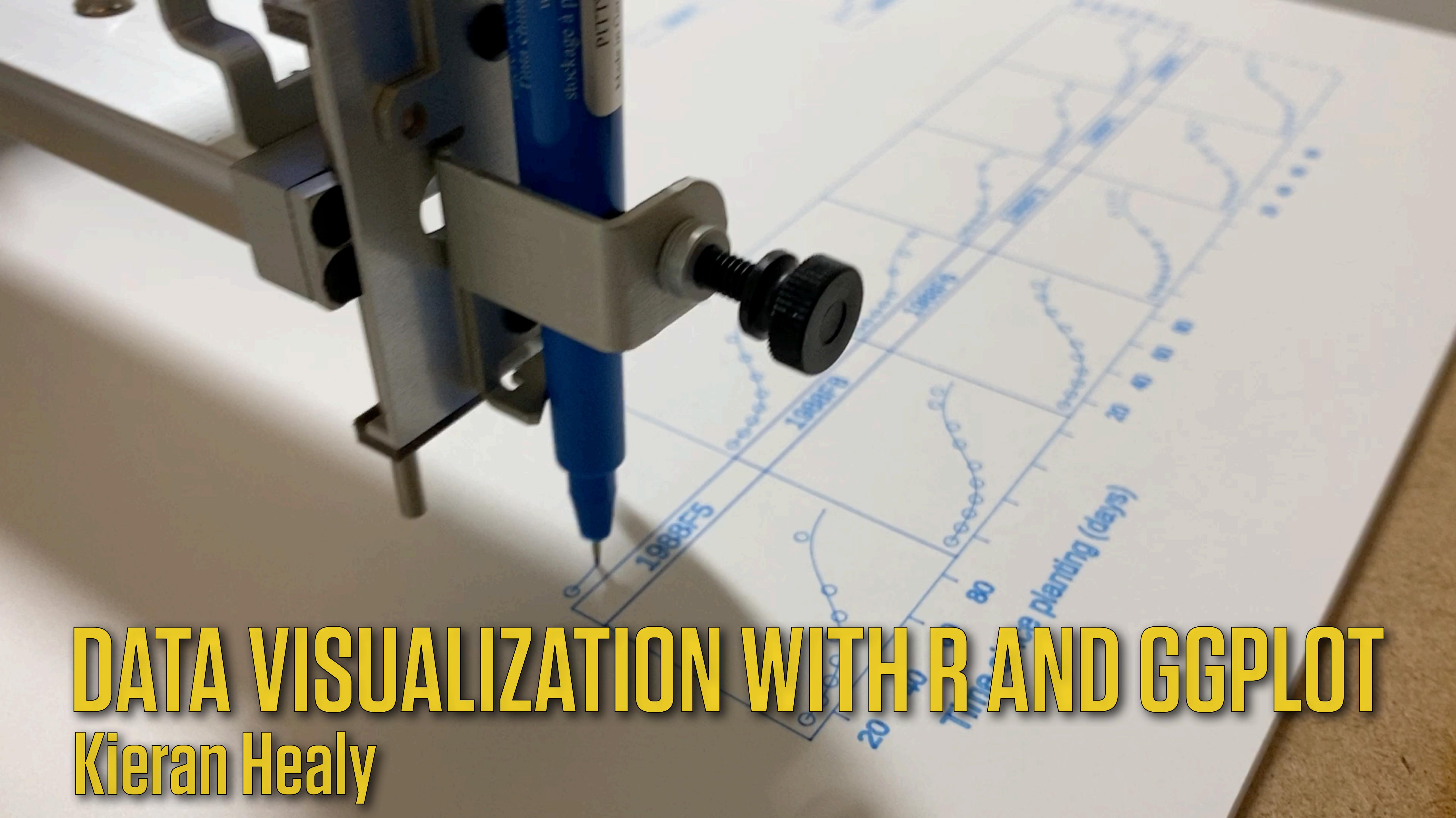
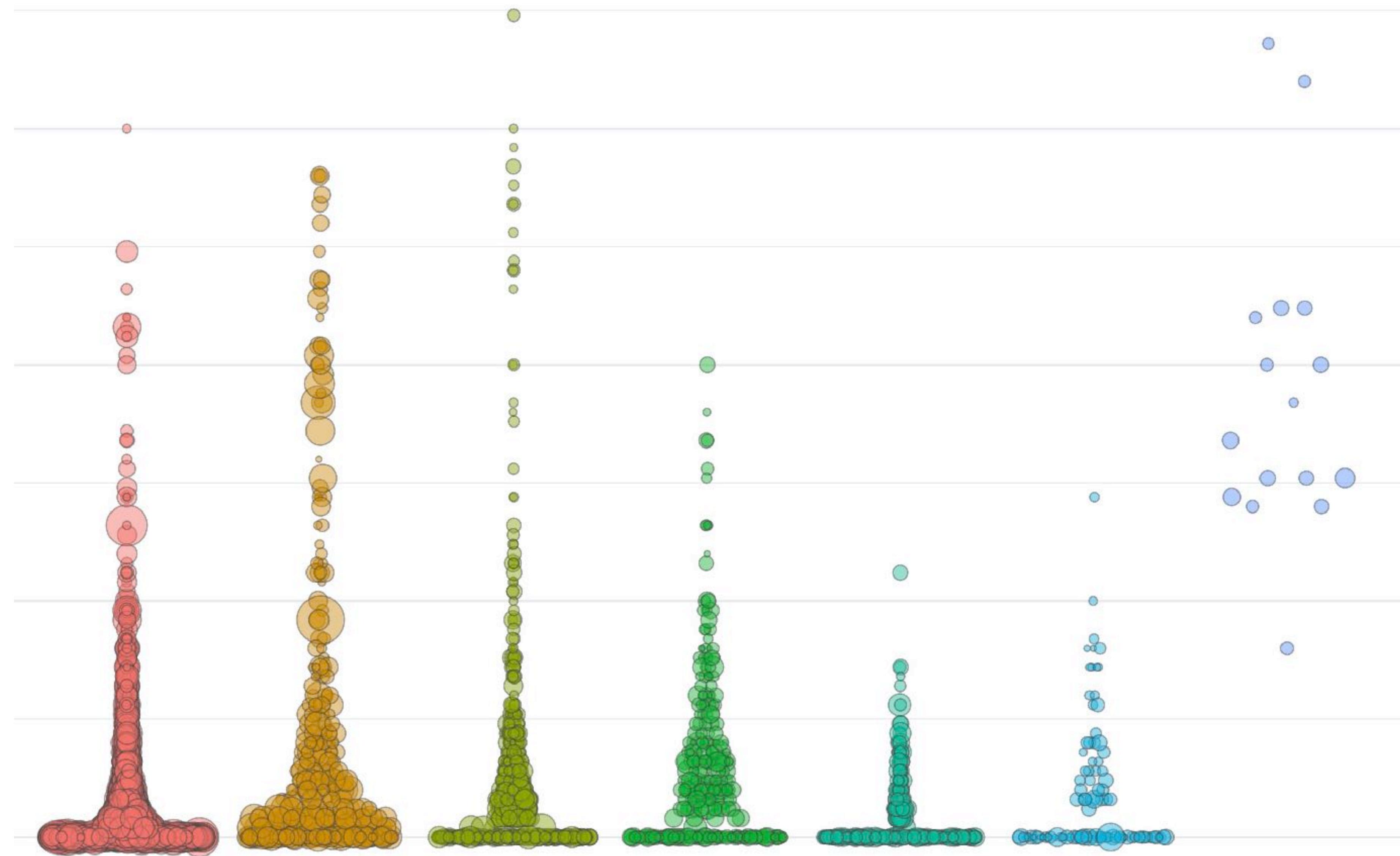


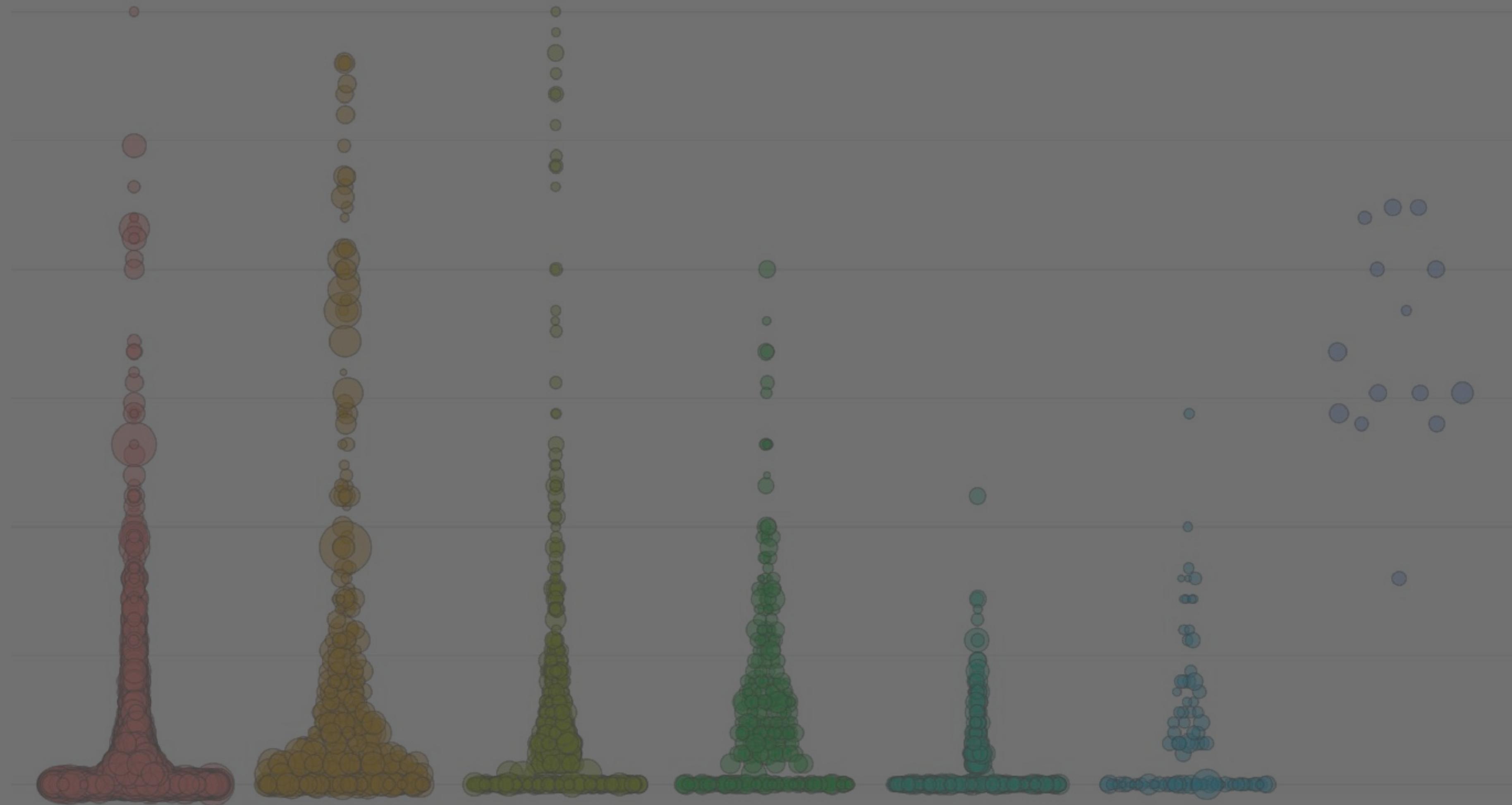
DATA VISUALIZATION WITH R AND GGPLOT

Kieran Healy





DATA VISUALIZATION





RStudio

Get Up & Running

Console

~ /Desktop/

> |

Get the Notes from inside RStudio

```
# Read the options for where to put it  
usethis::use_course("http://kjhealy.co/sh0820.zip")
```

Or just download from your Browser



kjhealy.co/sh0820.zip

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

stathorizons_0820 - master - RStudio

Console Jobs ~/Documents/courses/stathorizons_0820/

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

knitr hook "anchor" is now available
Loading required package: testthat

Attaching package: 'testthat'

The following object is masked from 'package:devtools':
  test_file

> |
```

RStudio

RStudio

The screenshot shows the RStudio interface with several panes:

- Code Editor (highlighted by a red box):** Displays an R Markdown file named "00_open_me_first.Rmd". The code includes setup instructions, a note about knitting, and a section for installing packages.
- Environment:** Shows the Global Environment pane which is currently empty.
- Files:** Shows a file browser with the following contents:

Name	Size	Modified
..	40 B	Jul 21, 2020, 11:16 AM
.gitignore	1.8 KB	Jul 23, 2020, 2:15 PM
00_open_me_first.Rmd	4 KB	Jul 23, 2020, 10:18 AM
01_introduction.Rmd	4 KB	Jul 23, 2020, 10:25 AM
02_get_started.Rmd	5.8 KB	Jul 21, 2020, 11:16 AM
03_make_a_plot.Rmd	5 KB	Jul 23, 2020, 11:27 AM
04_show_the_right_numbers.Rmd	10.8 KB	Jul 23, 2020, 11:12 AM
05_tables_and_labels.Rmd	14.7 KB	Jul 23, 2020, 11:54 AM
06_models.Rmd	12.5 KB	Jul 21, 2020, 11:16 AM
07_maps.Rmd	21.7 KB	Jul 21, 2020, 11:16 AM
08_refine_plots.Rmd	17 KB	Jul 23, 2020, 12:13 PM
09_supplementary_material.Rmd		
assets		
data		
figures		
keynote		
LICENSE.md	18.1 KB	Jul 21, 2020, 11:16 AM
materials		
README.md	5.7 KB	Jul 21, 2020, 11:16 AM
slides		
stathorizons_0820.Rproj	205 B	Jul 23, 2020, 2:19 PM

- Console:** Displays R session output, including package loading and a warning about a masked object from devtools.



RStudio

File

Edit

Code

View

Plots

Session

Build

Debug

Profile

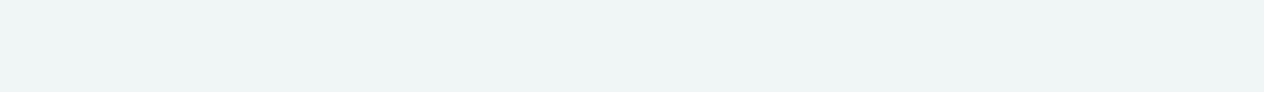
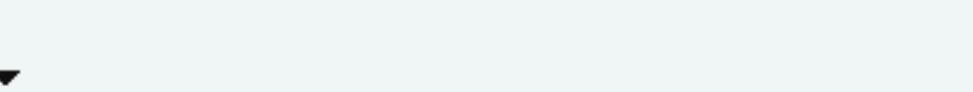
Tools

Window

Help



stathorizons_0820 - master - RStudio



00_open_me_first.Rmd



Insert



Run



```
1 ---  
2 title: "Installation and Setup"  
3 author: "Kieran Healy"  
4 date: "7/23/2020"  
5 output: html_document  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10```  
11  
12## Start Here  
13  
14When you click the Knit button above (it looks like a ball of wool with  
a knitting needle in it) a document will be generated that runs the R code  
below. The code will install some add-on packages for the course. An  
internet connection is required. You should only have to run this code once.  
15  
16```{r}
```

Environment

History

Connecti



Import Dataset



Global Environment

Files

Plots

Packages

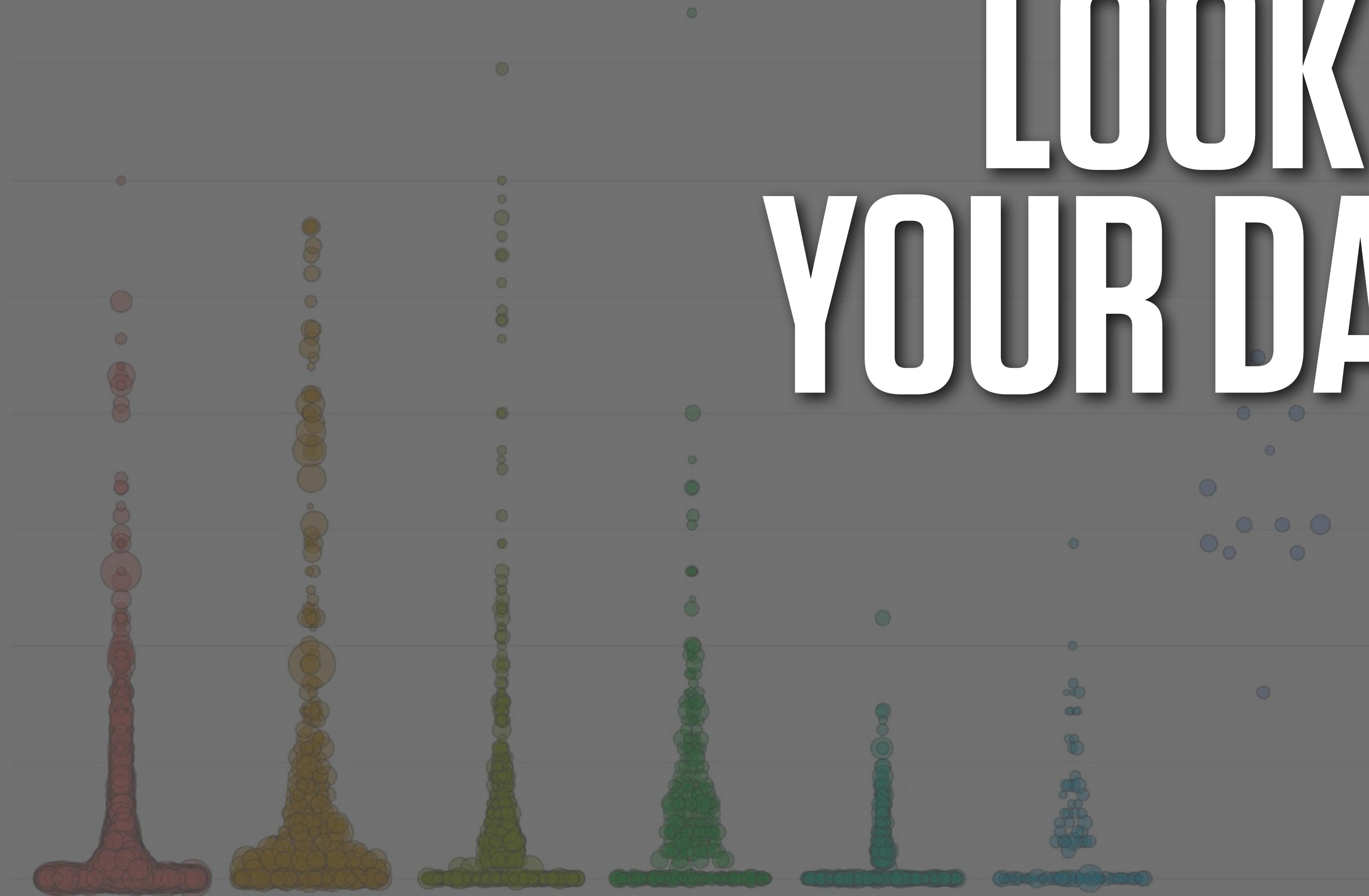
Help

New Folder

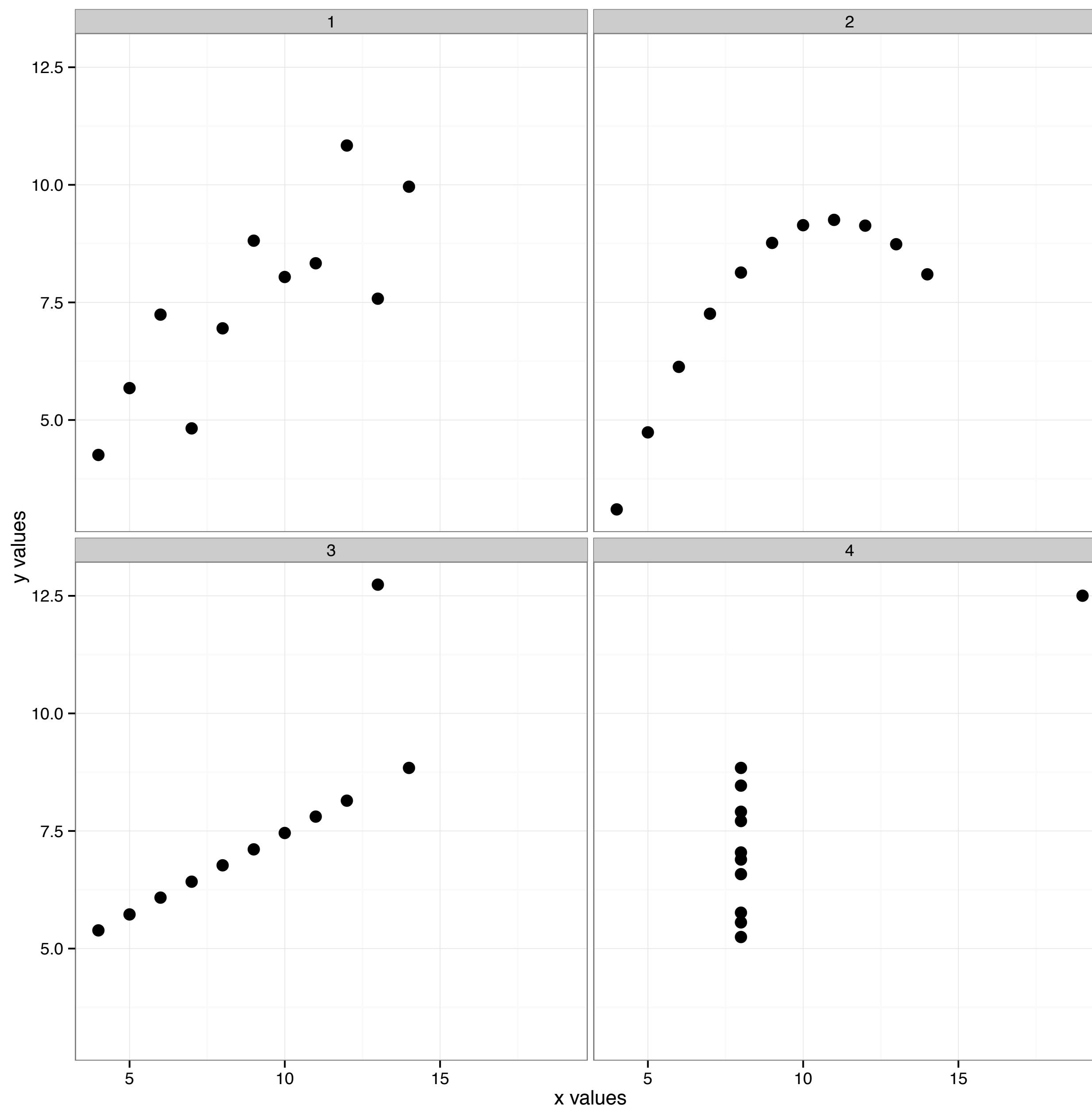
Delete

Rer

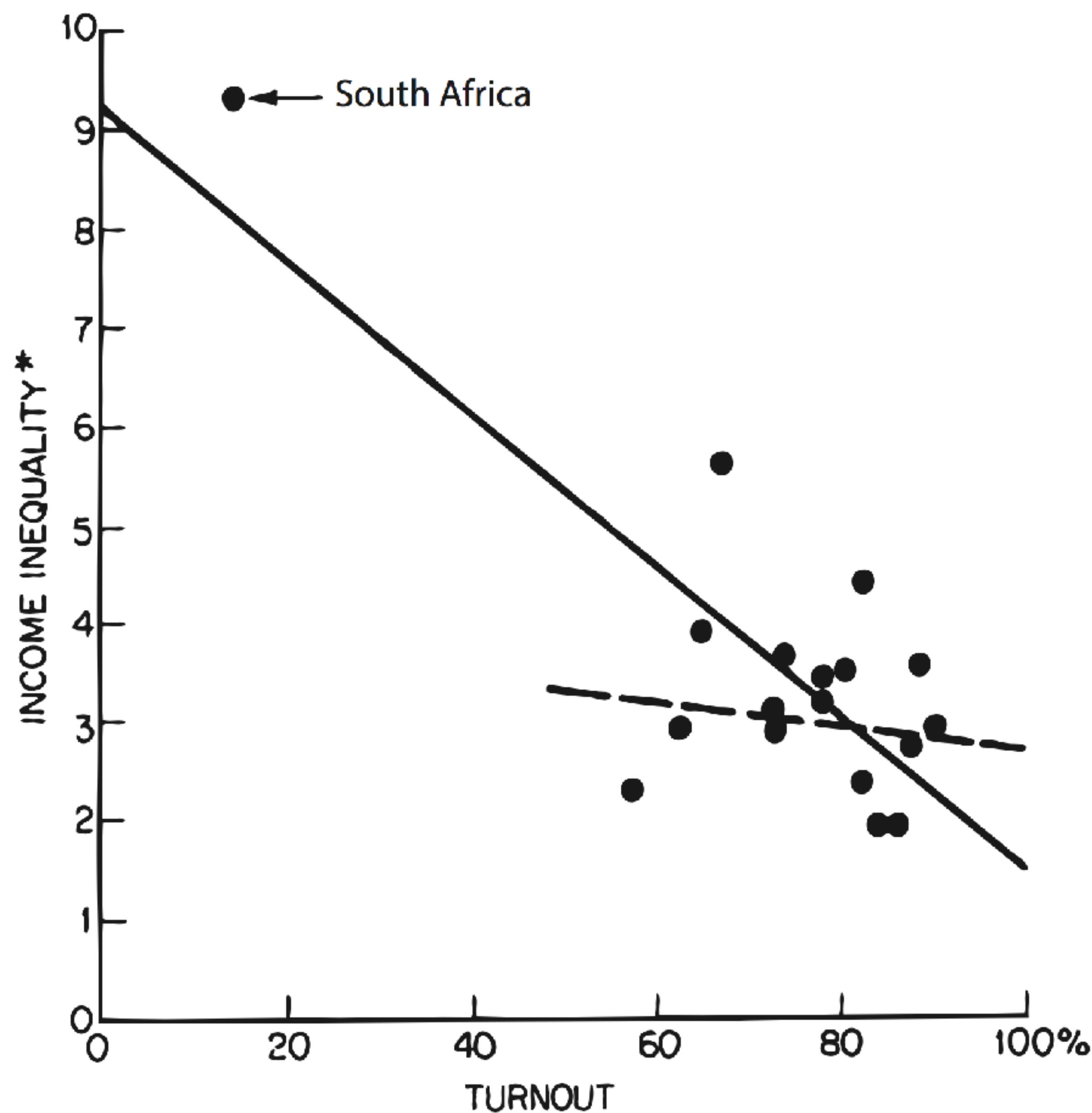
LOOK AT YOUR DATA



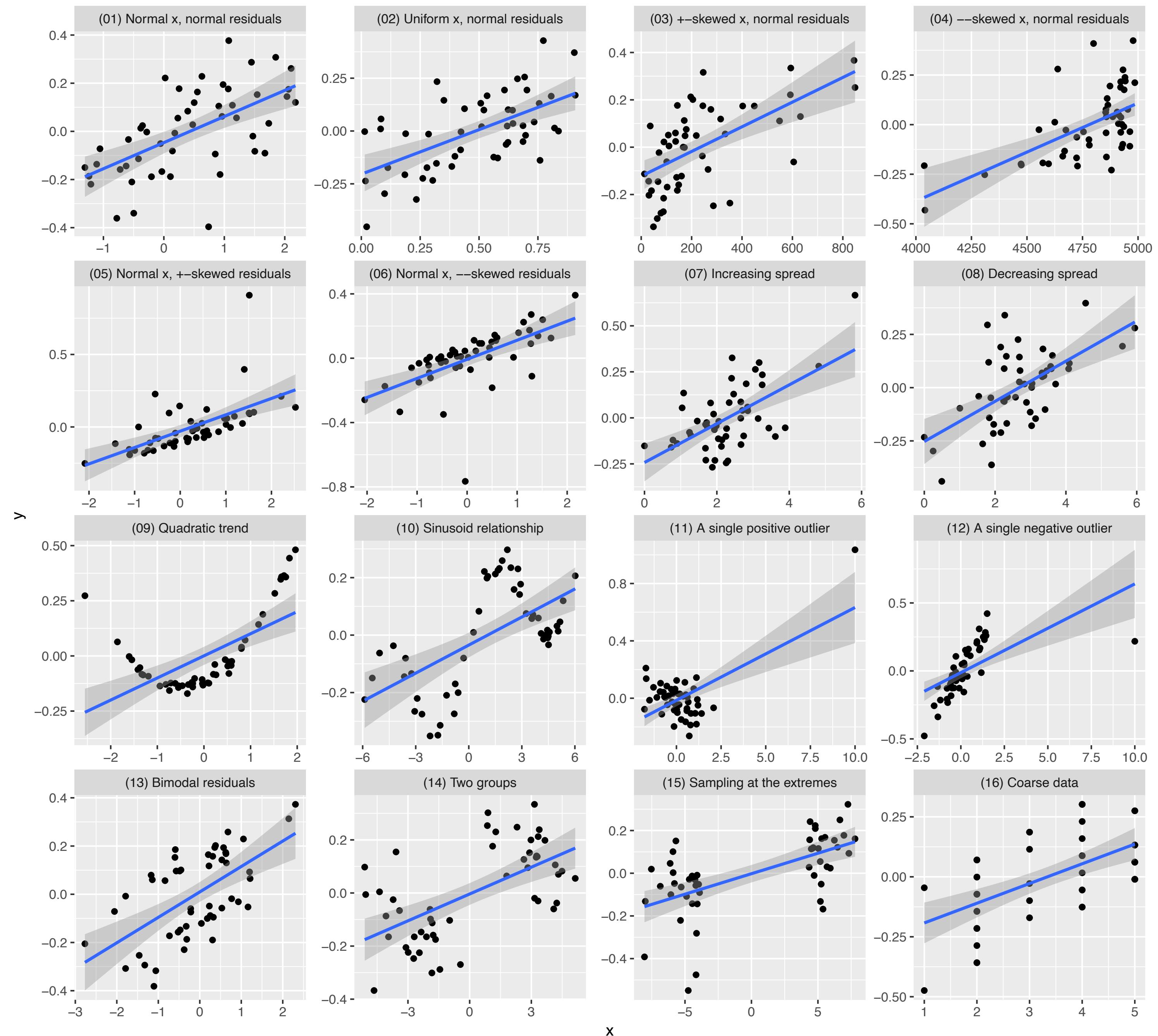
SEEING THINGS

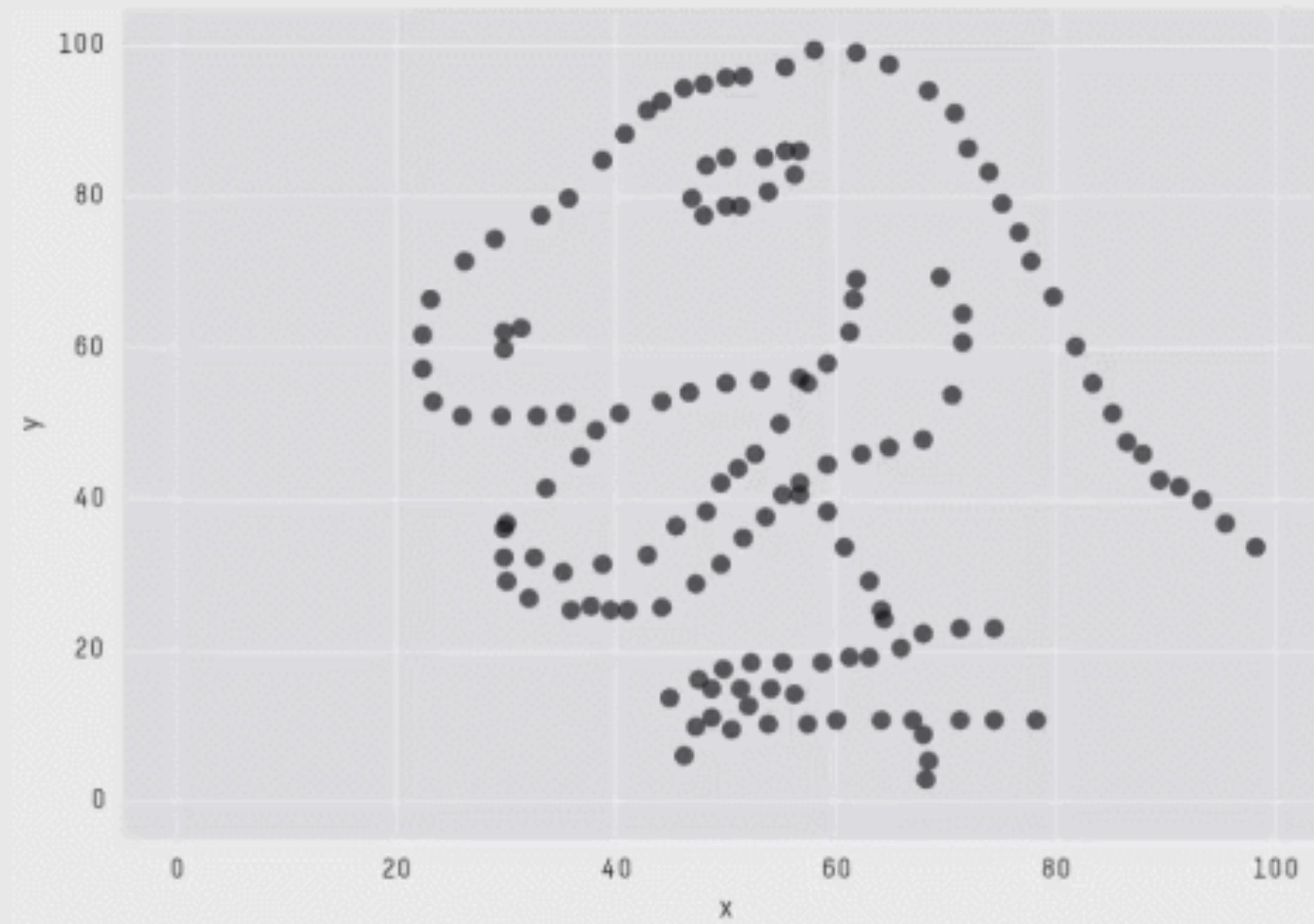


For all panels, N=11; Mean=7.5; Regression: $Y=3 + 0.5(X)$; SE of slope estimate: 0.118, t=4.24; Sum of Squares (X-X): 100; r=0.82.



Key. — Bivariate slope including South Africa ($N = 18$)
— Bivariate slope excluding South Africa ($N = 17$)

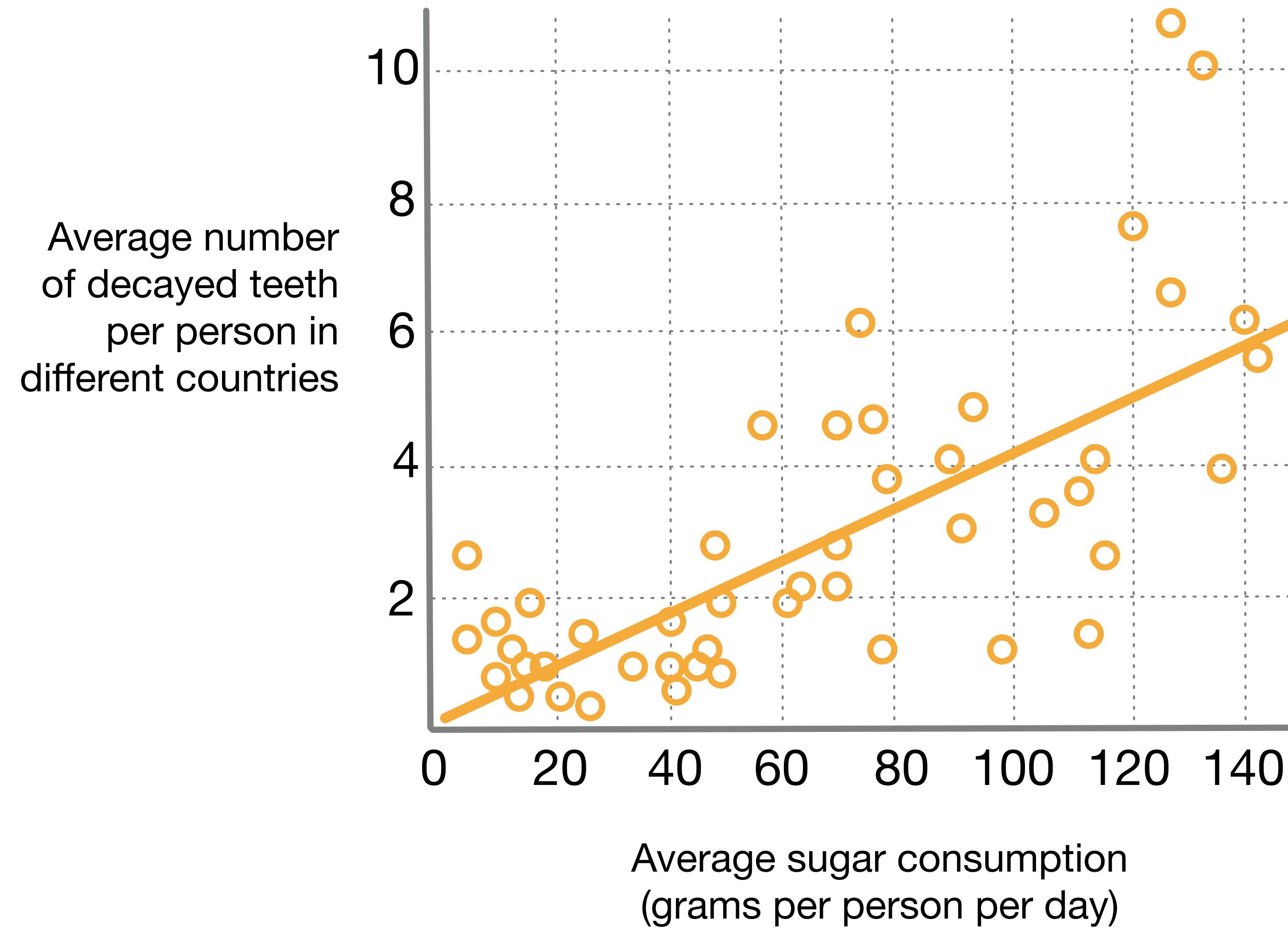




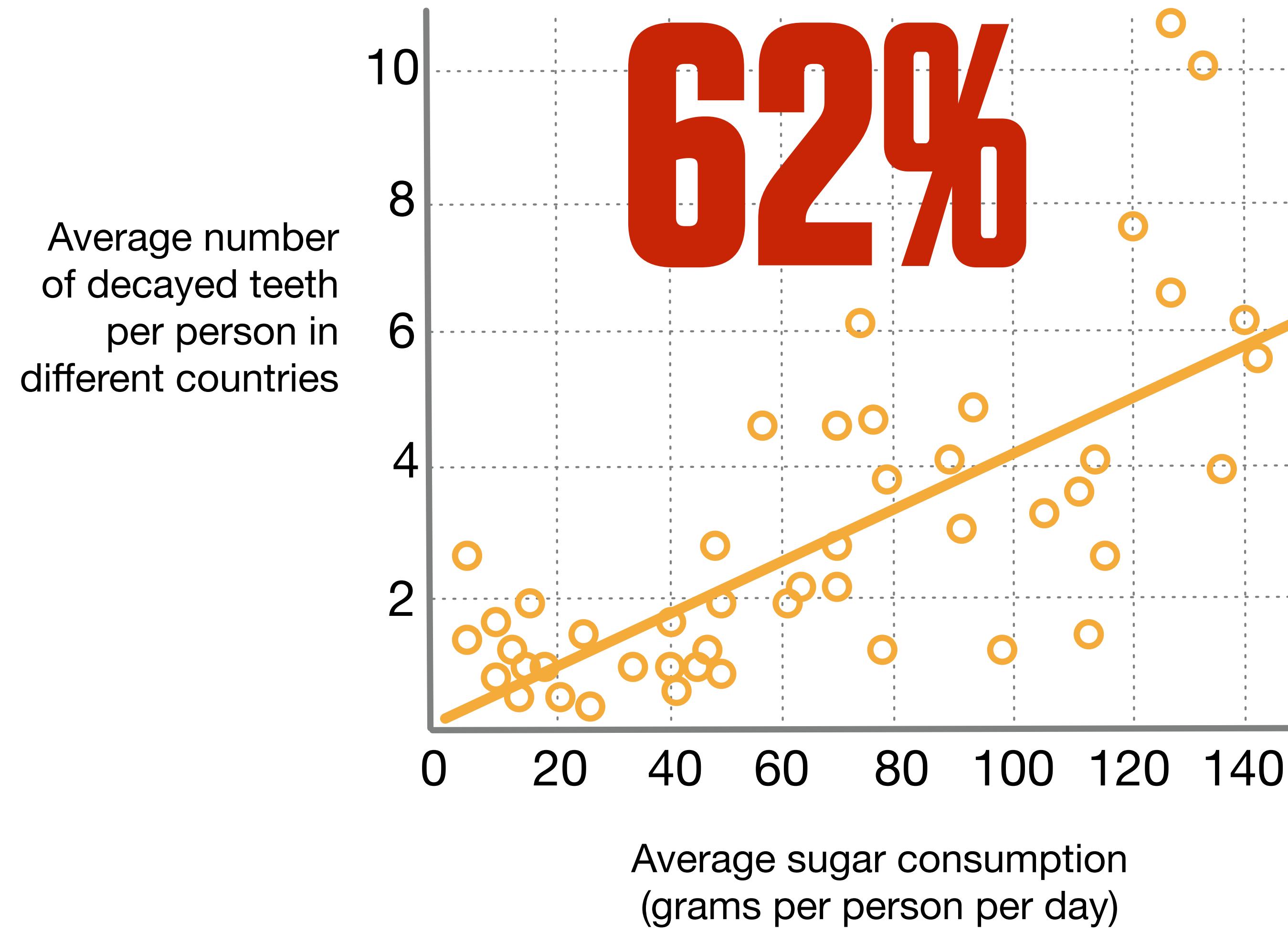
X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Cairo; Matejka & Fitzmaurice

Which of the following statements best describes the data in the graph below?



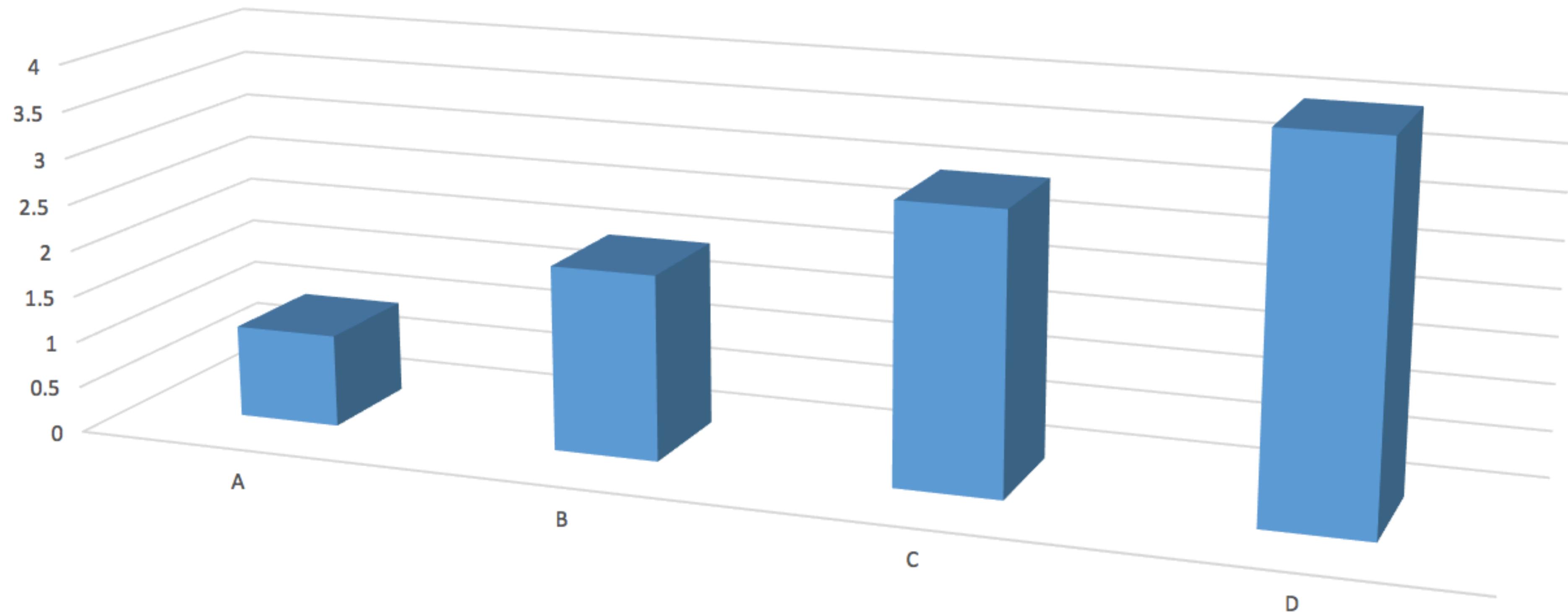
Which of the following statements best describes the data in the graph below?

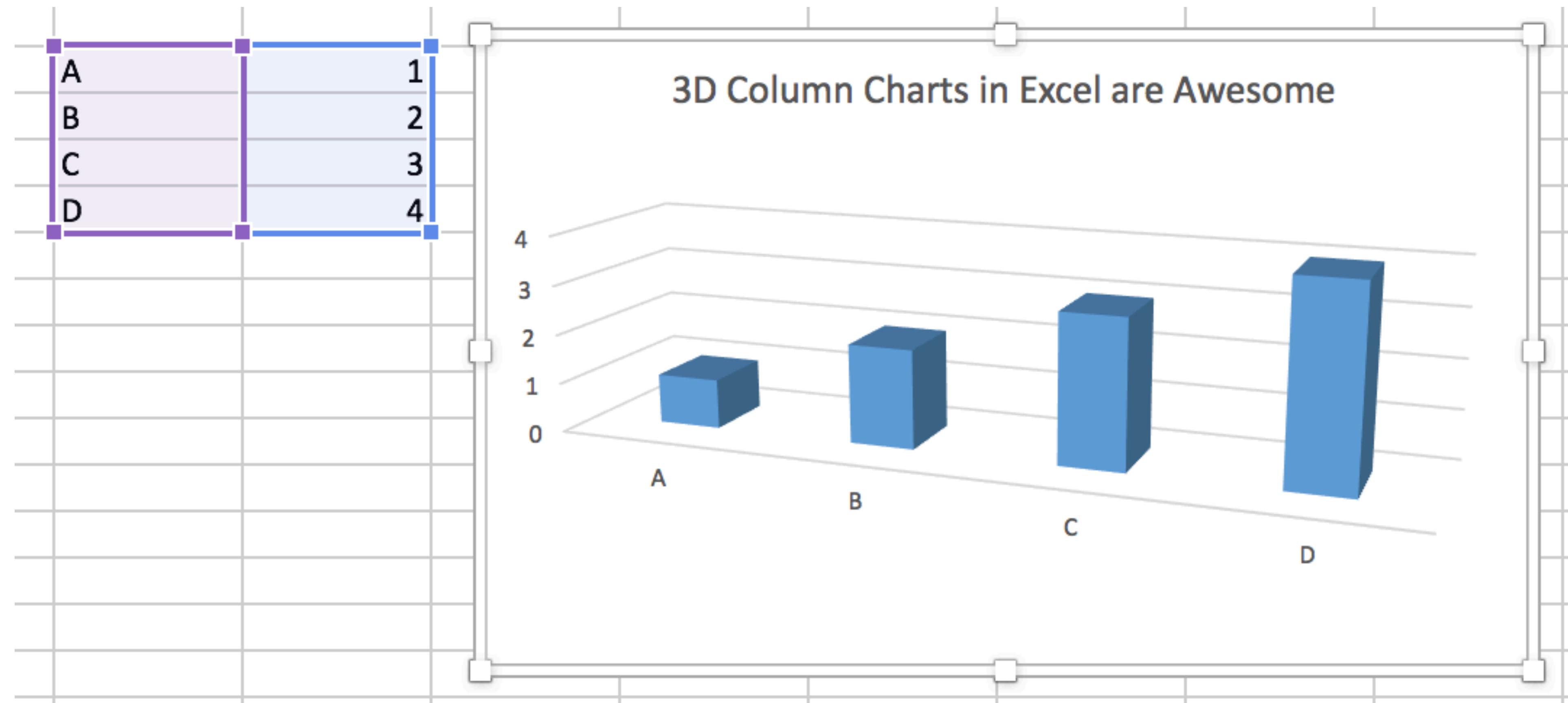


- A. In recent years, the rate of cavities has increased in many countries
- B. In some countries, people brush their teeth more frequently than in other countries
- C. The more sugar people eat, the more likely they are to get cavities**
- D. In recent years, the consumption of sugar has increased in many countries

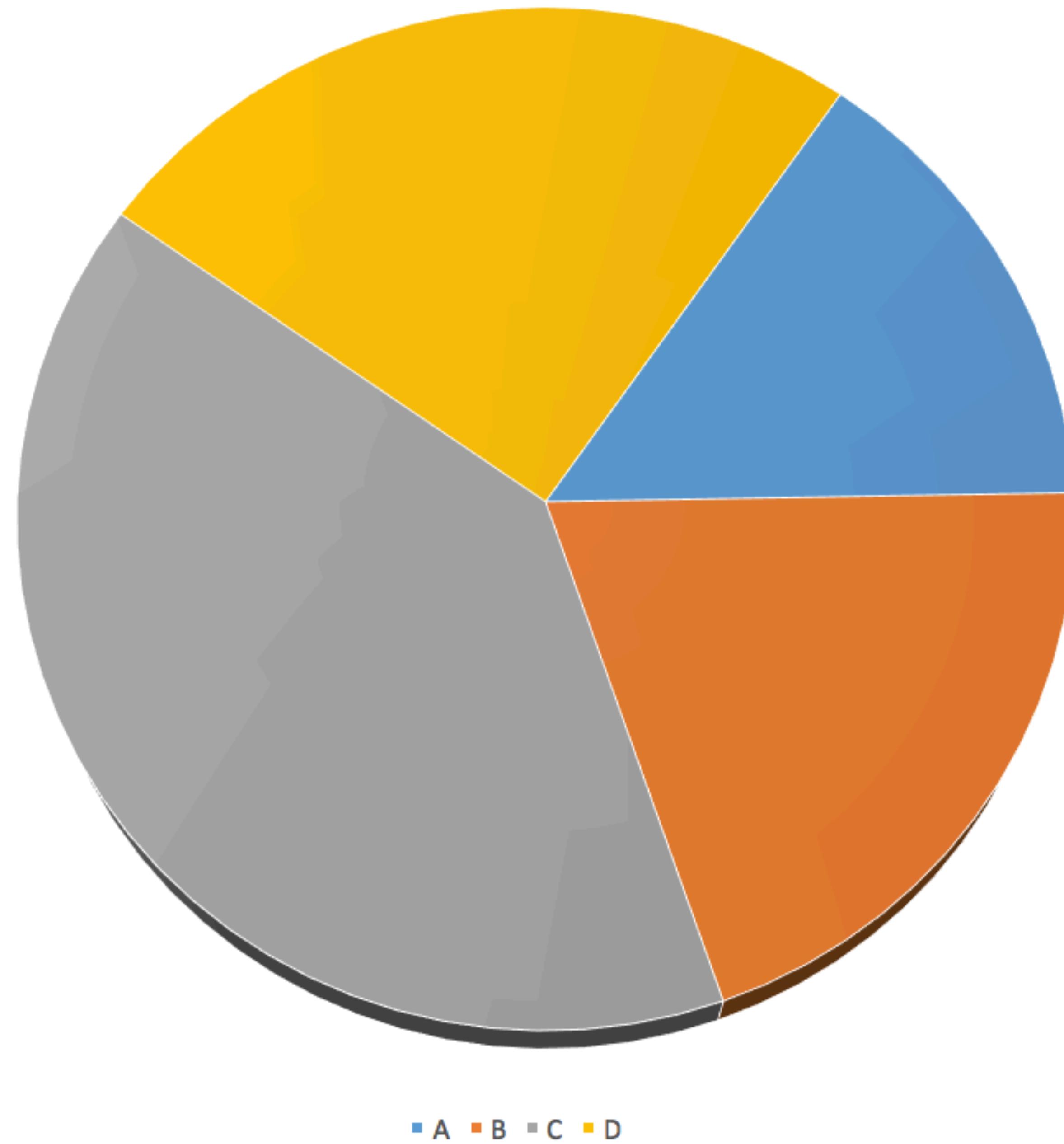
NOT
SEEING THINGS

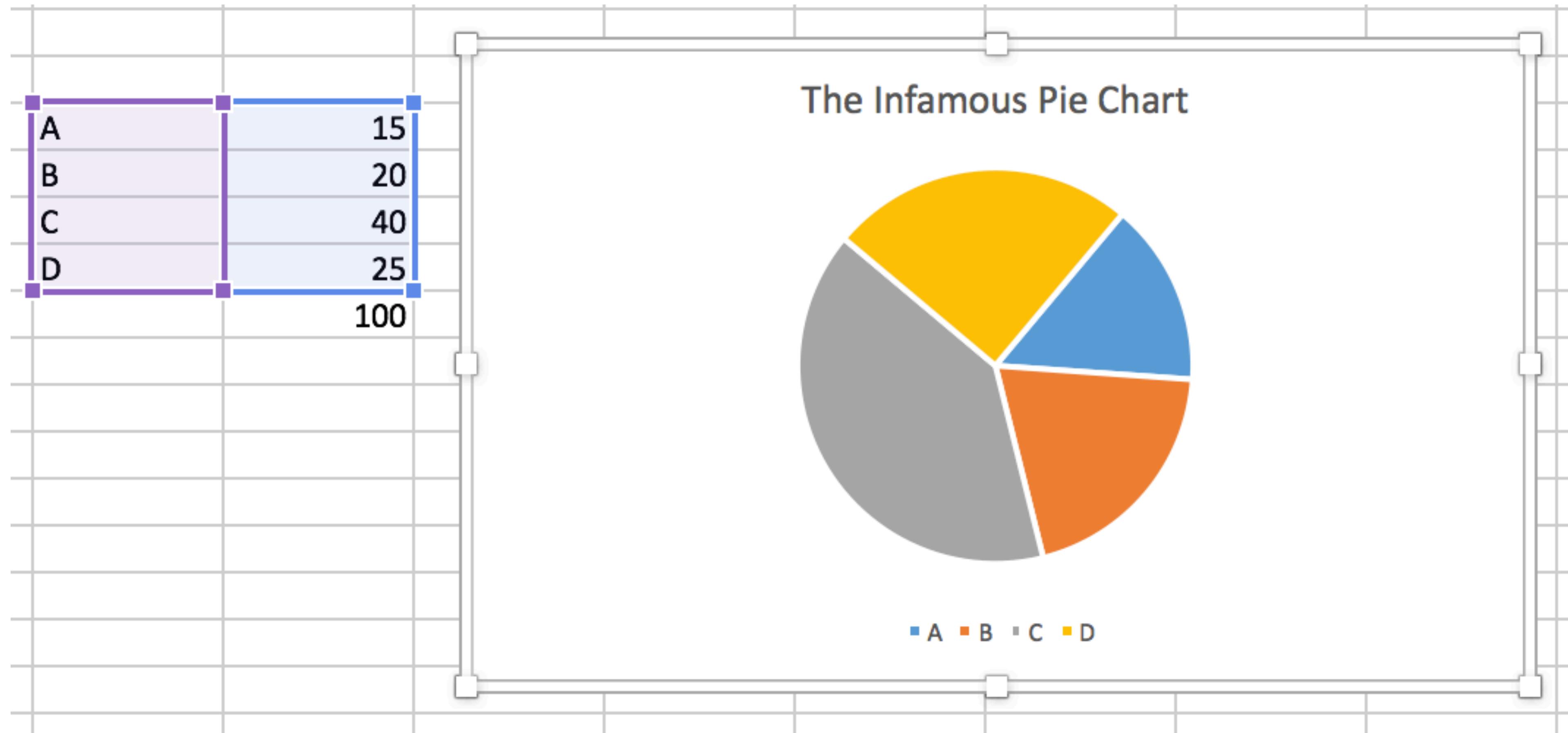
3D Column Charts in Excel are Awesome





The Infamous Pie Chart





BAD TASTE

BAD DATA

BAD PERCEPTION

TASTE

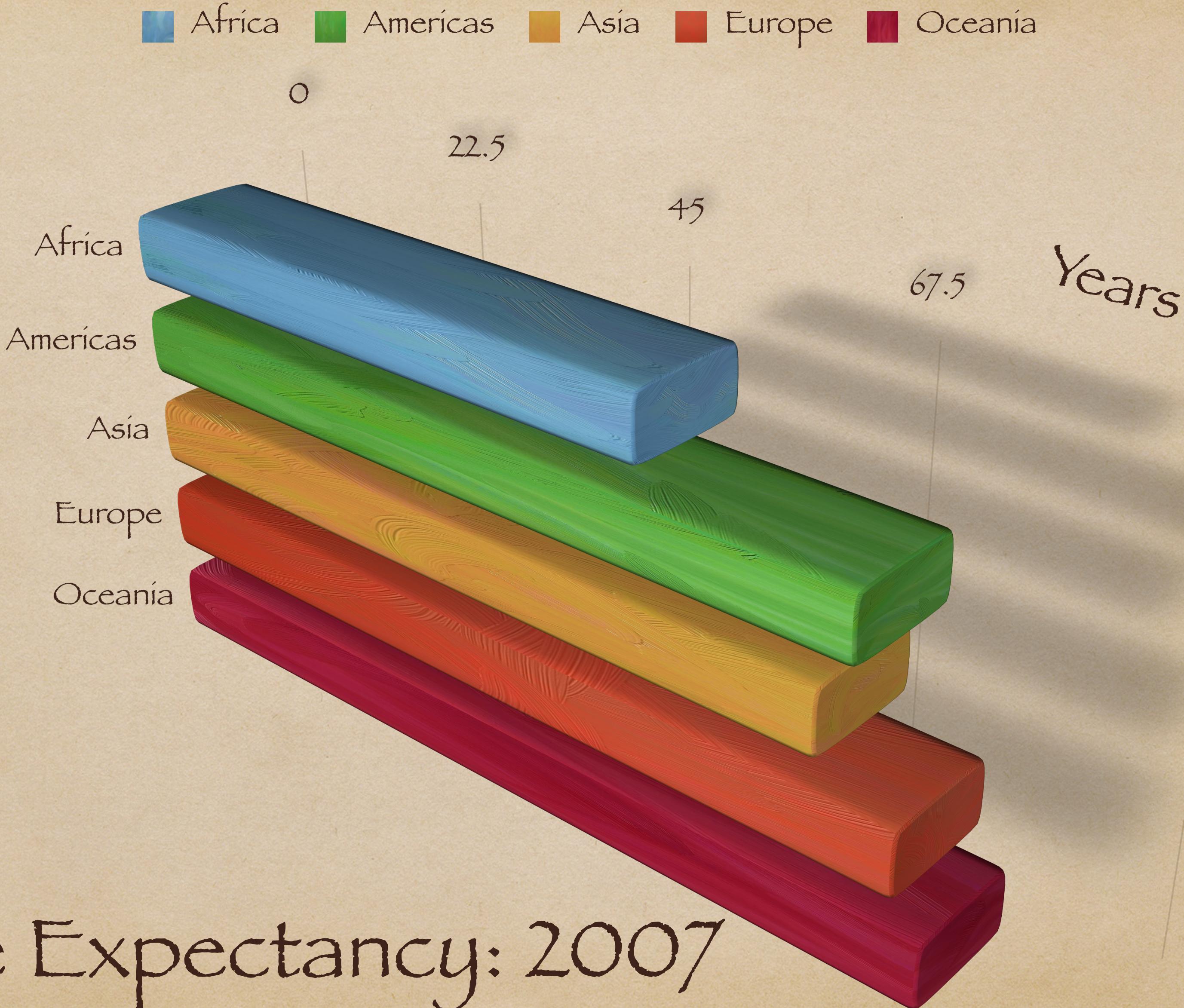
SIMPLIFY, SIMPLIFY?

The Visual Display
of Quantitative Information

EDWARD R. TUFTE

Life Expectancy: 2007

CONTINENT



Years

África Américas Asia Europa Oceania

0

22.5

45

67.5

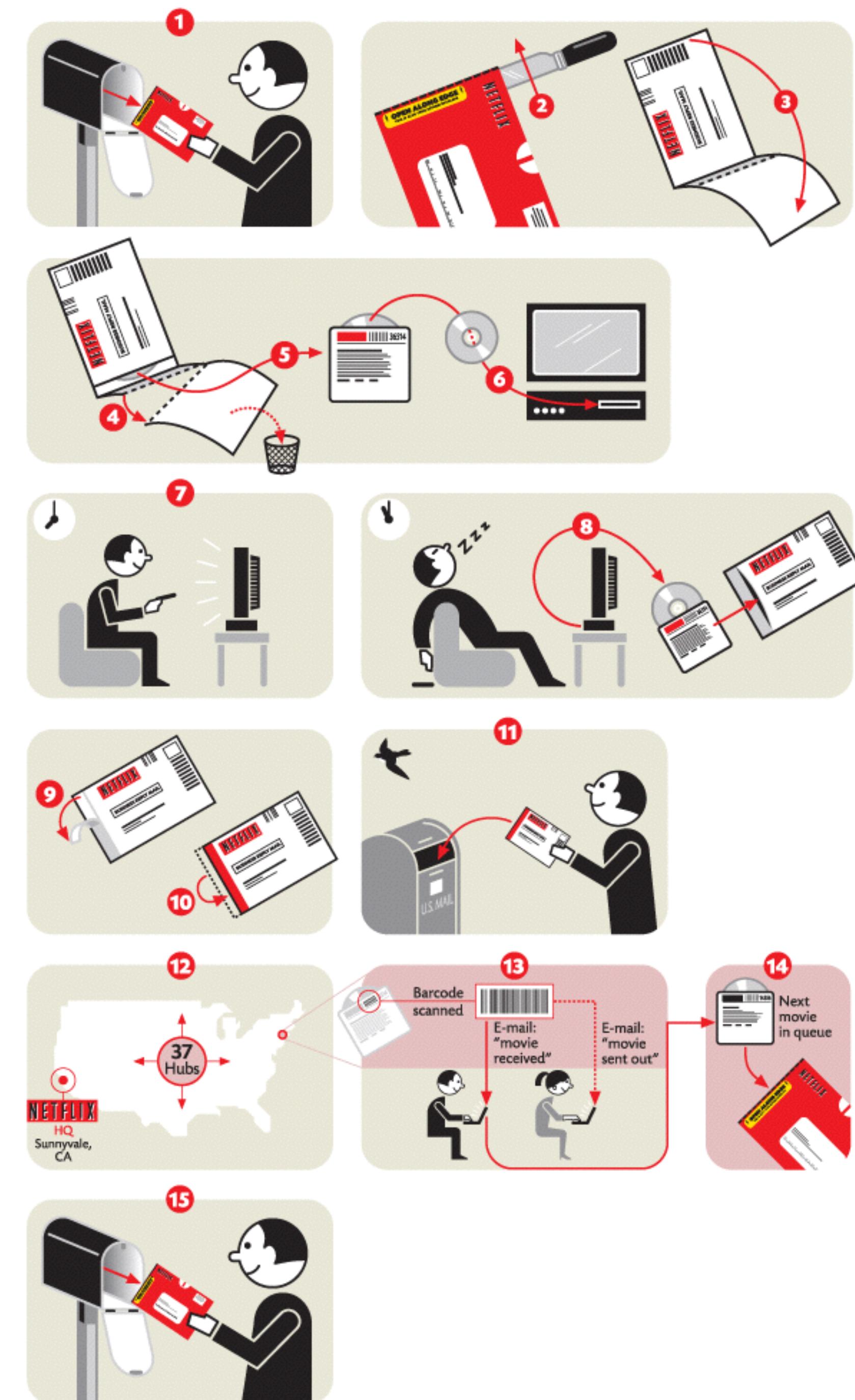
90

MONSTROUS COSTS

Total House and Senate campaign expenditures,
in millions



HOW NETFLIX WORKS

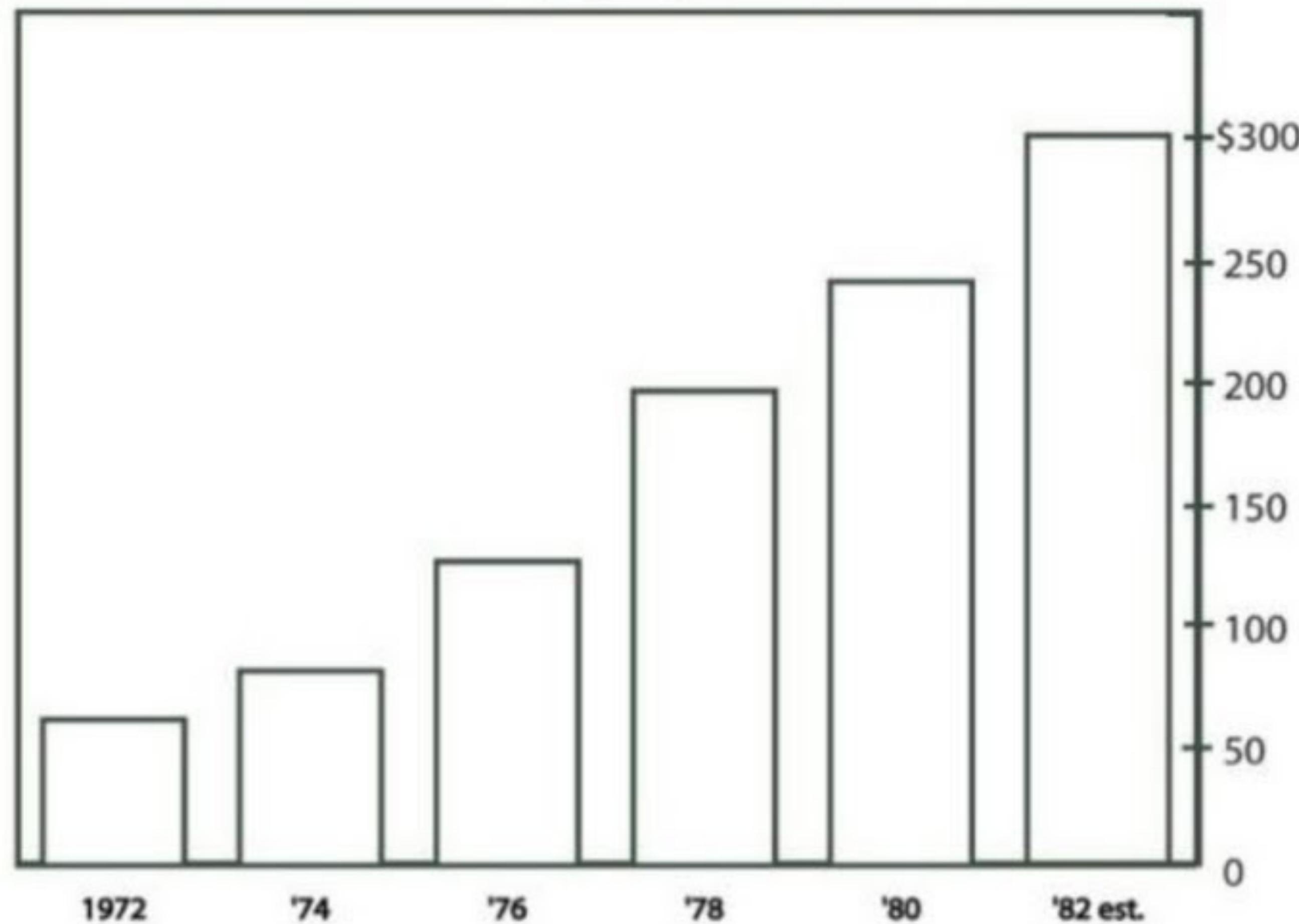


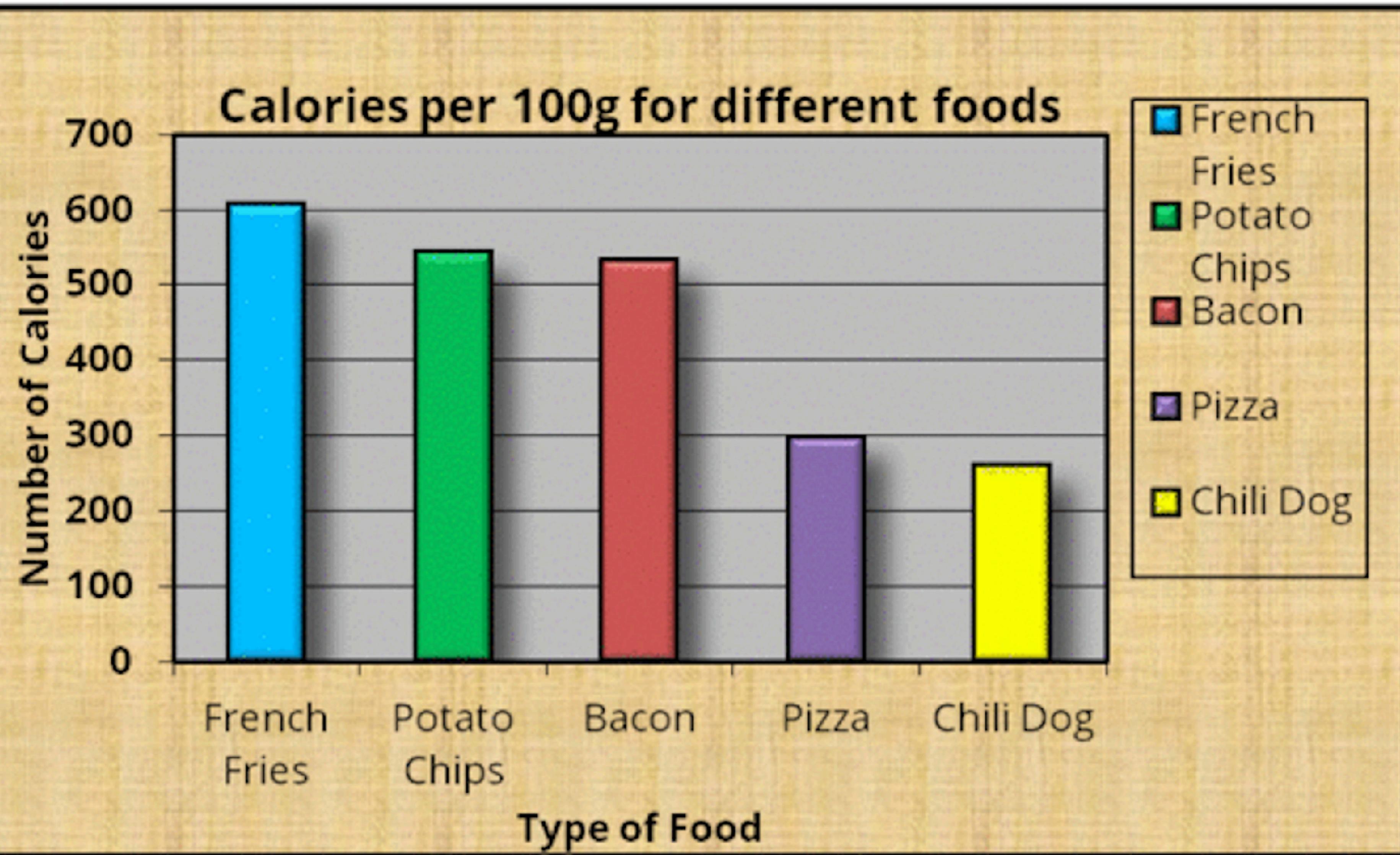
MONSTROUS COSTS

Total House and Senate campaign expenditures,
in millions

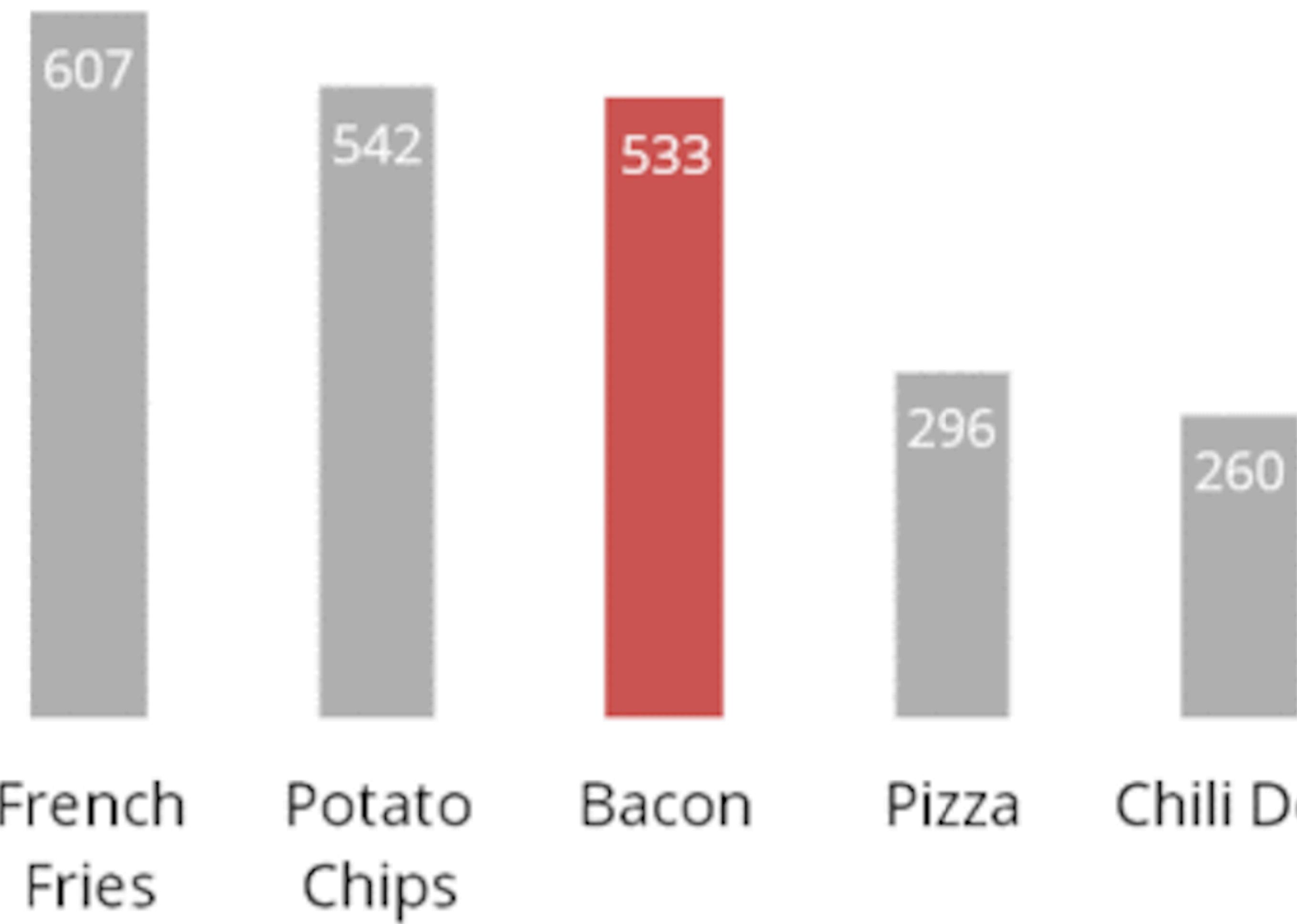


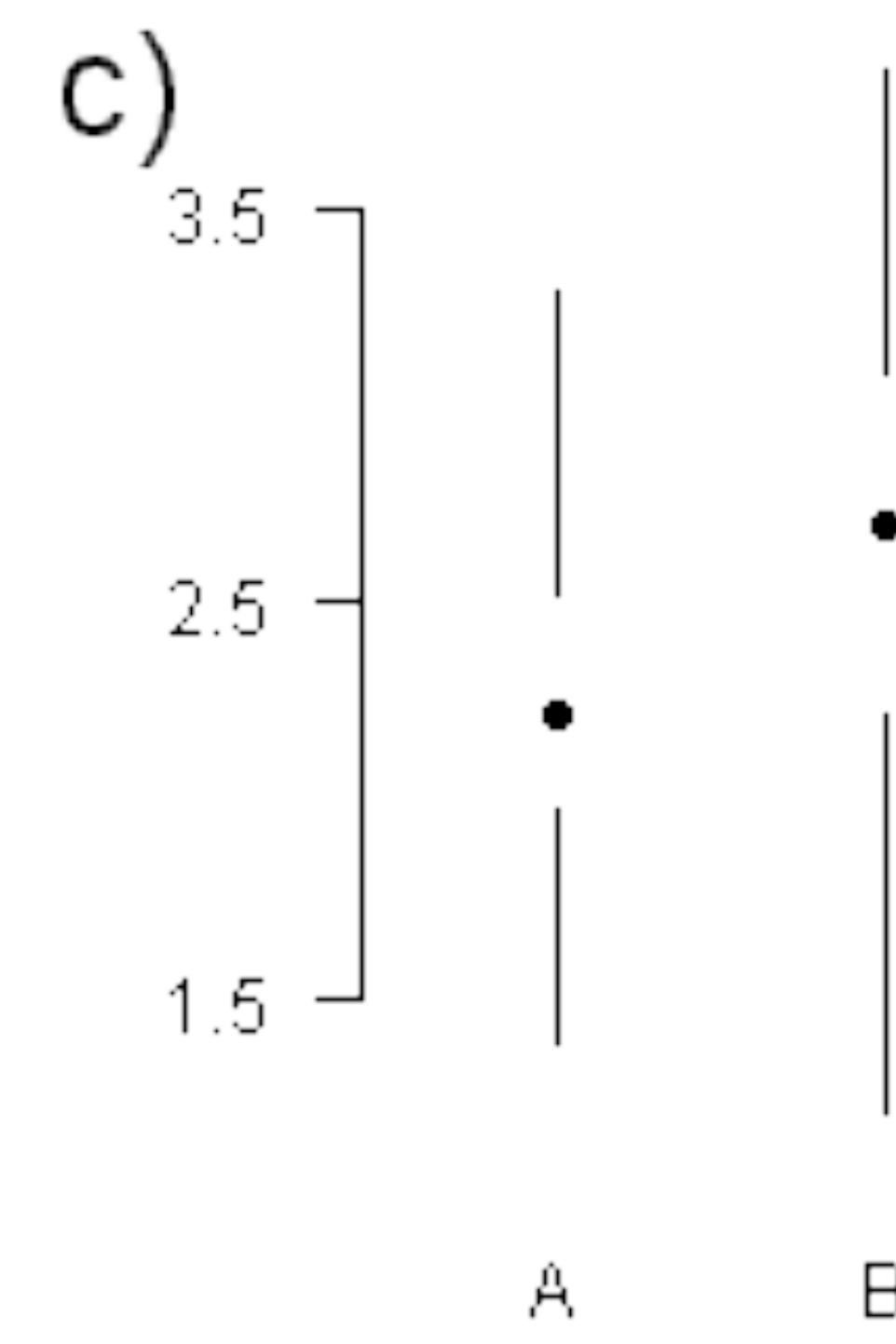
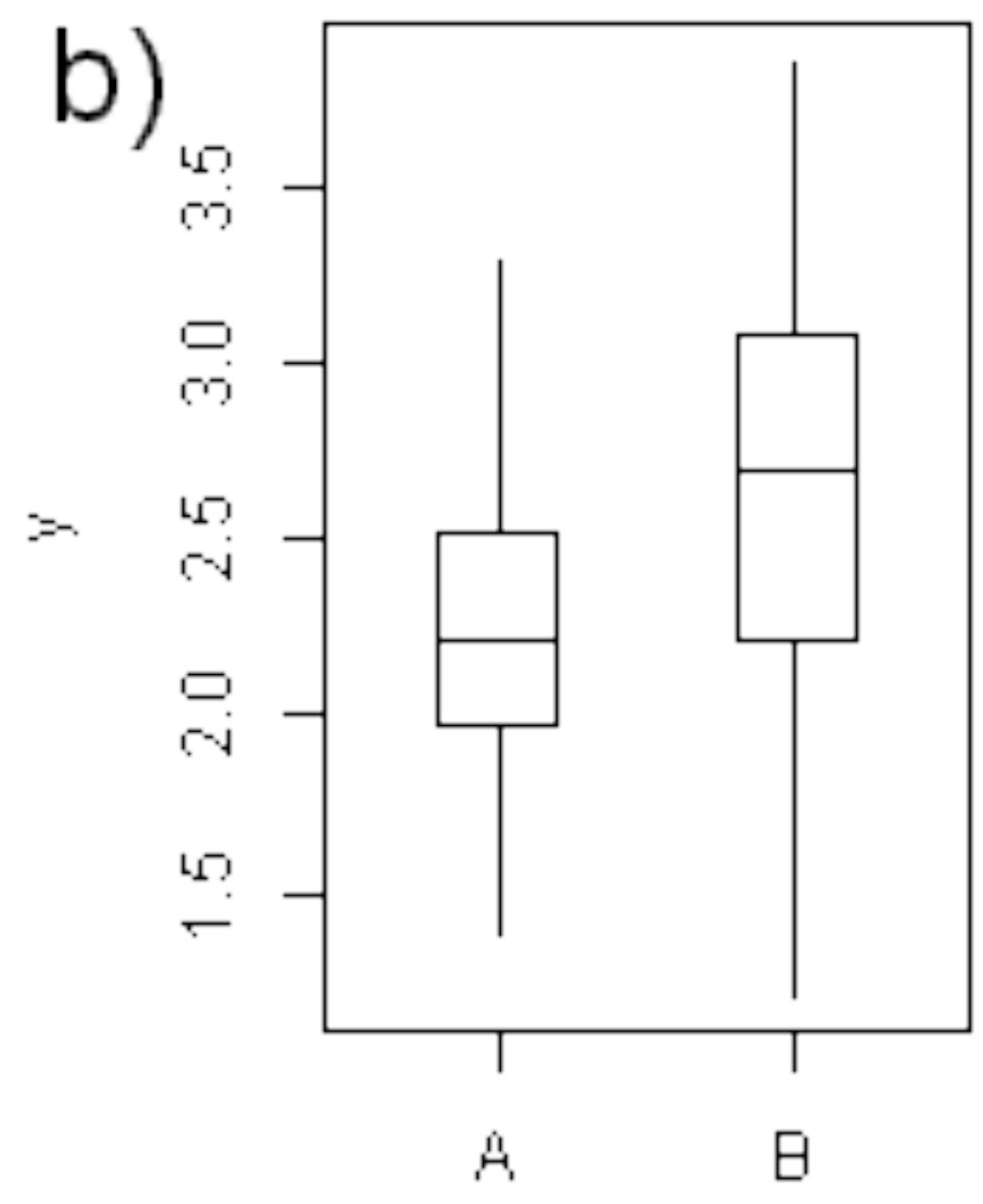
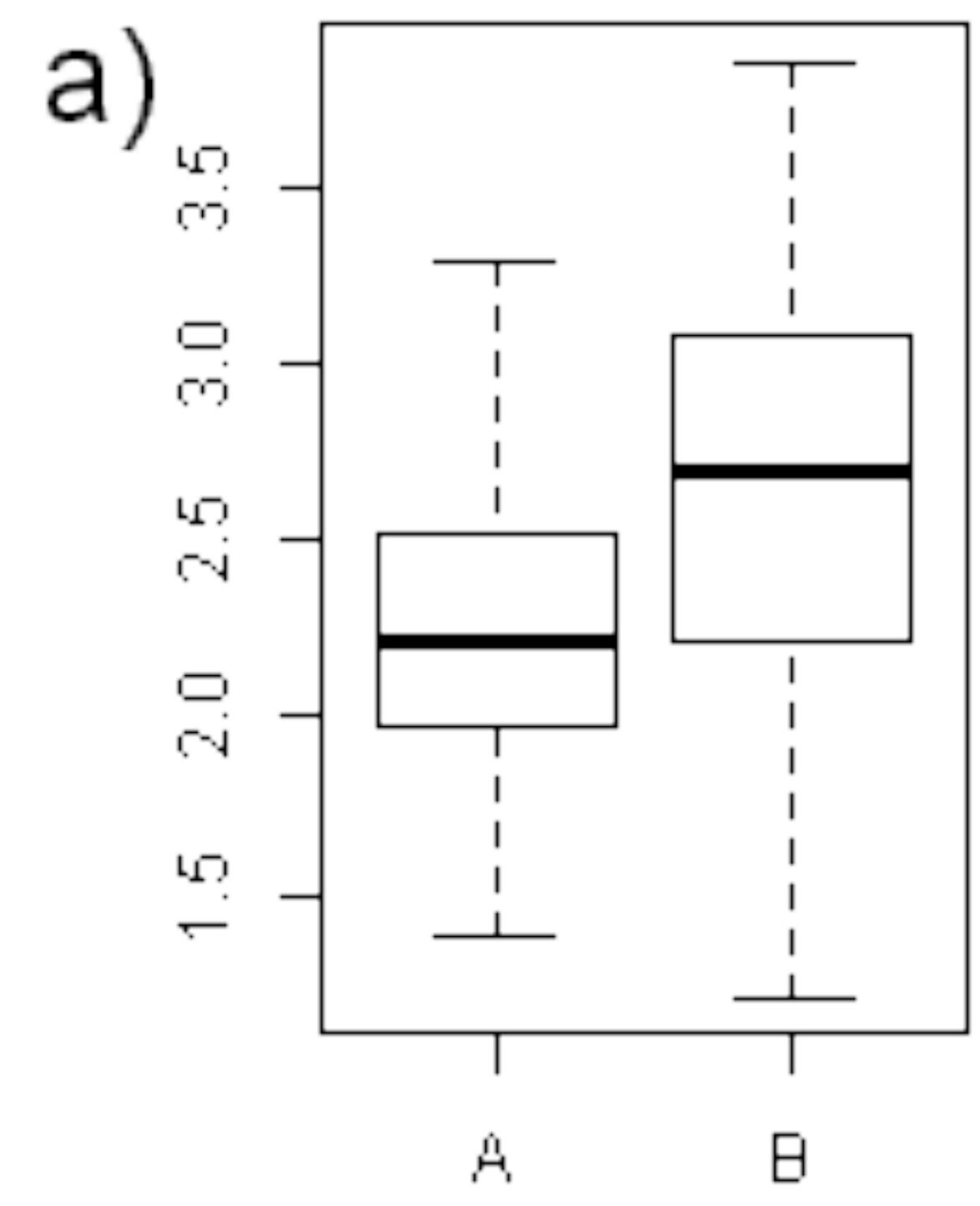
MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions

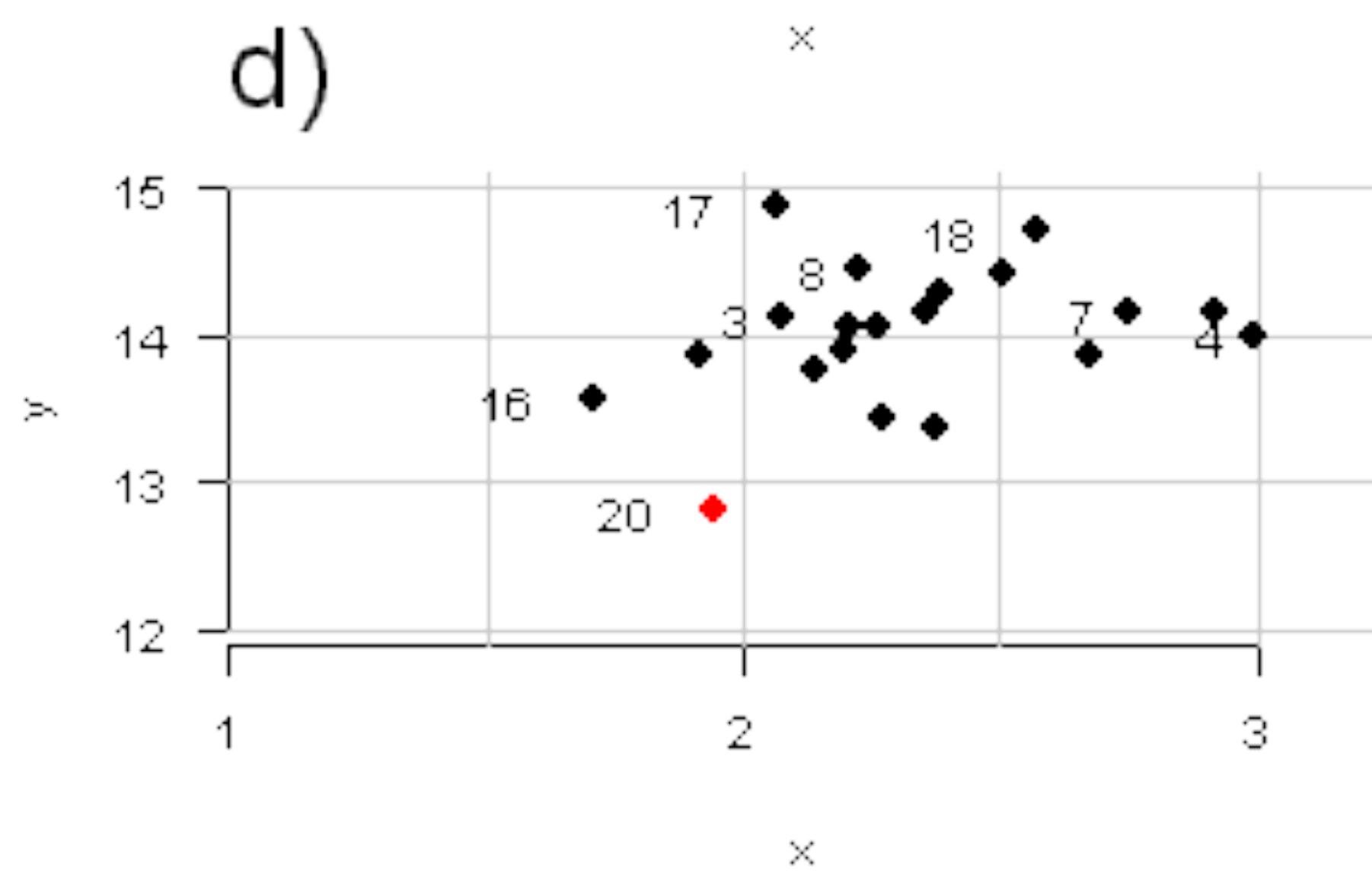
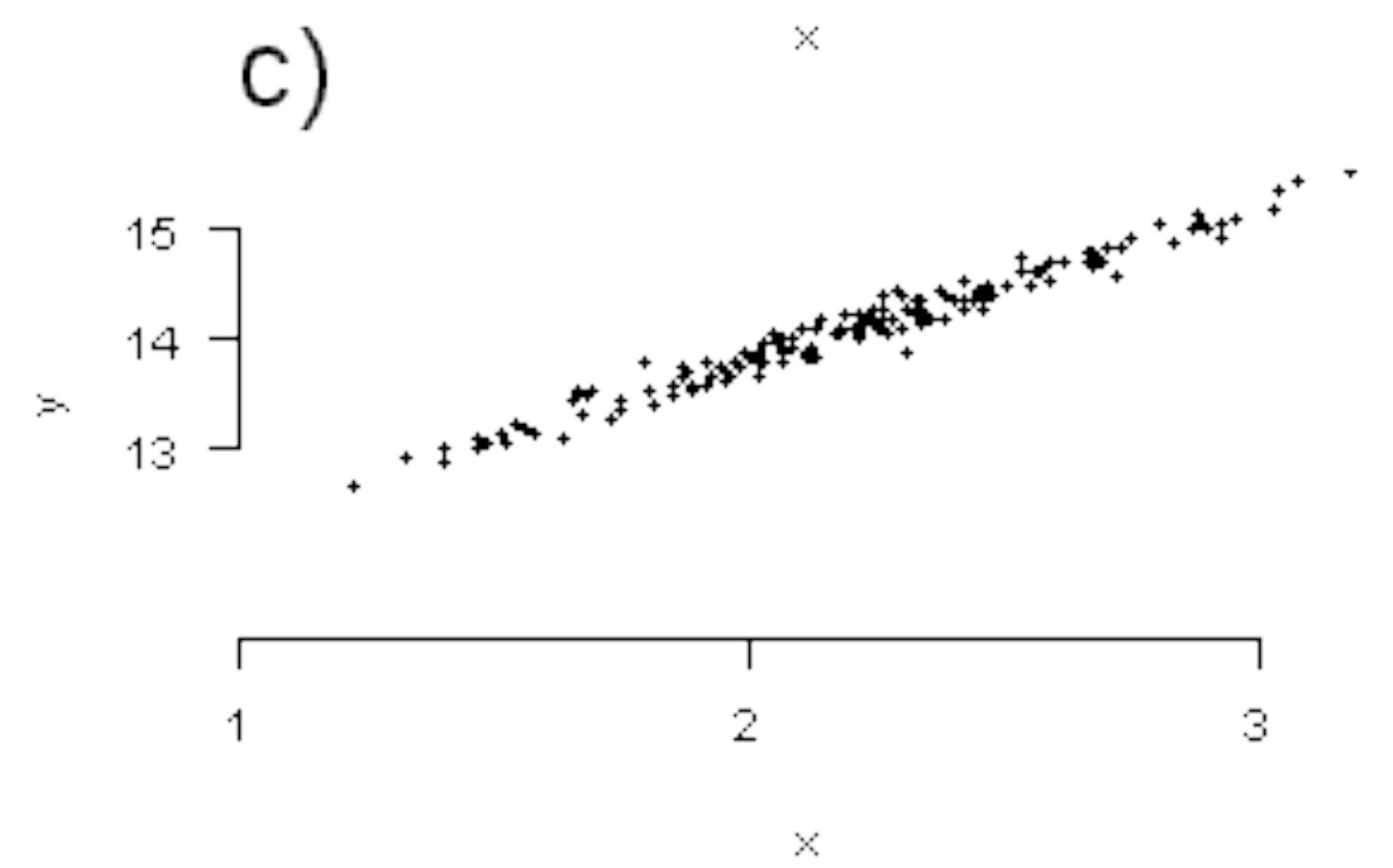
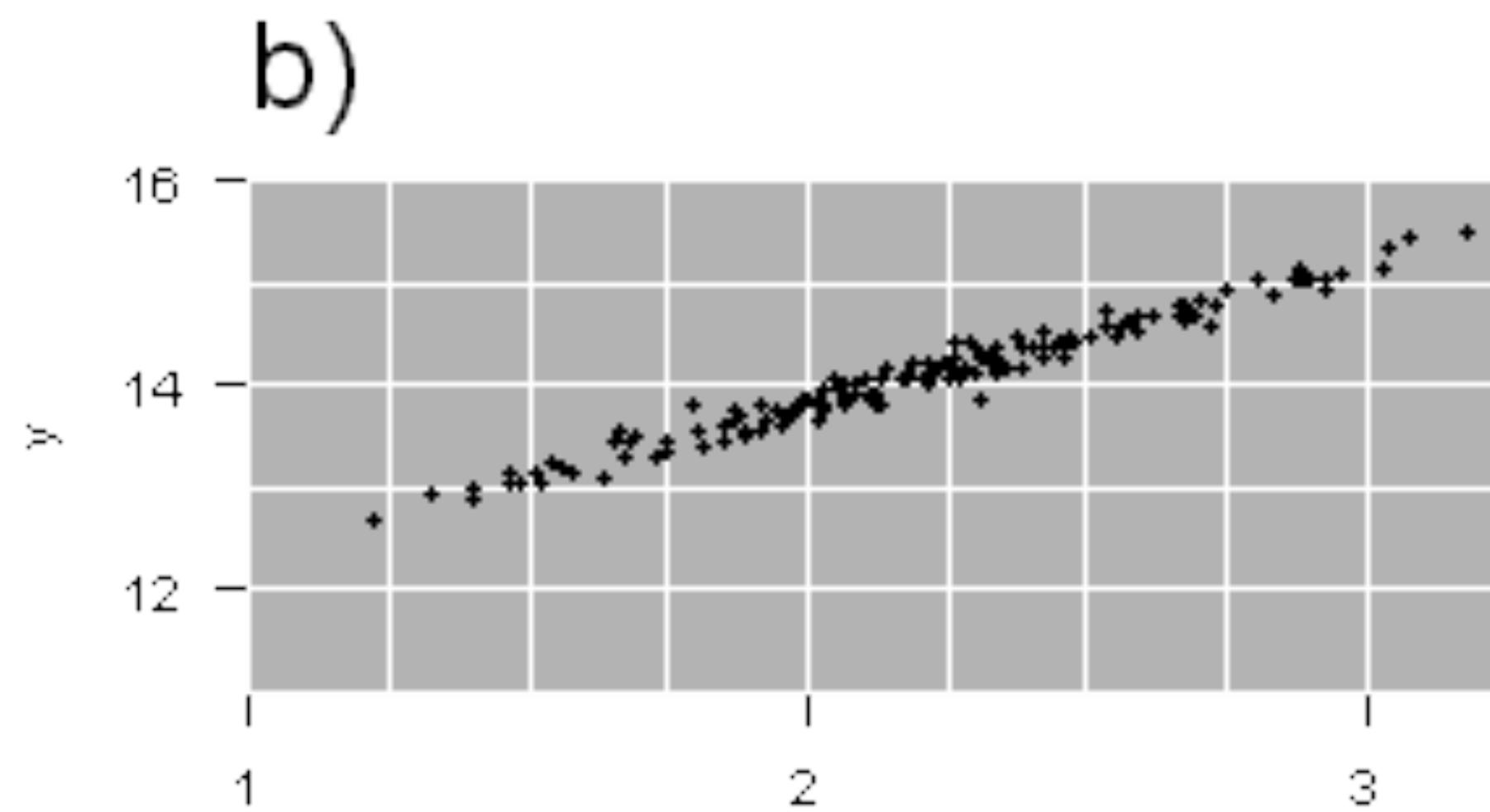
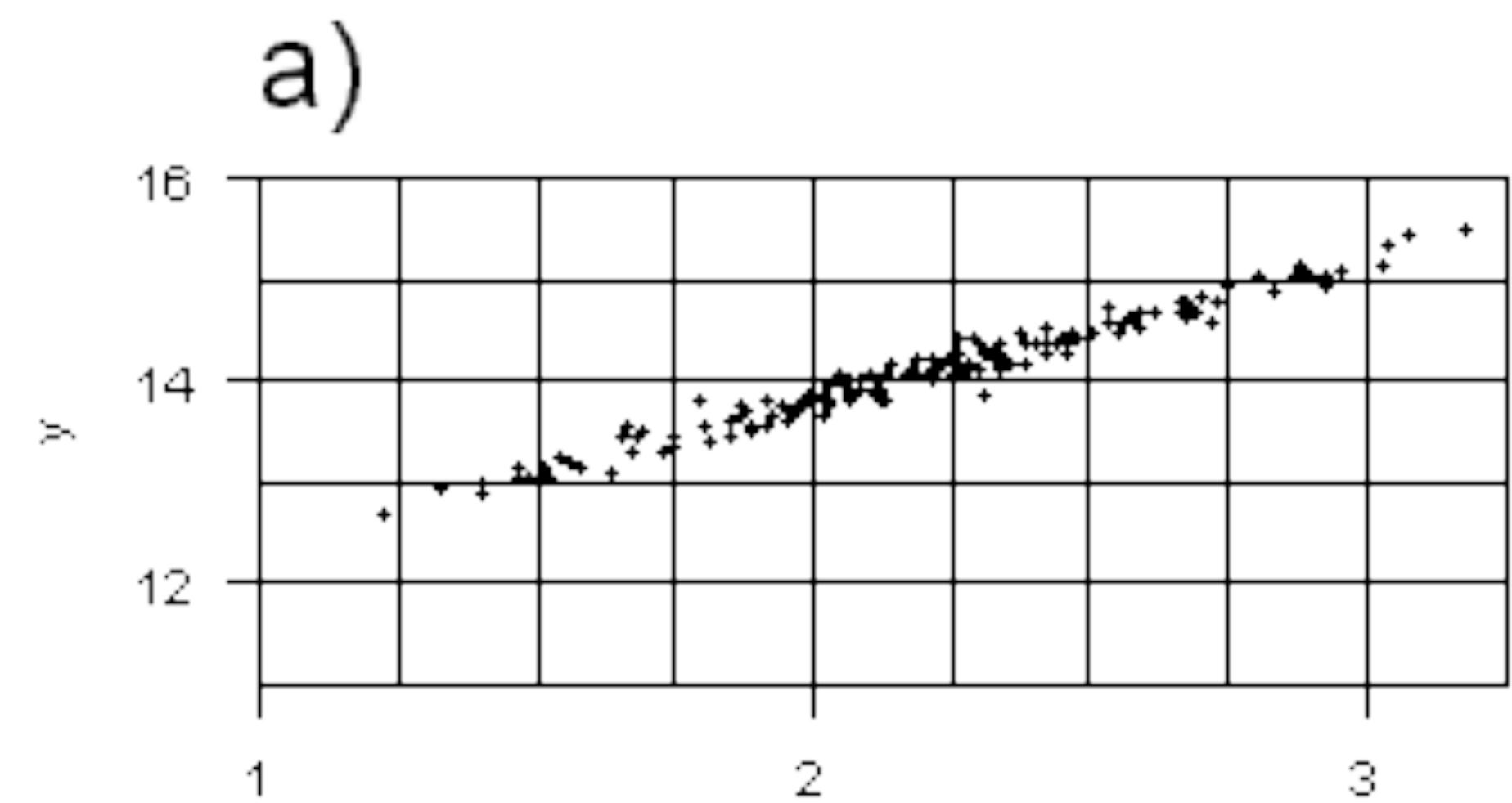




Calories per 100g

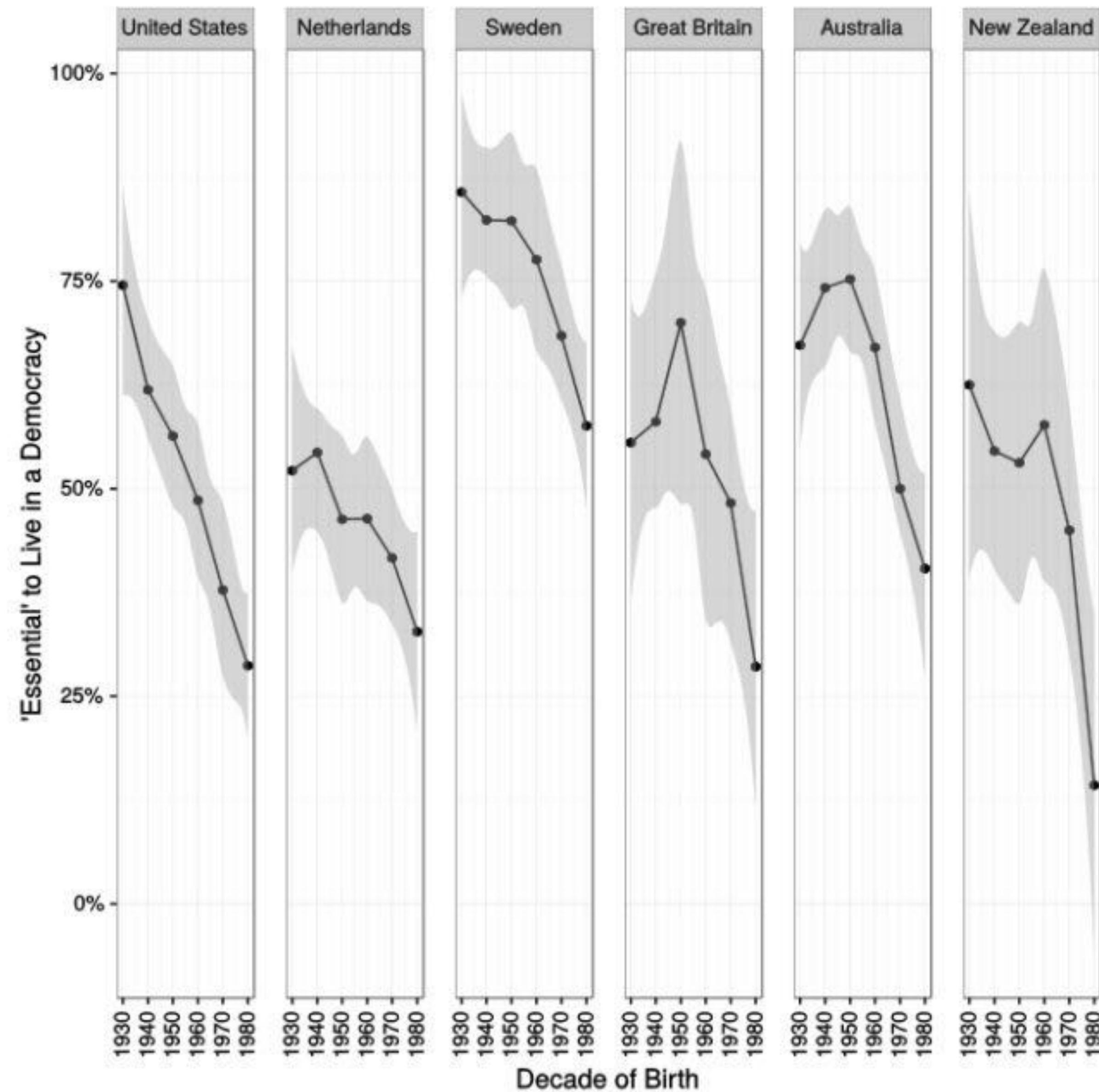




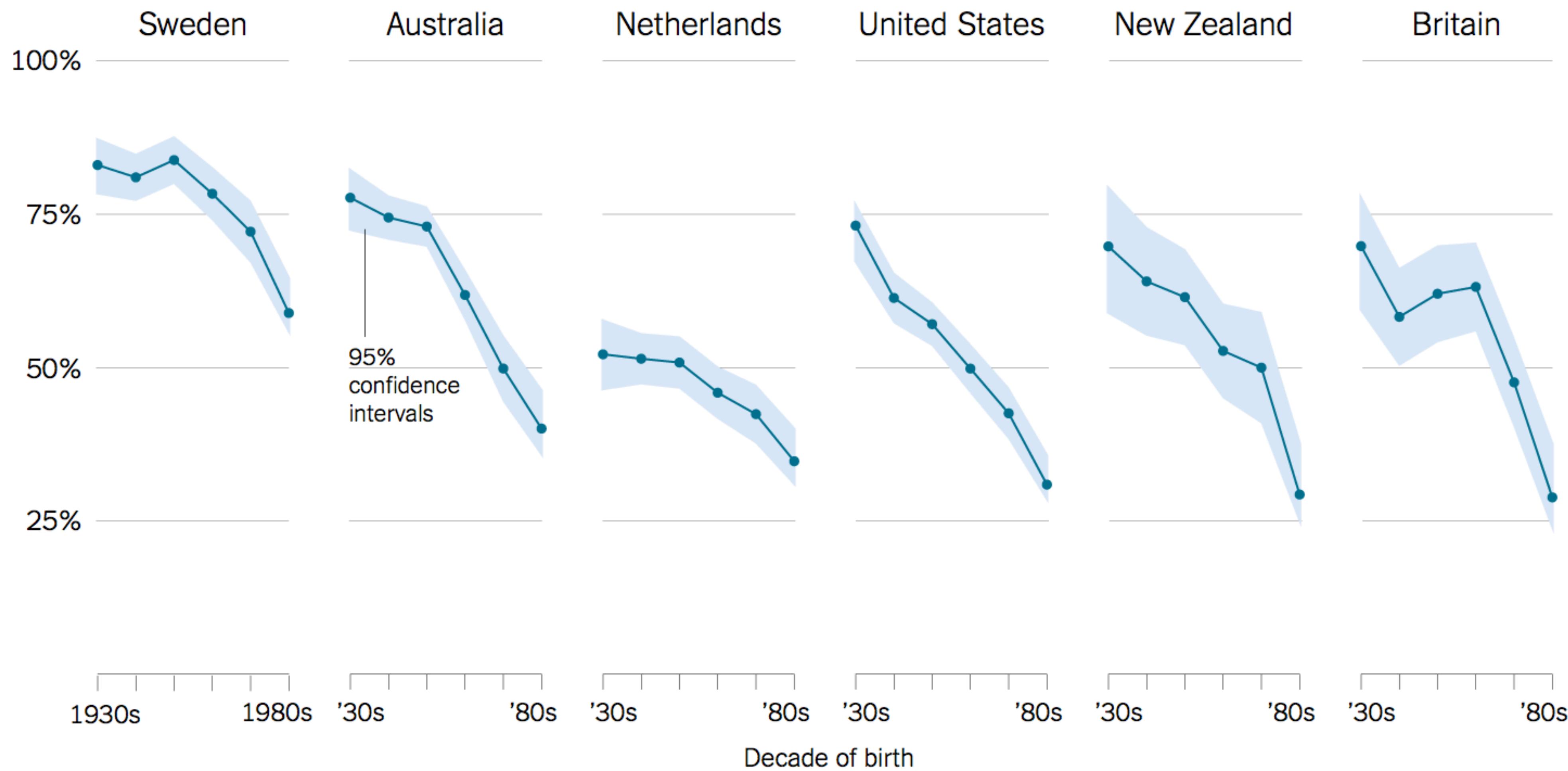


DATA

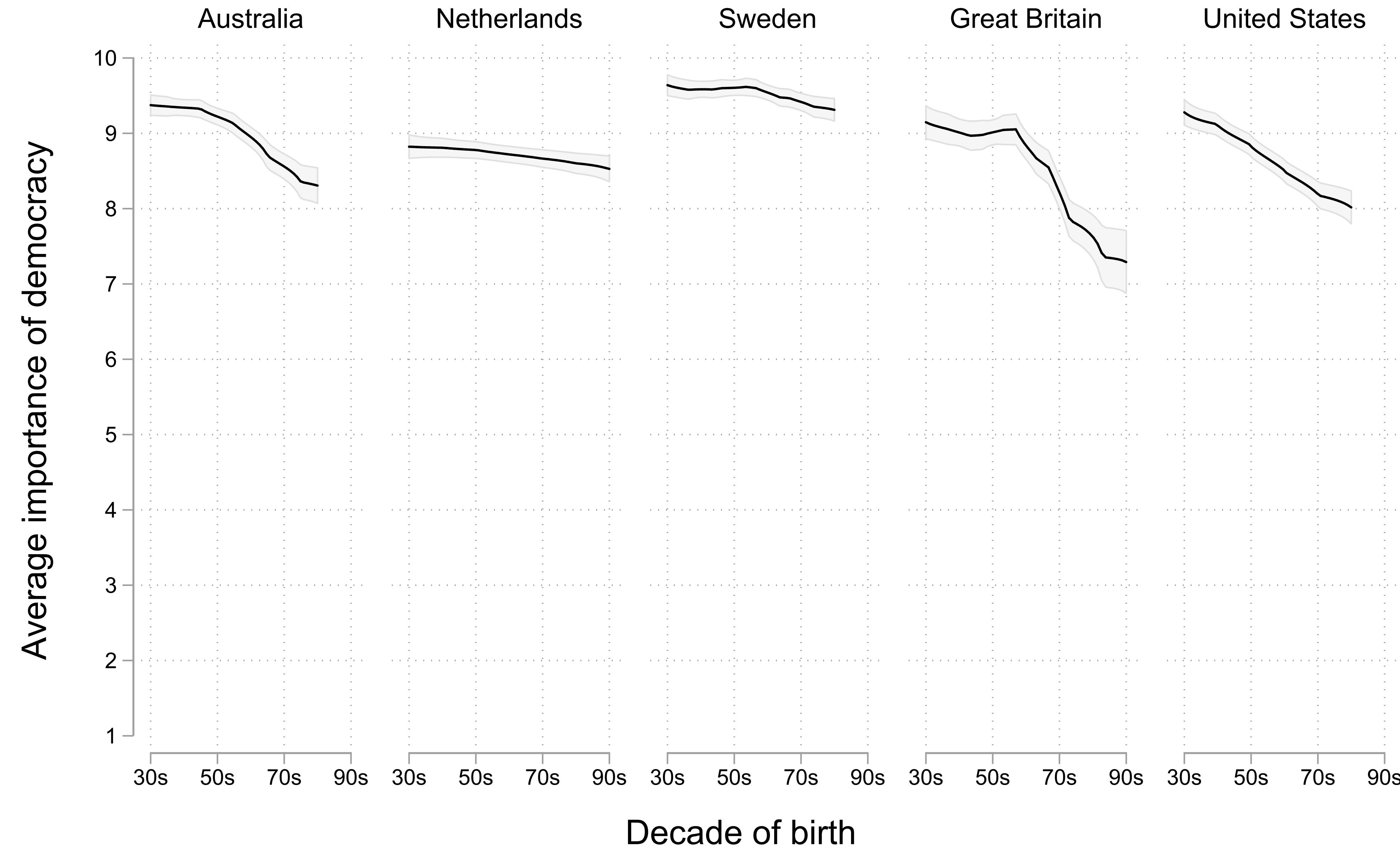
**JUNK-FREE
JUNK CHARTS**



Percentage of people who say it is “essential” to live in a democracy



Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times

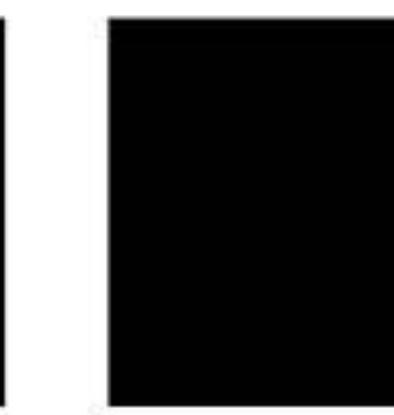
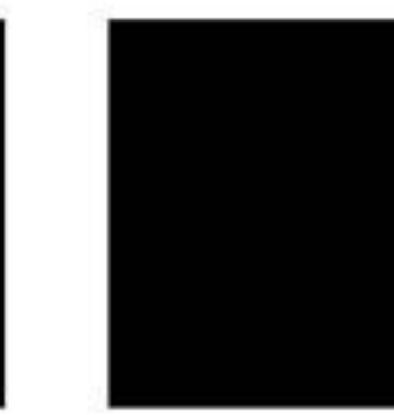


Graph by Erik Voeten, based on WVS 5

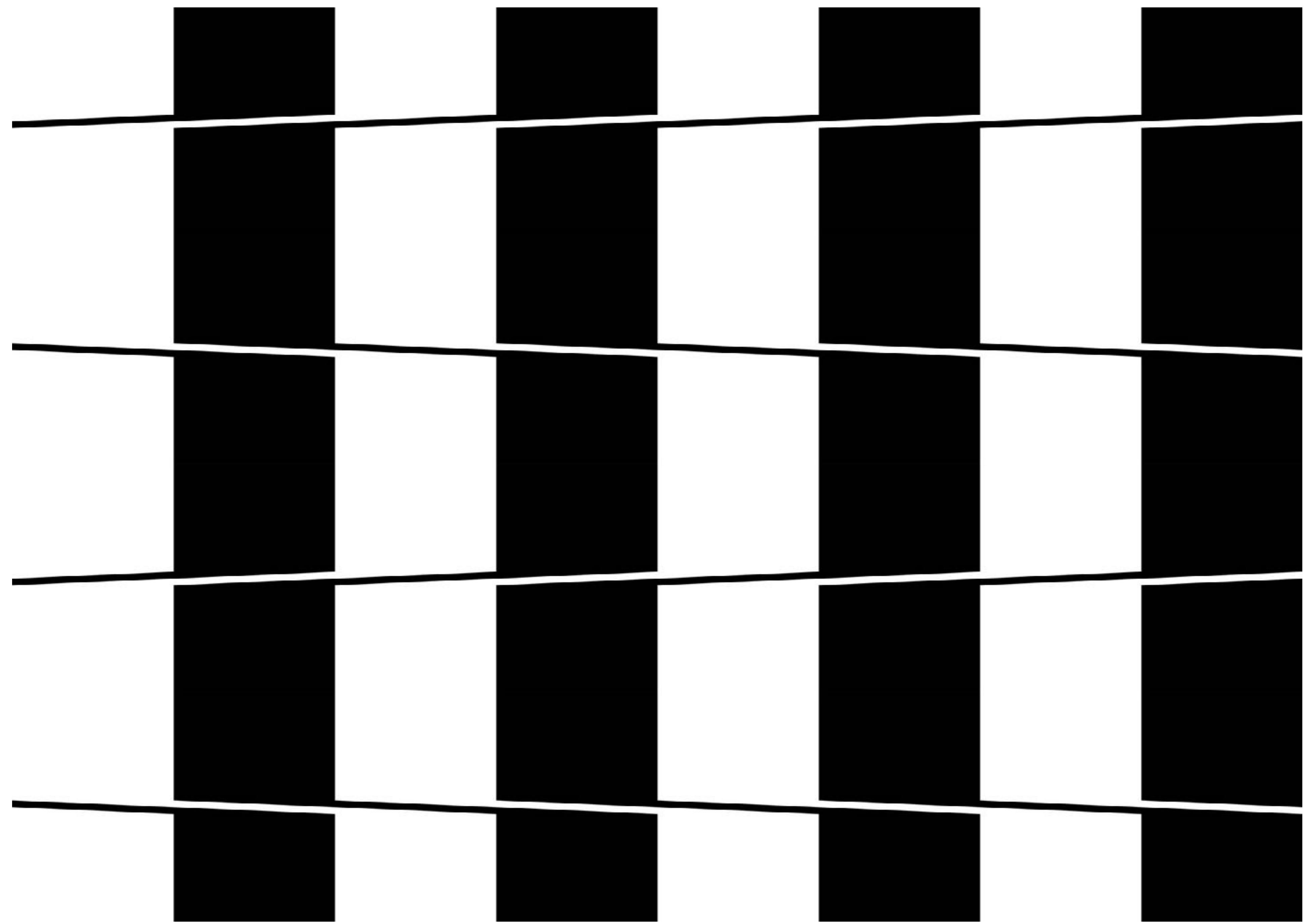
PERCEPTION

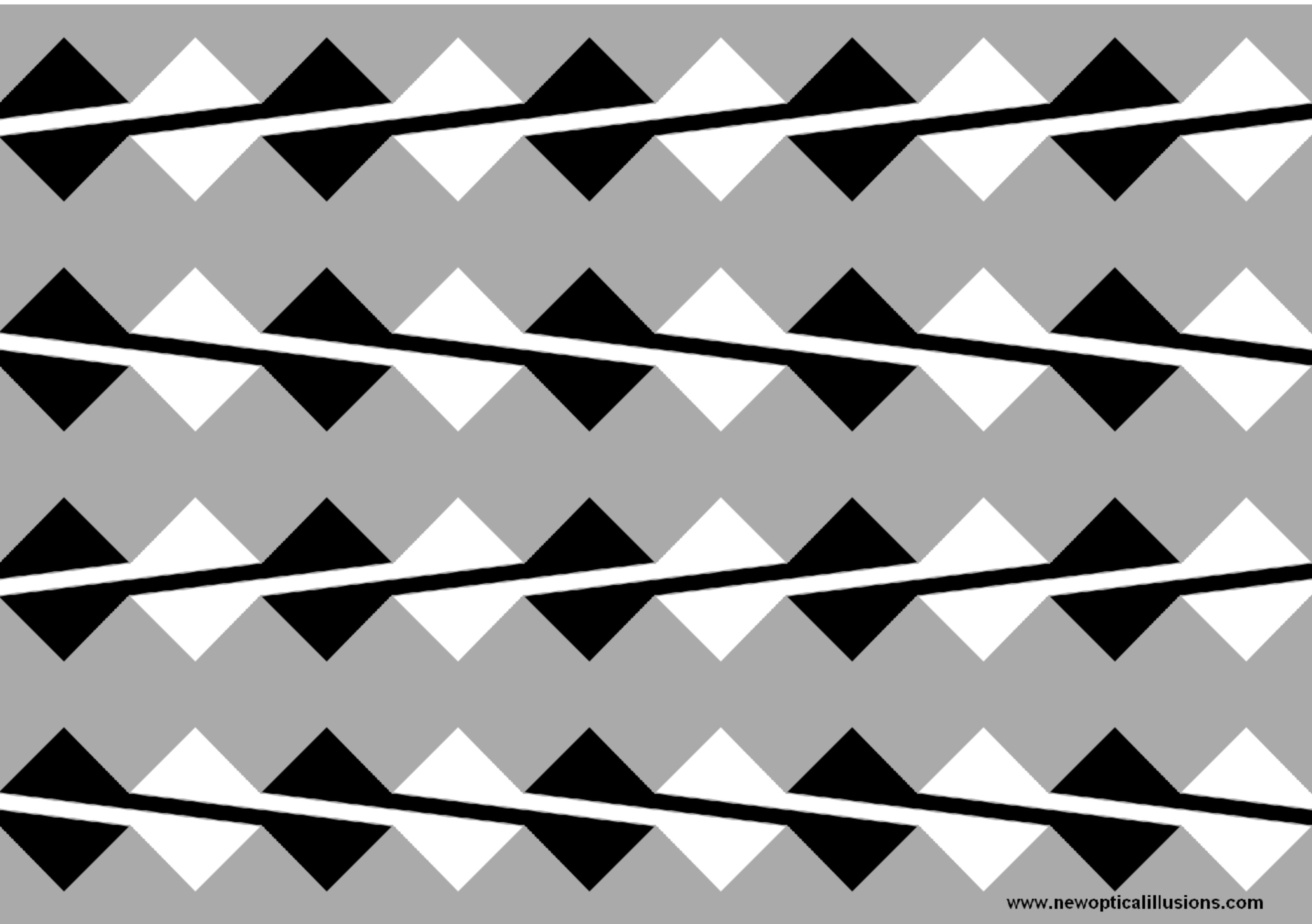
SIGHT AND INSIGHT

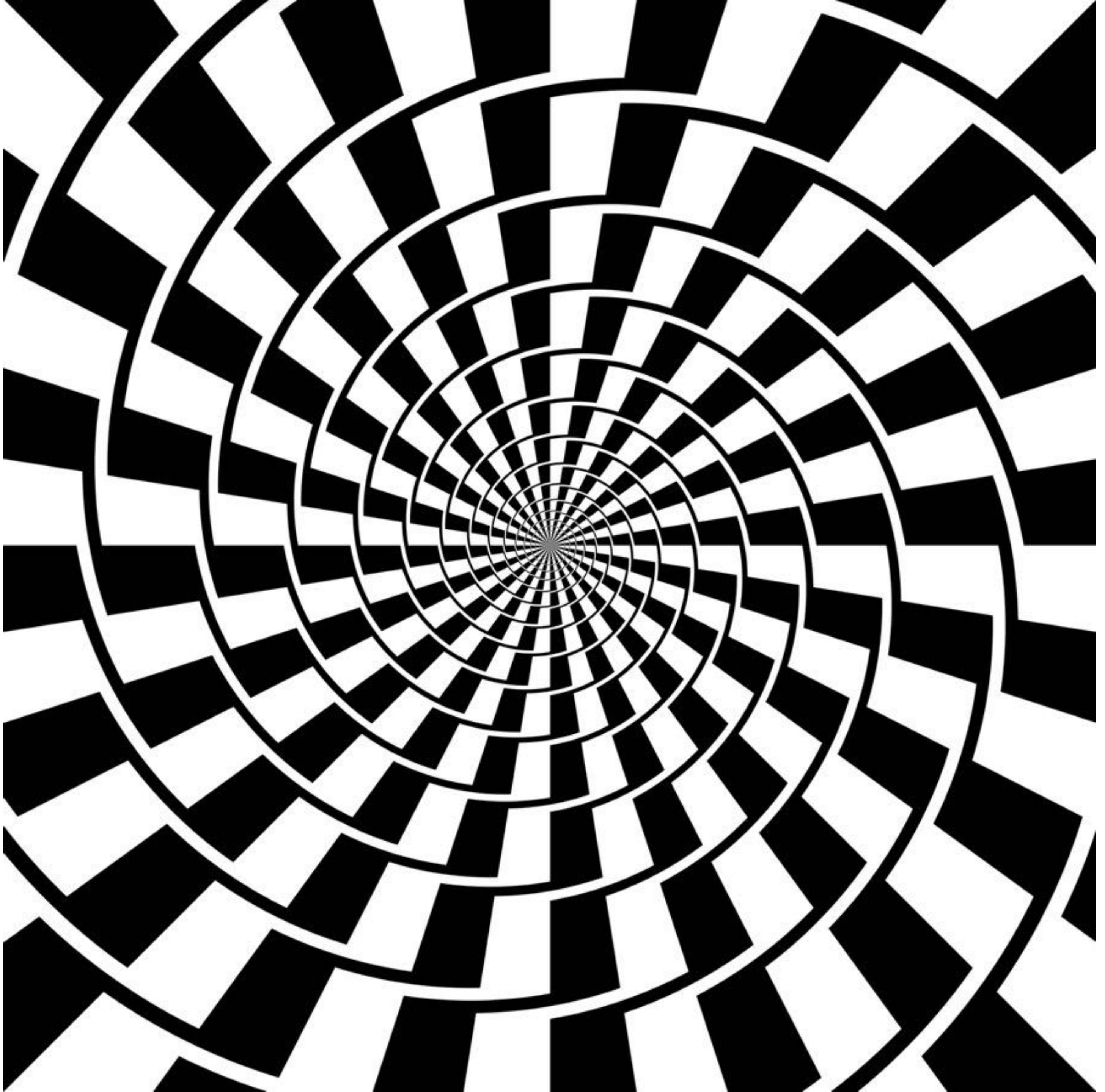
**Edges &
Contrasts**

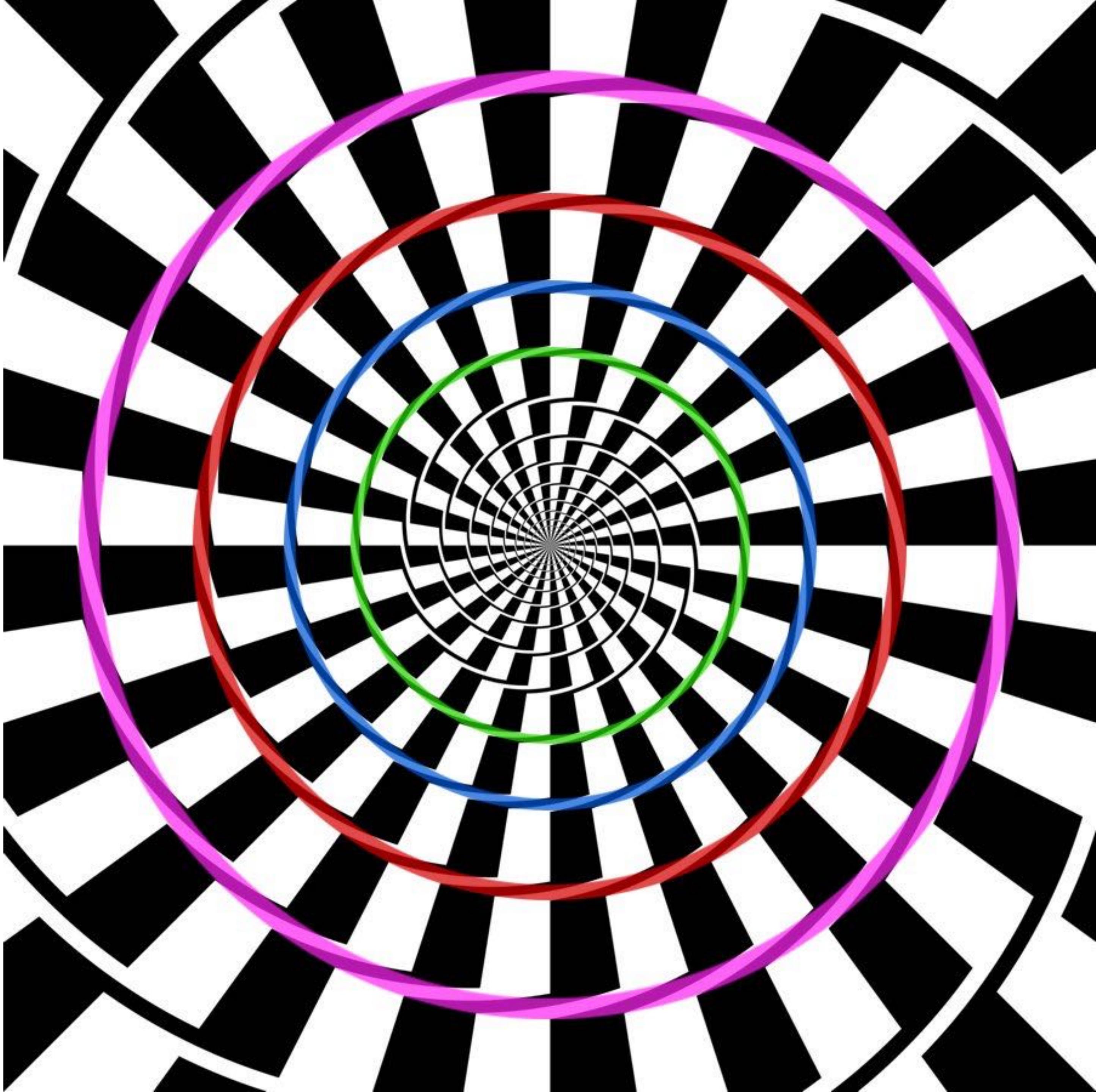


fin

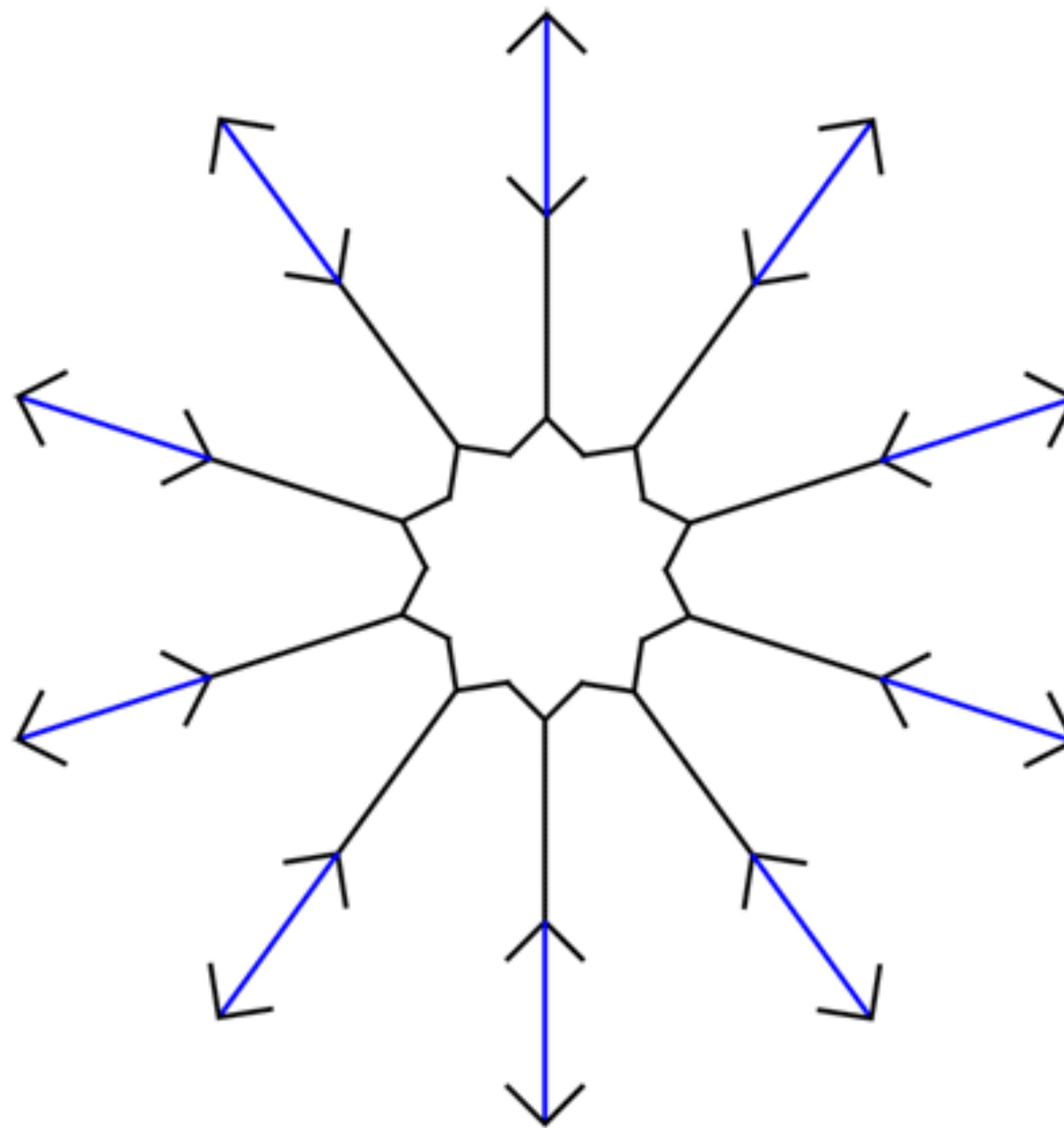






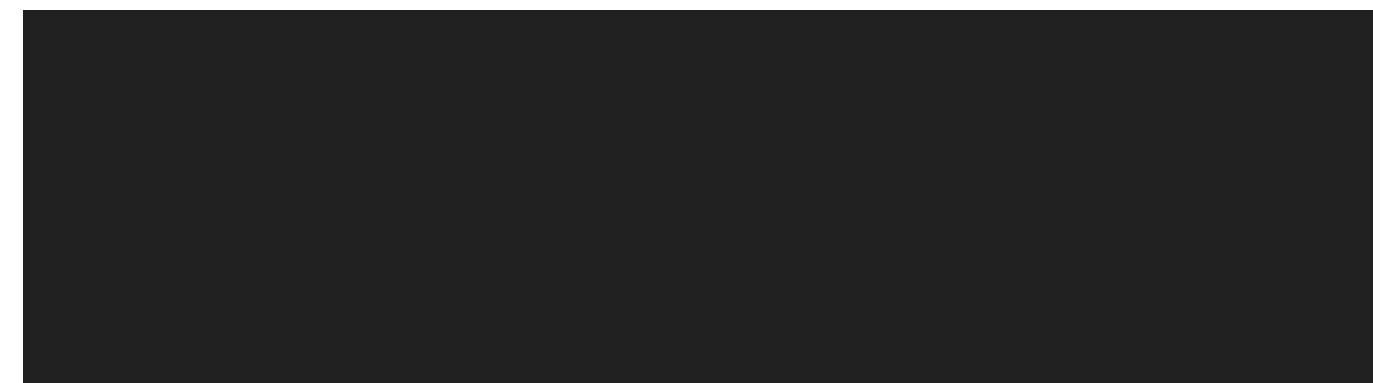


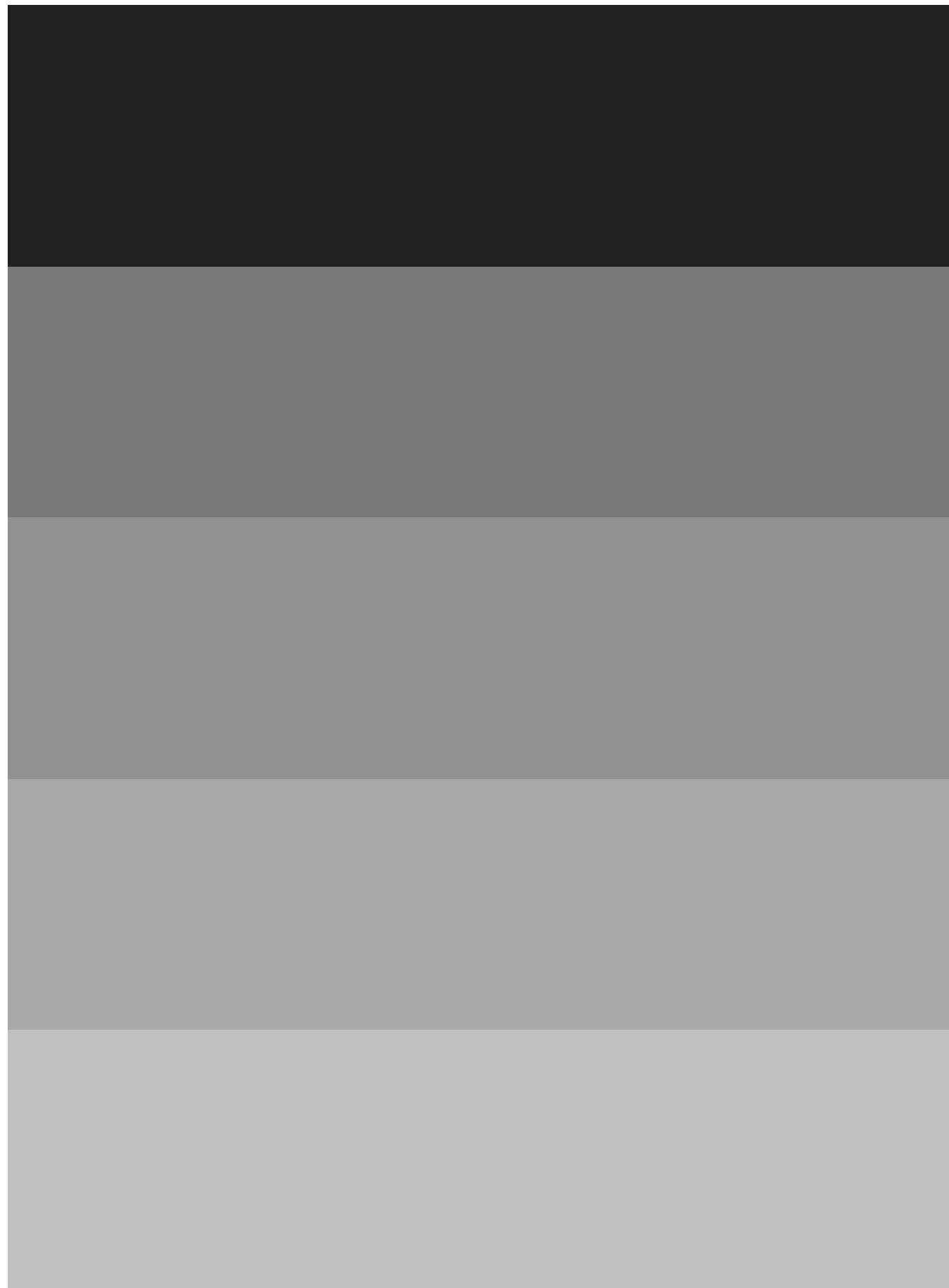
Sarcone's Dynamic Müller-Lyer Illusion



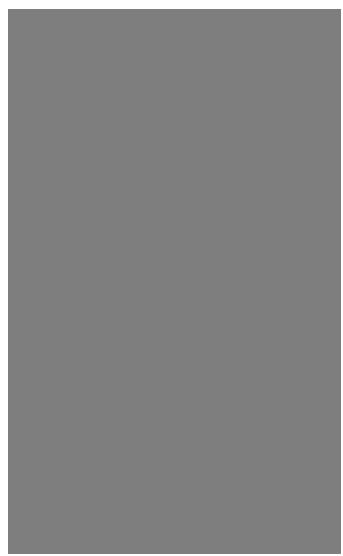
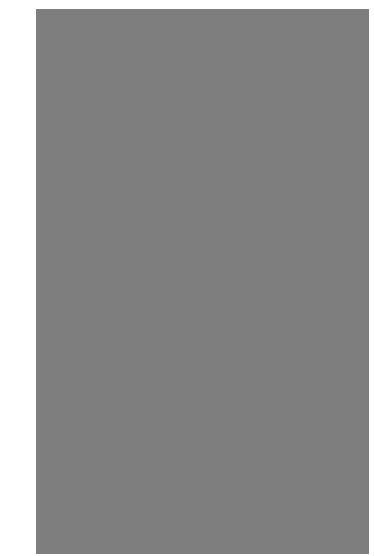


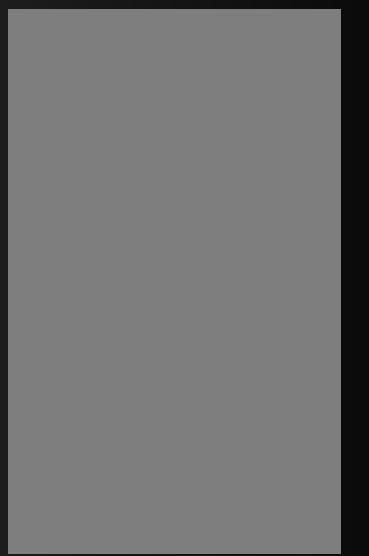
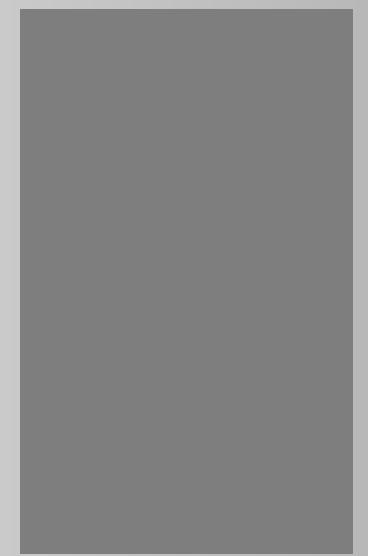
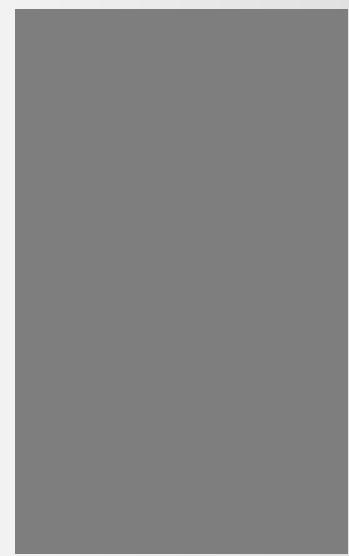
These are two perfectly
geometrical circles

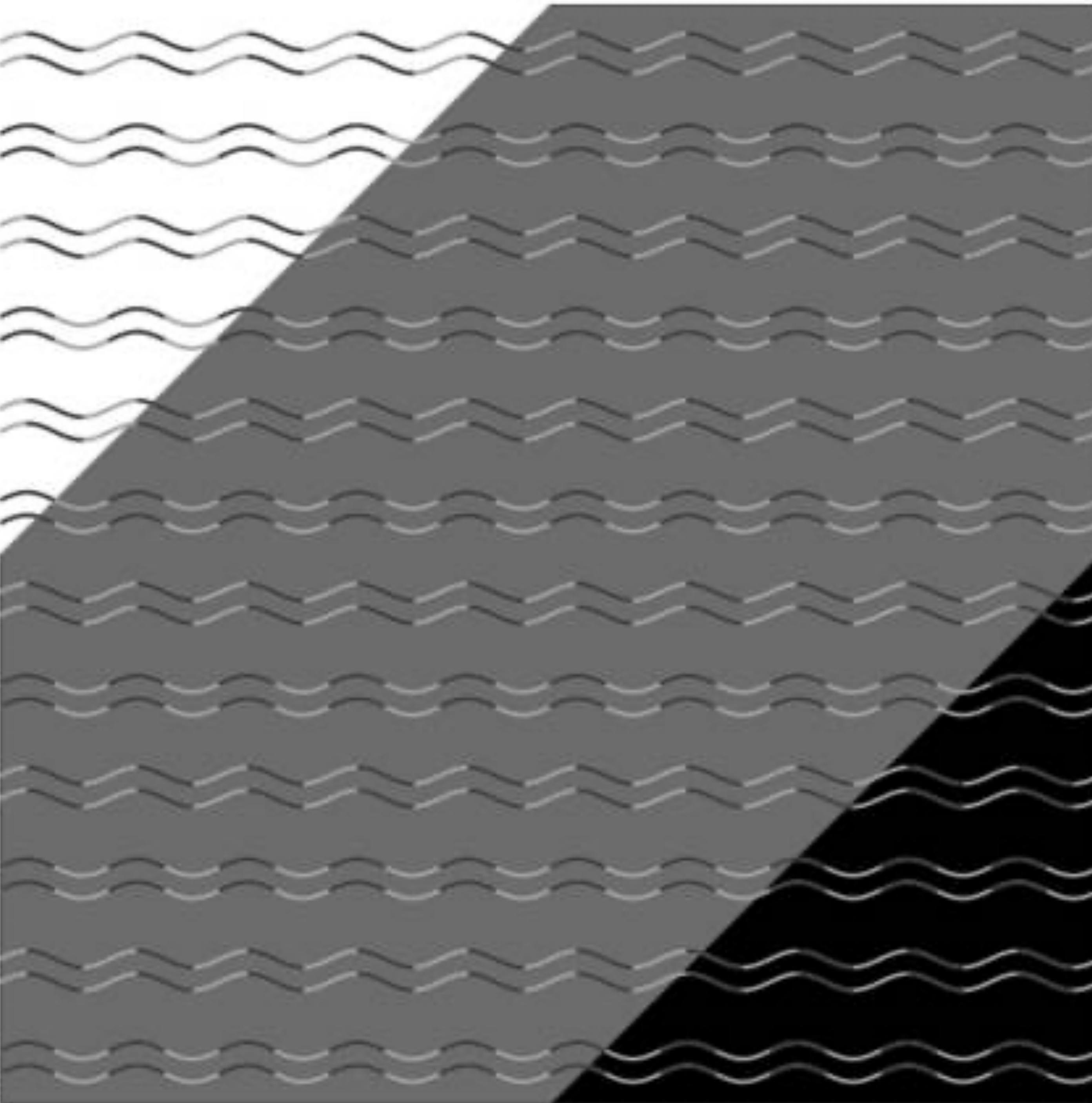


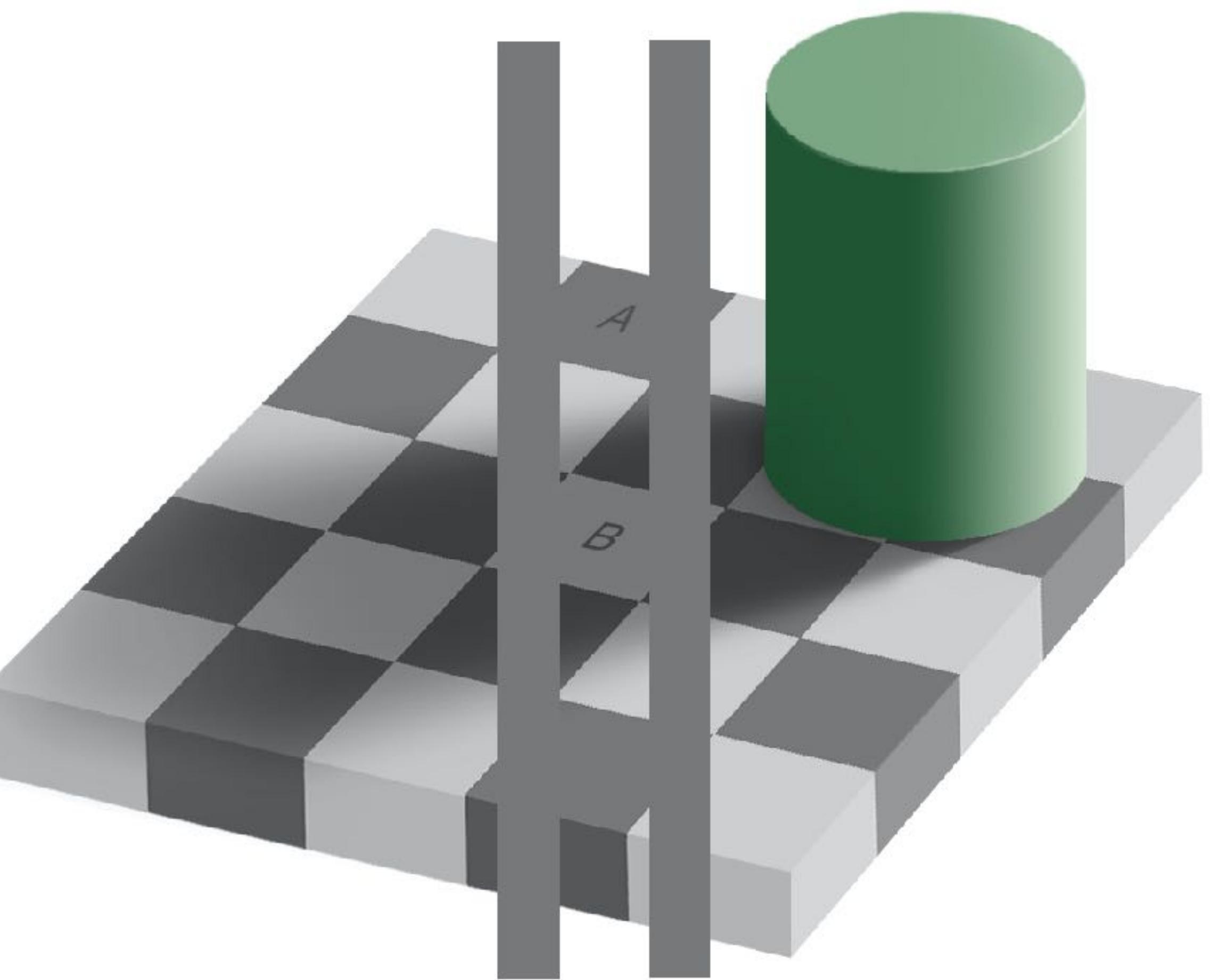
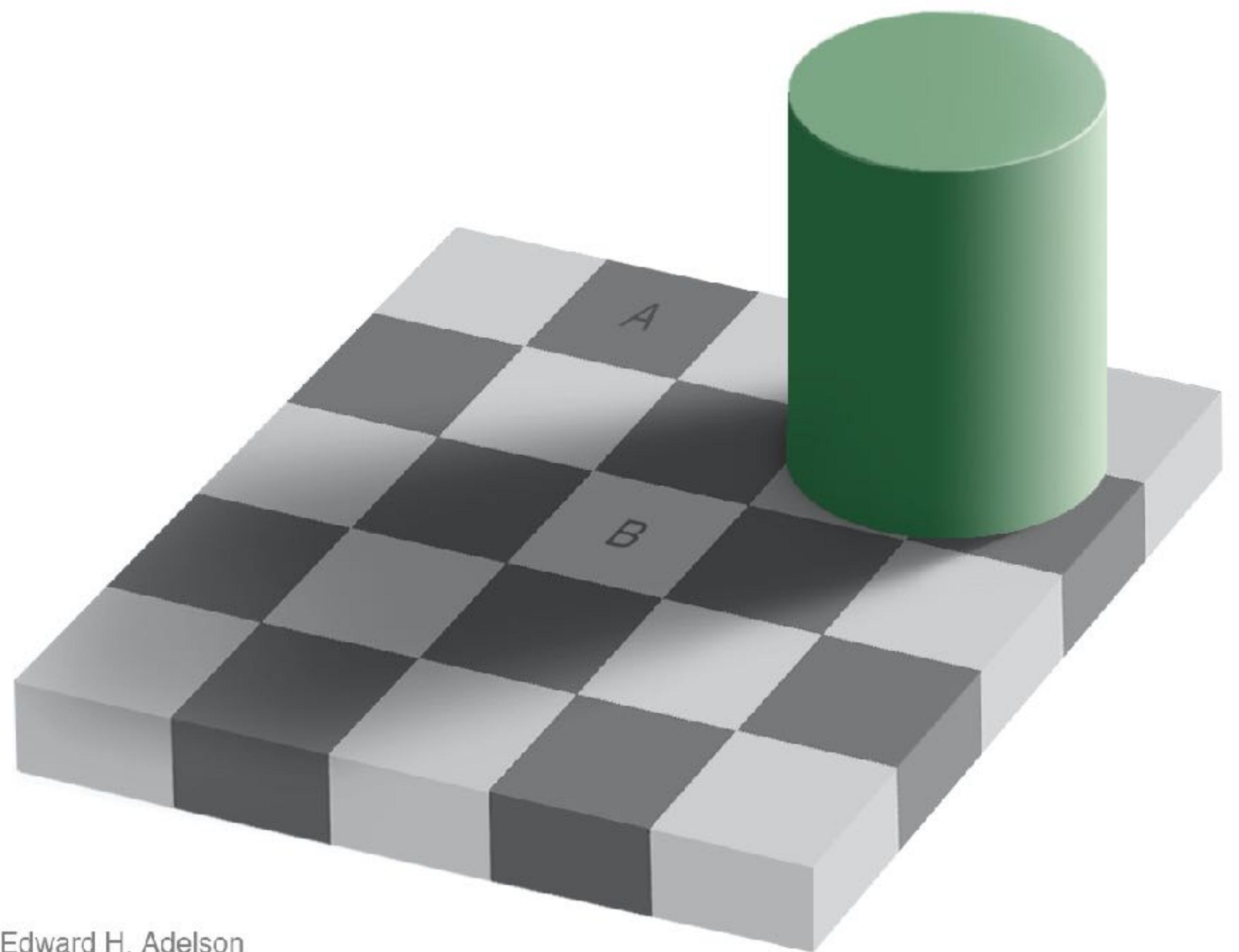




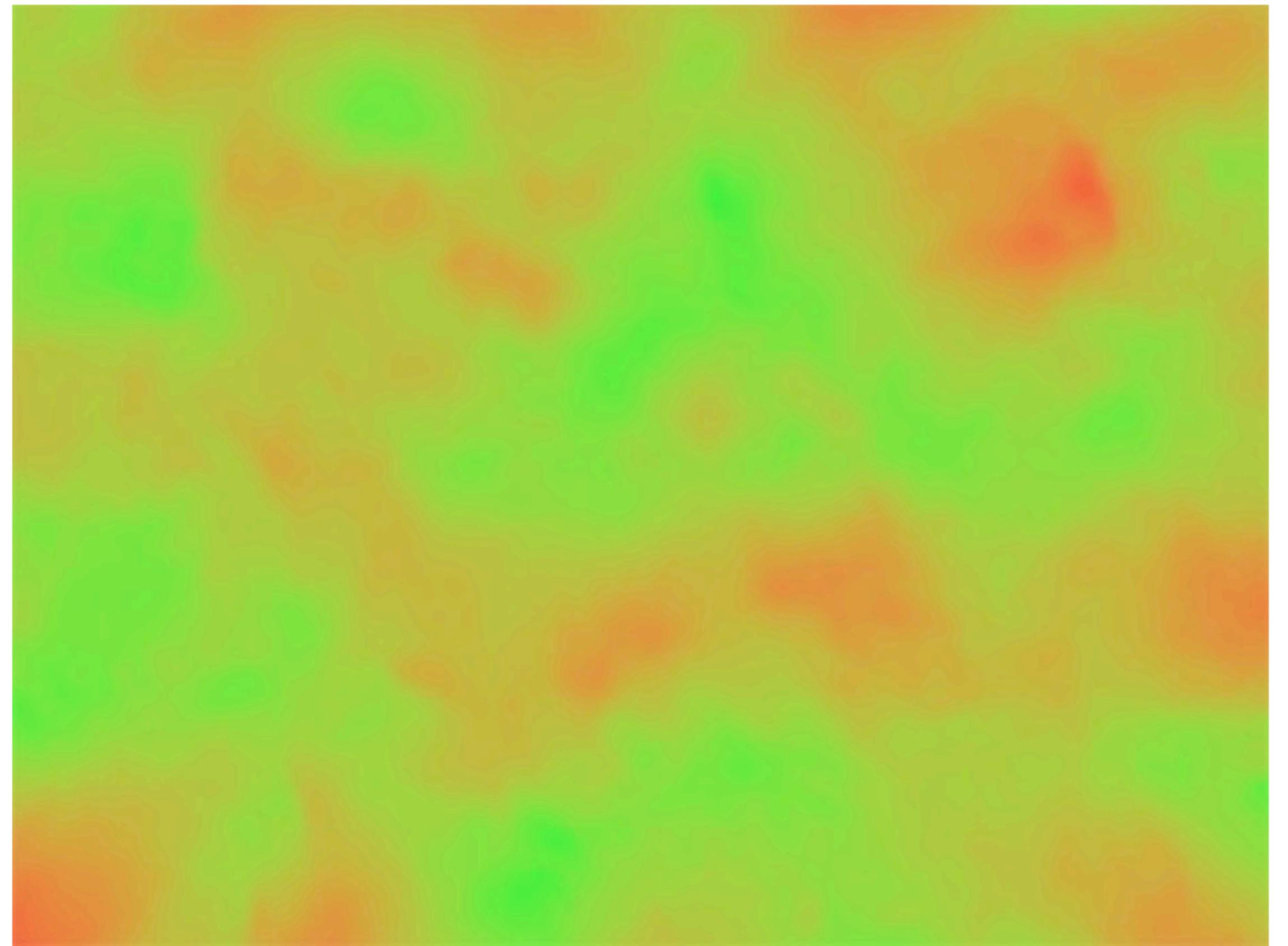
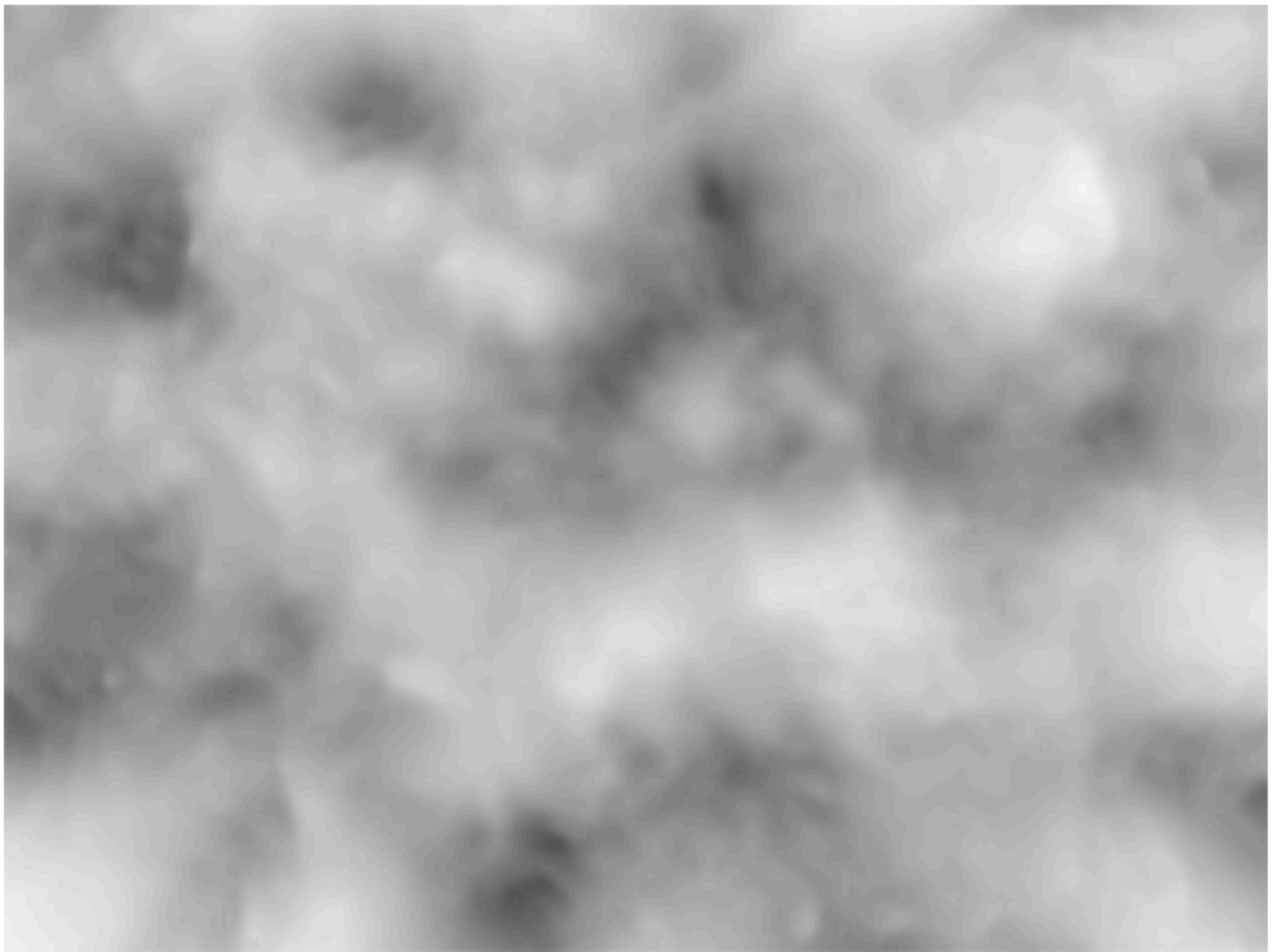


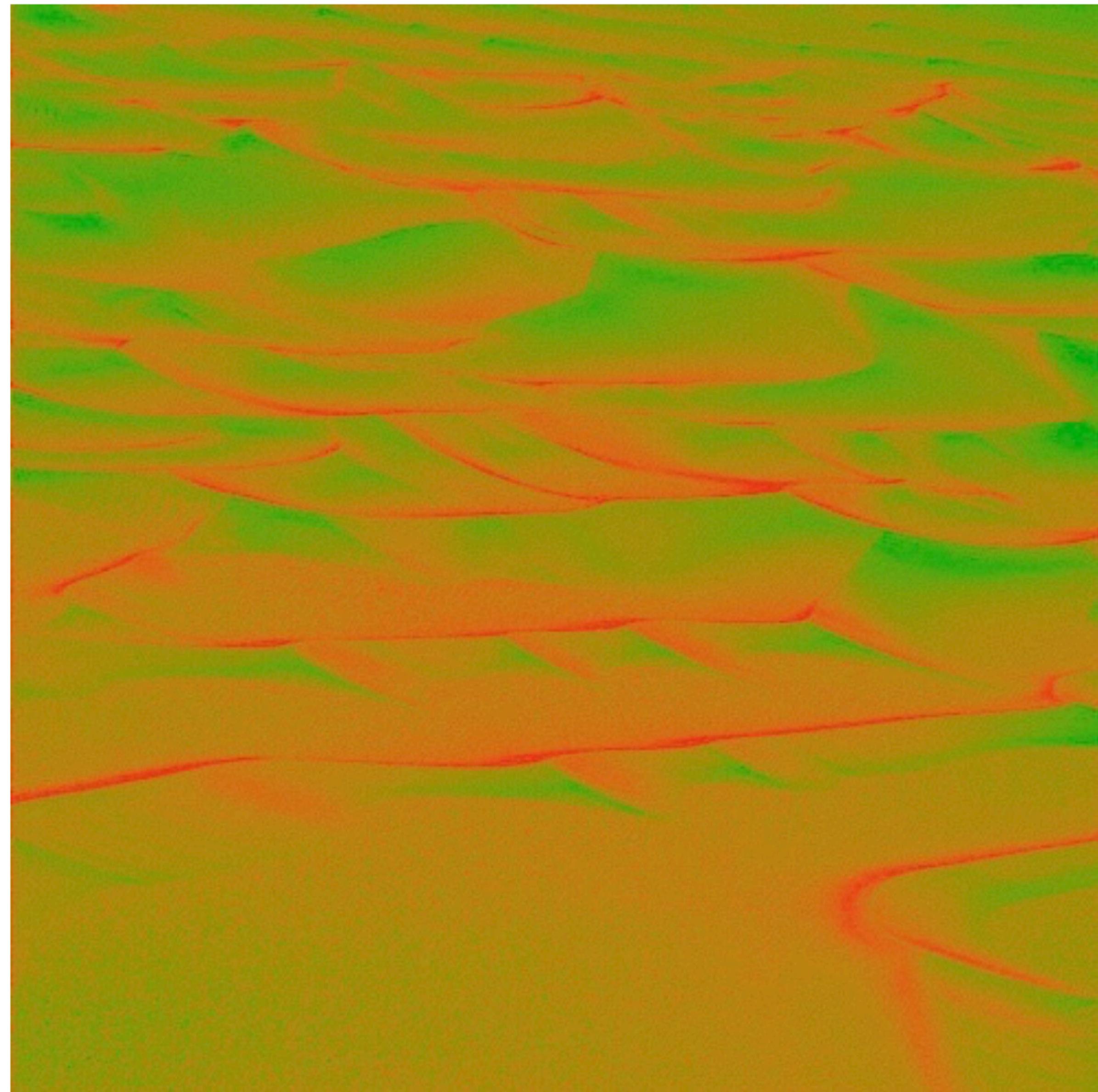


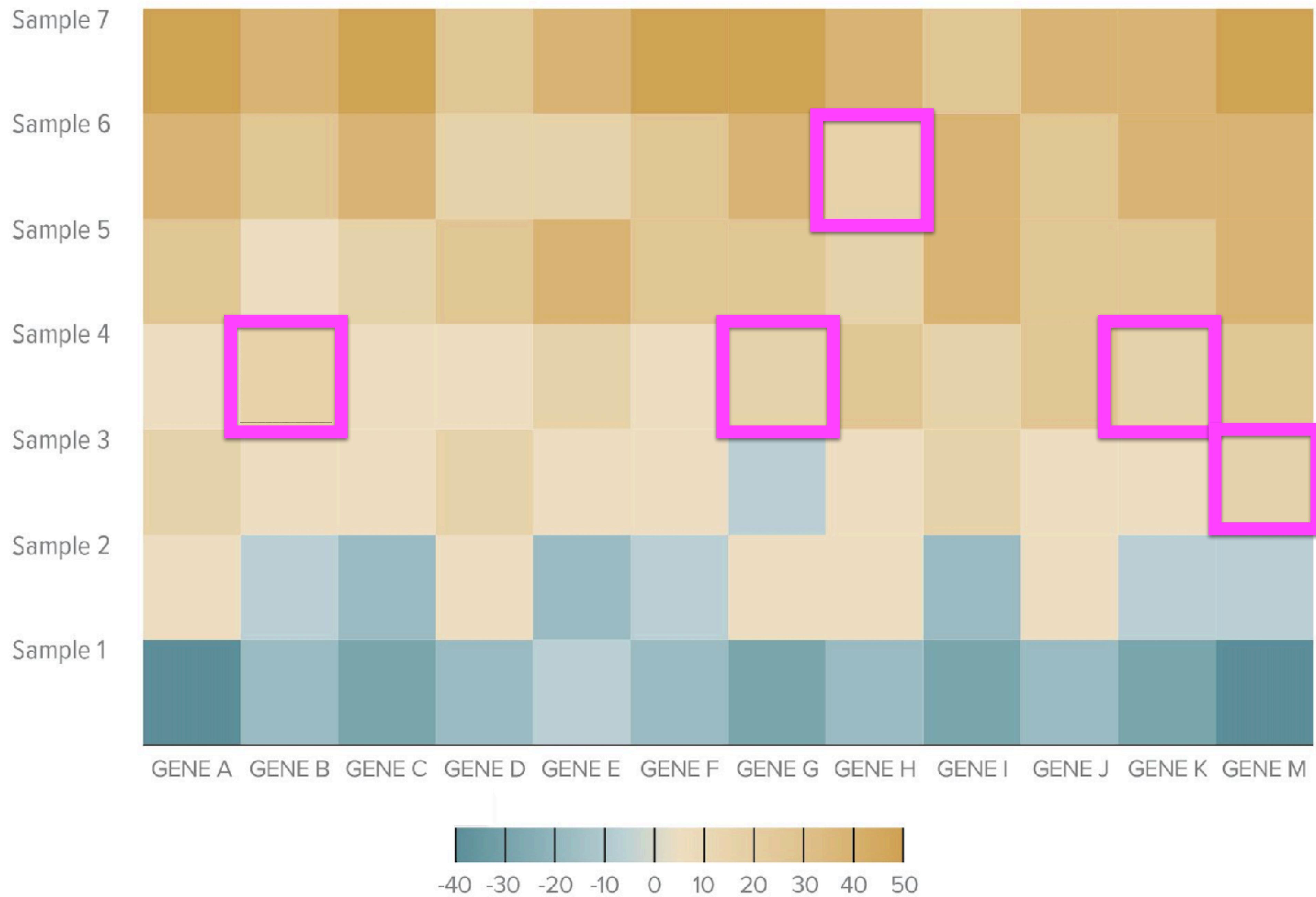


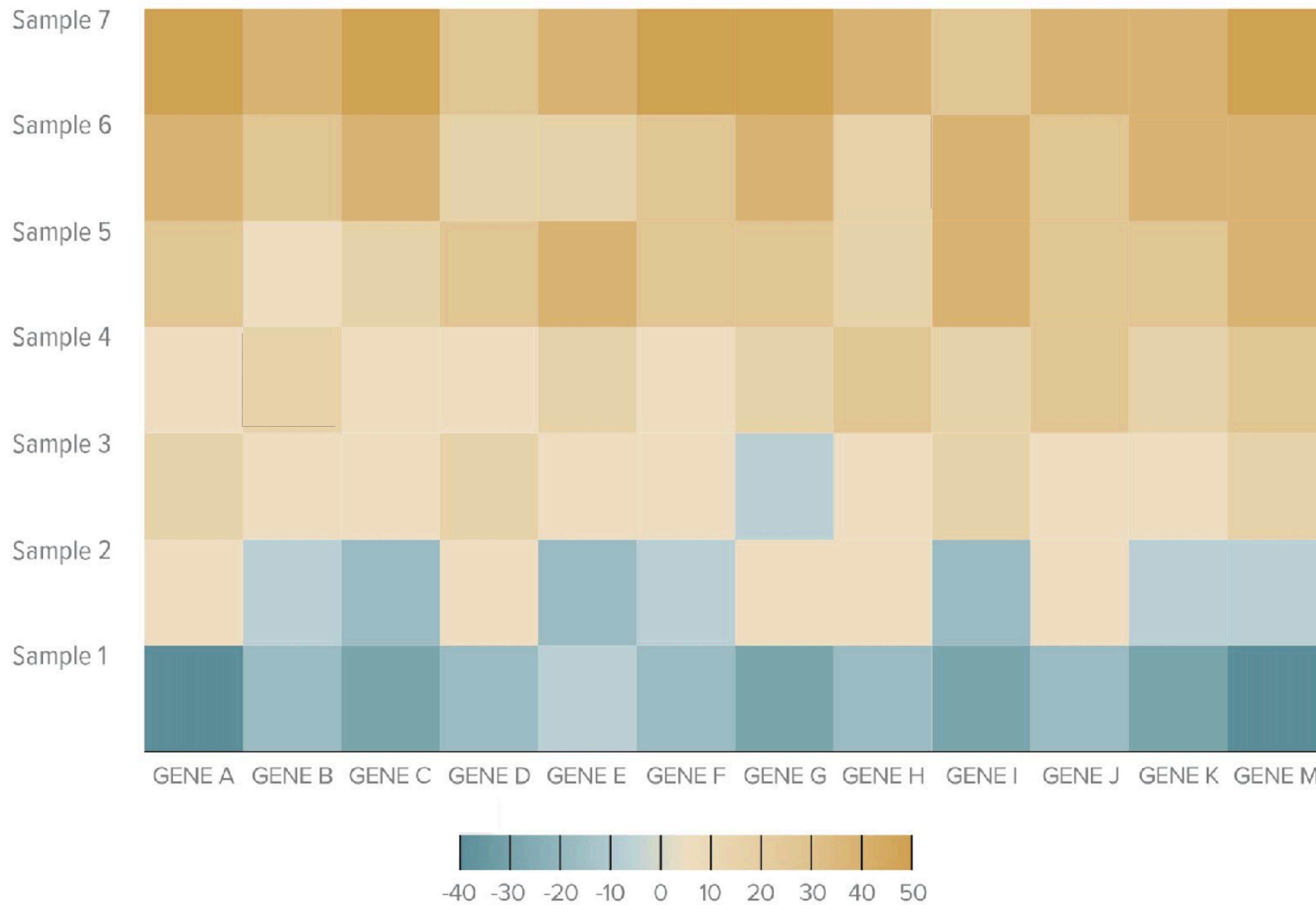


Edward H. Adelson





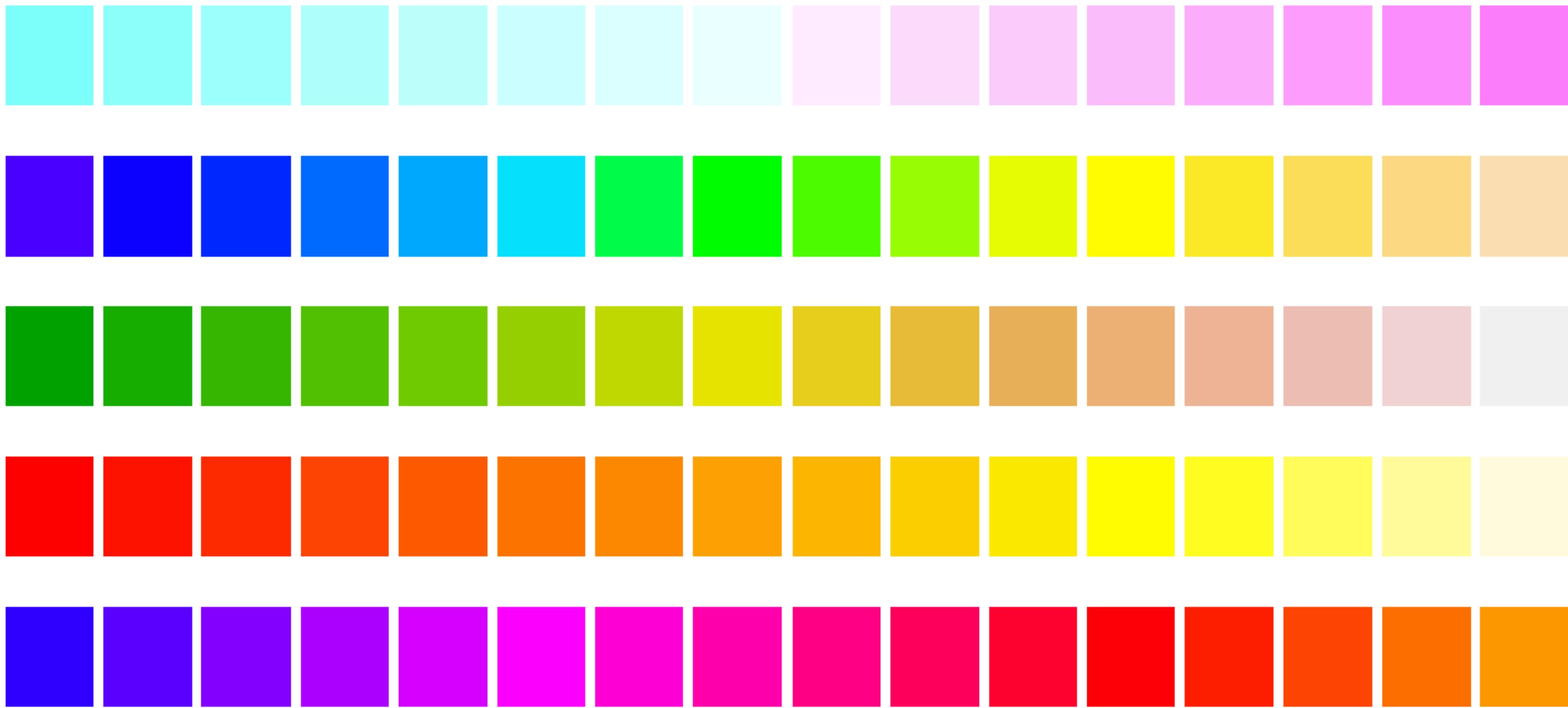




Color

+





viridis



magma



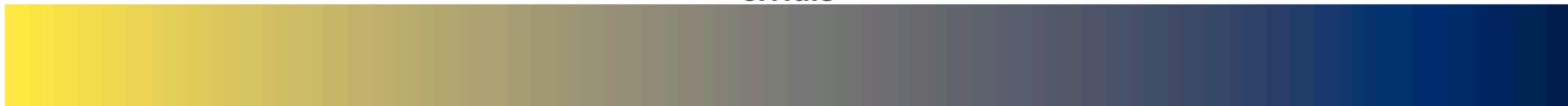
plasma

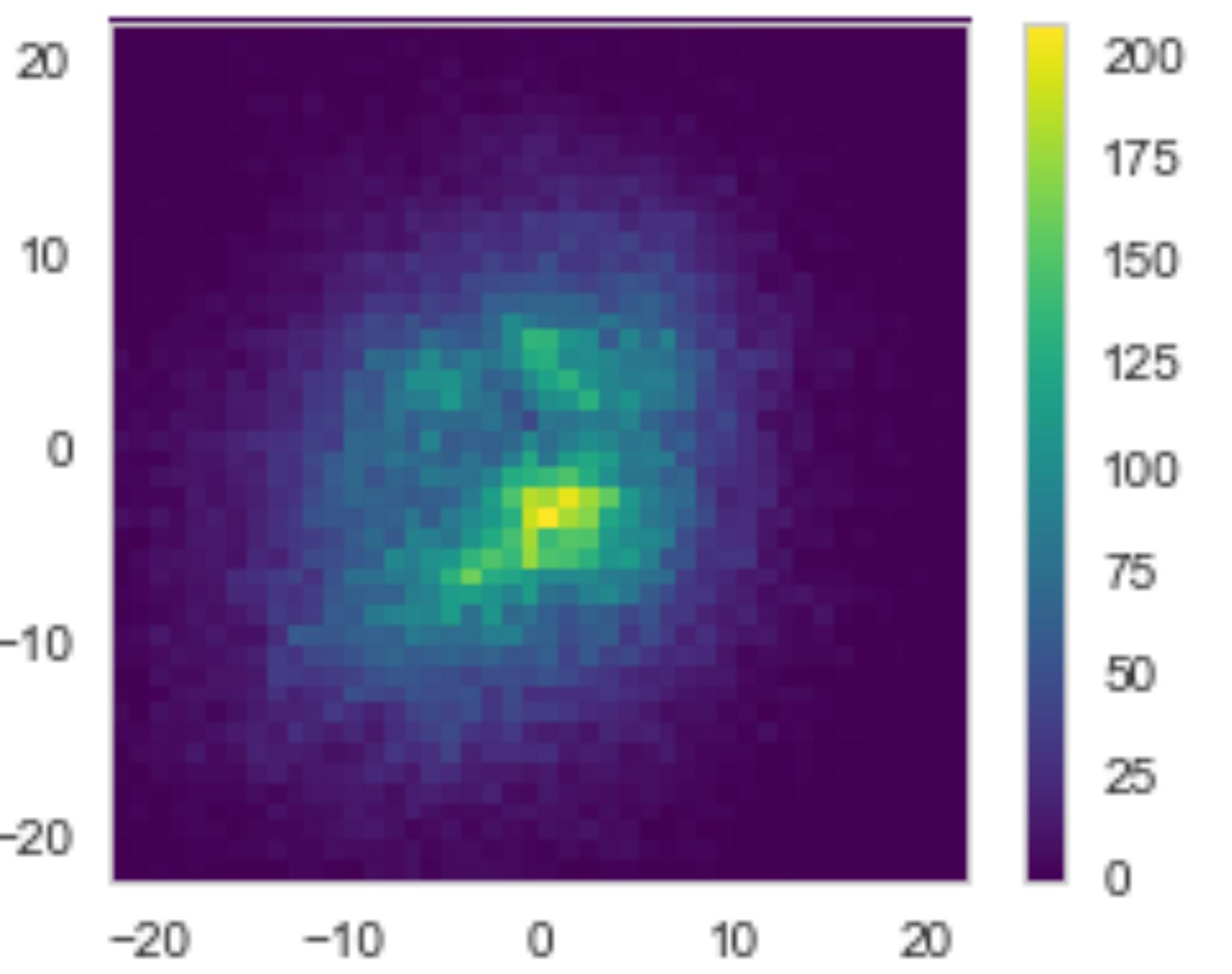
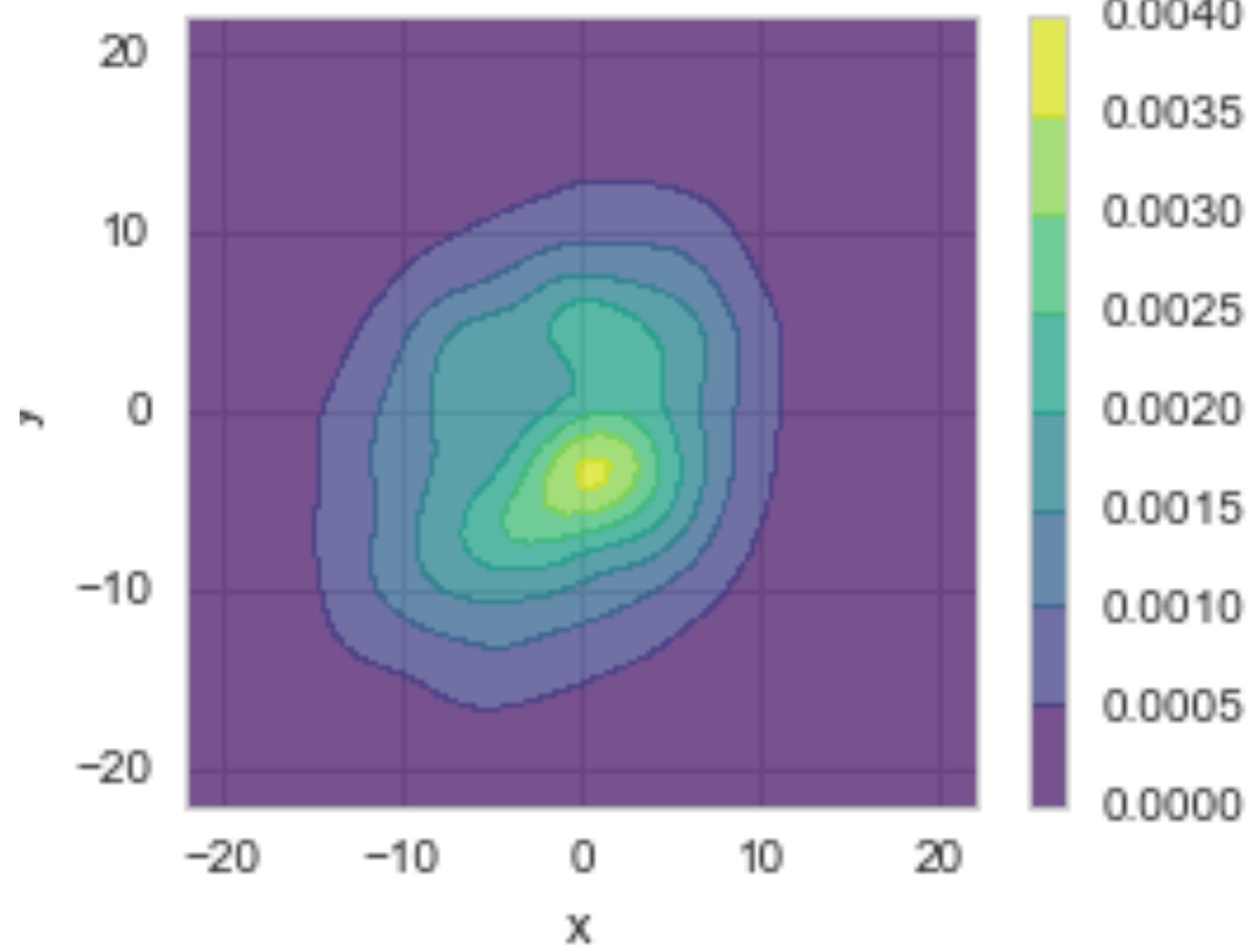
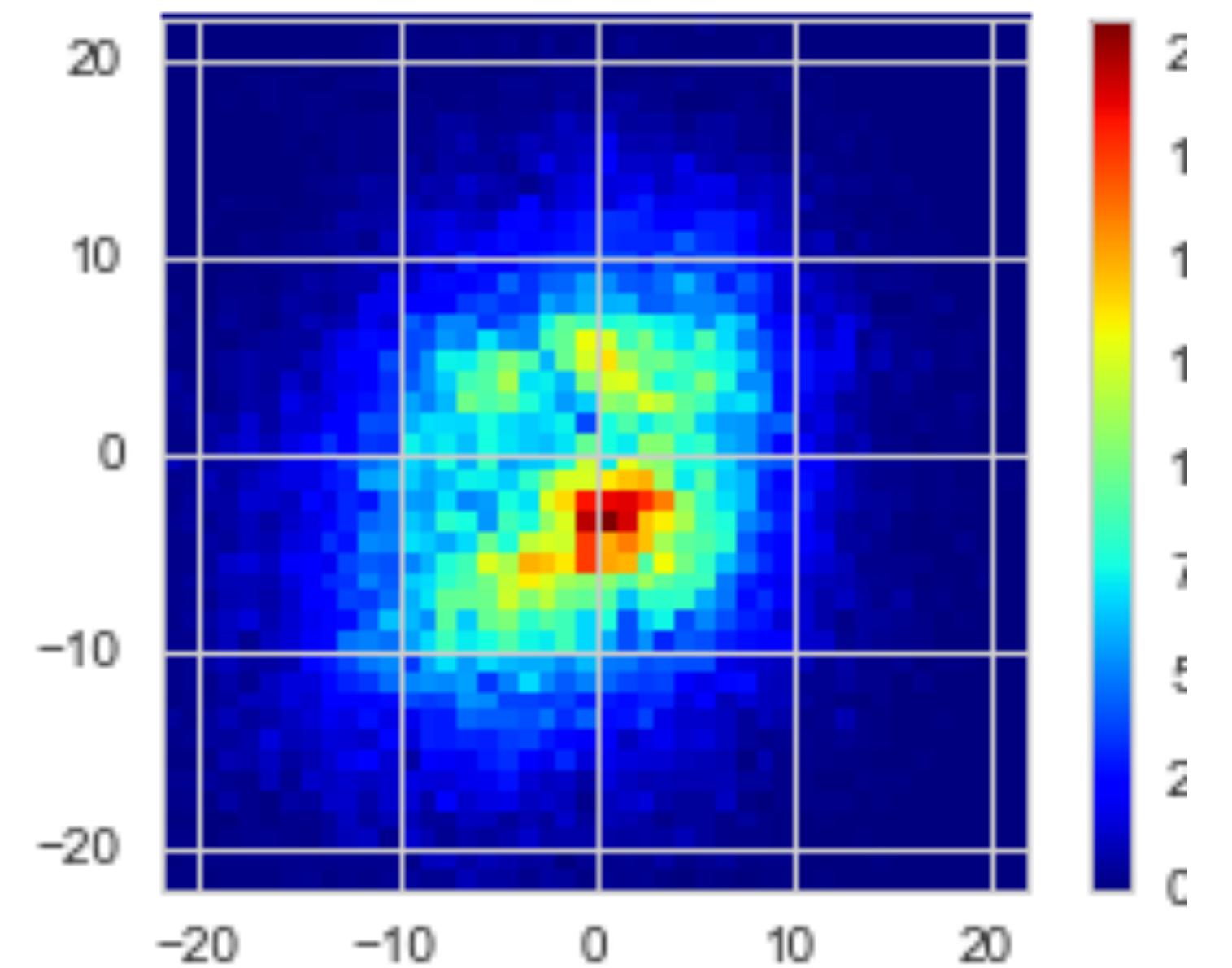
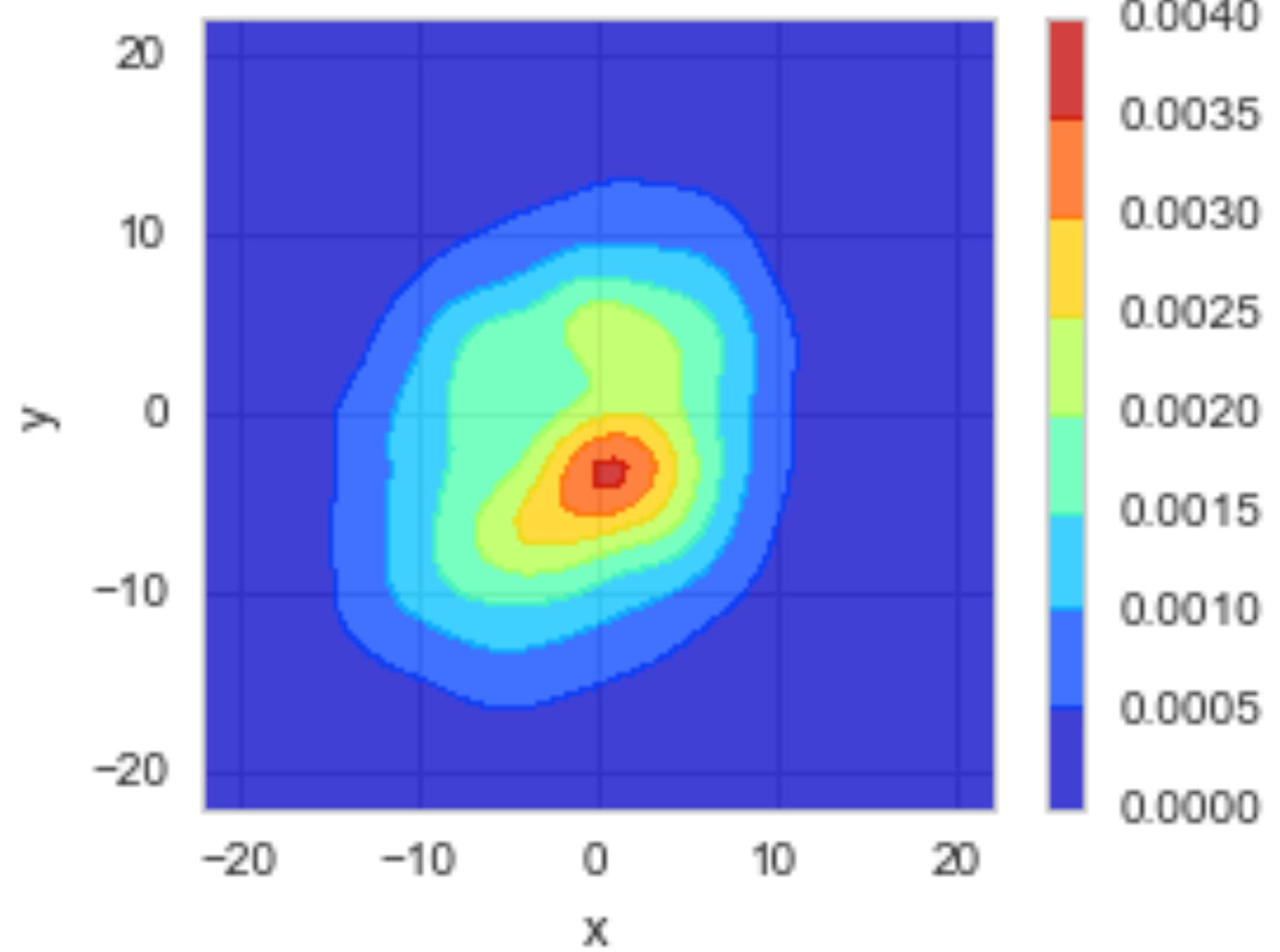


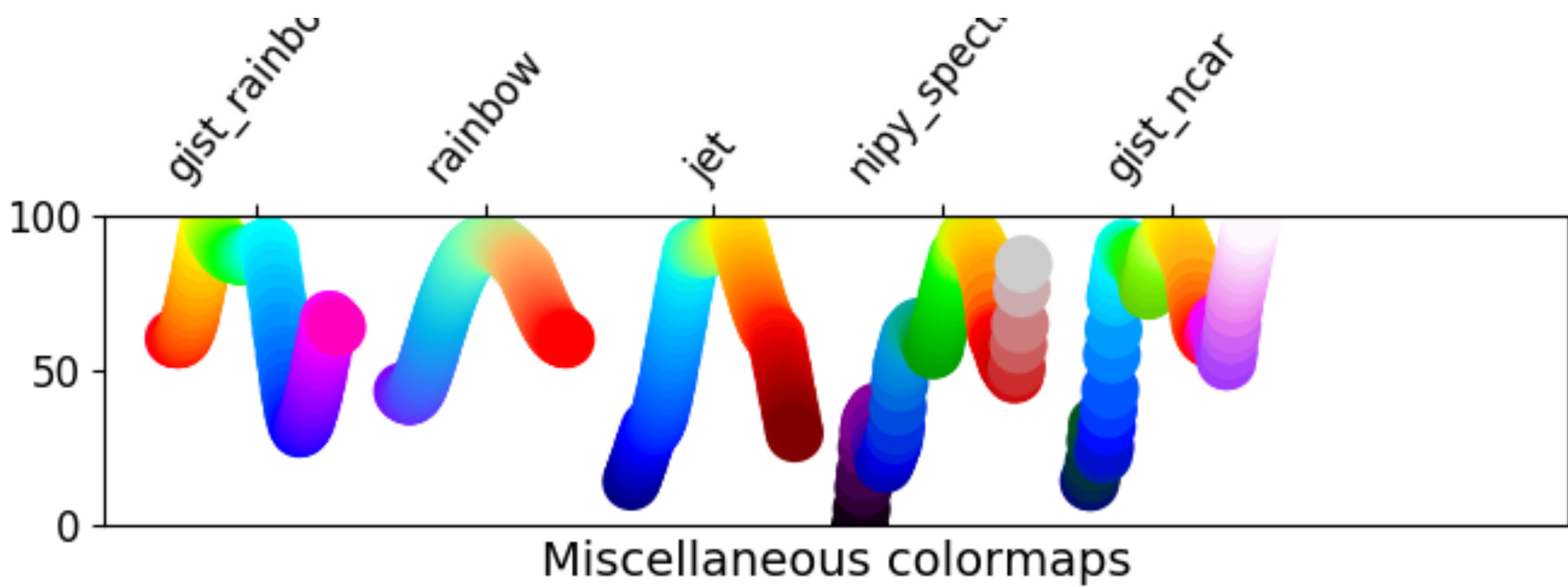
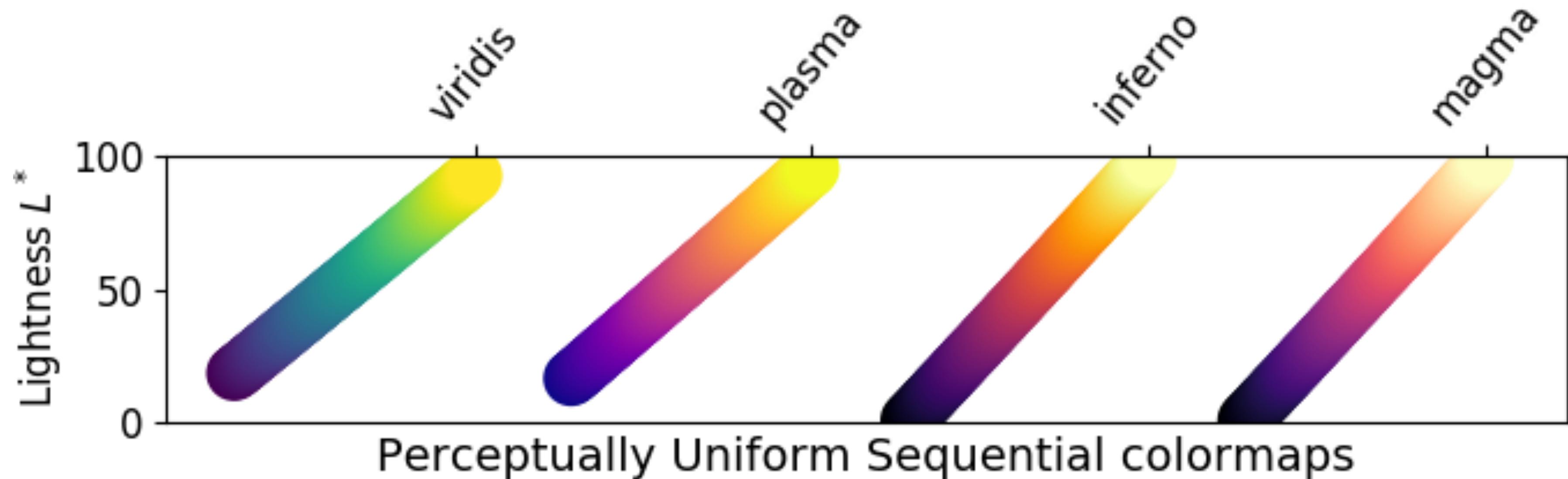
inferno

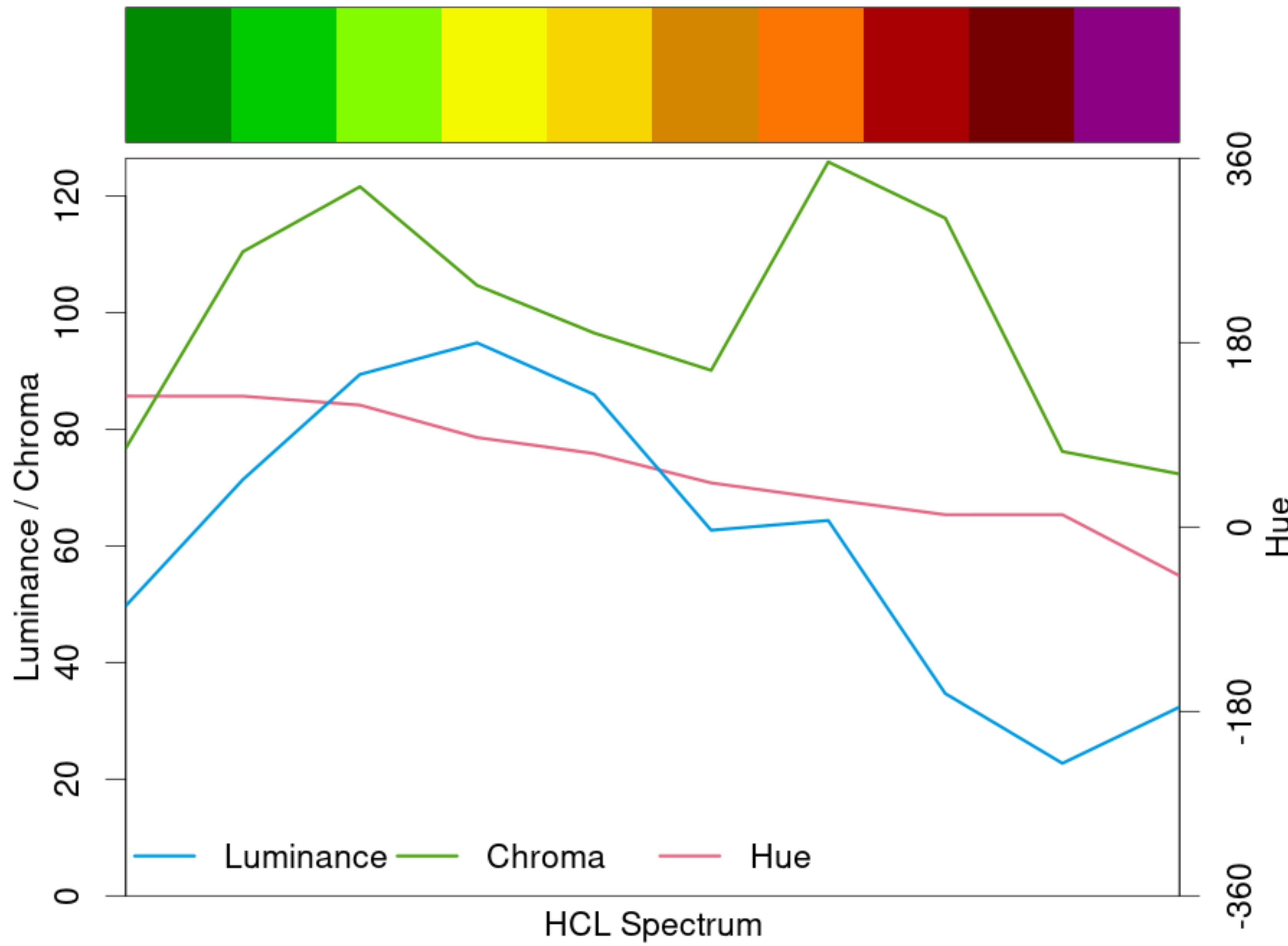


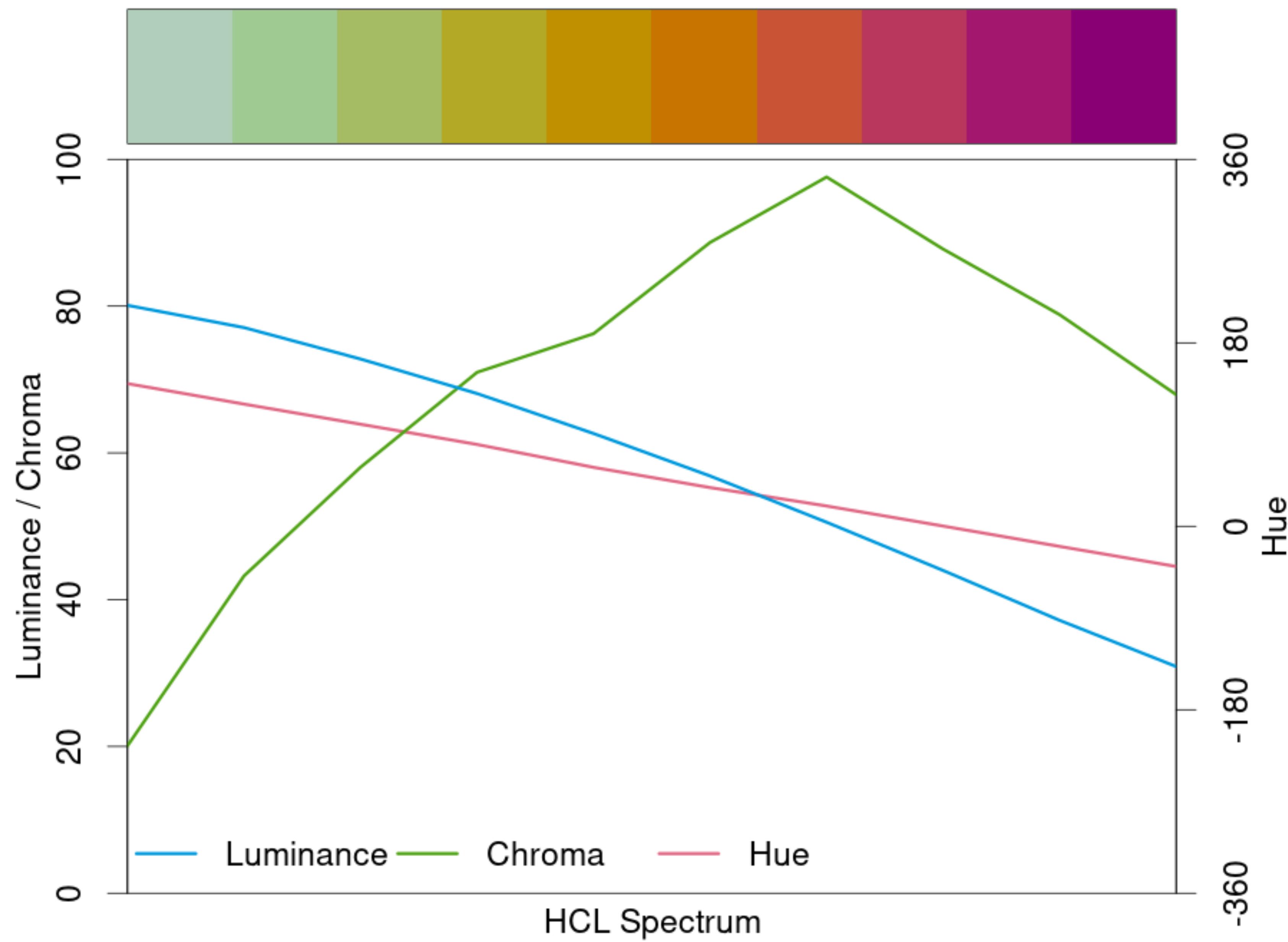
cividis

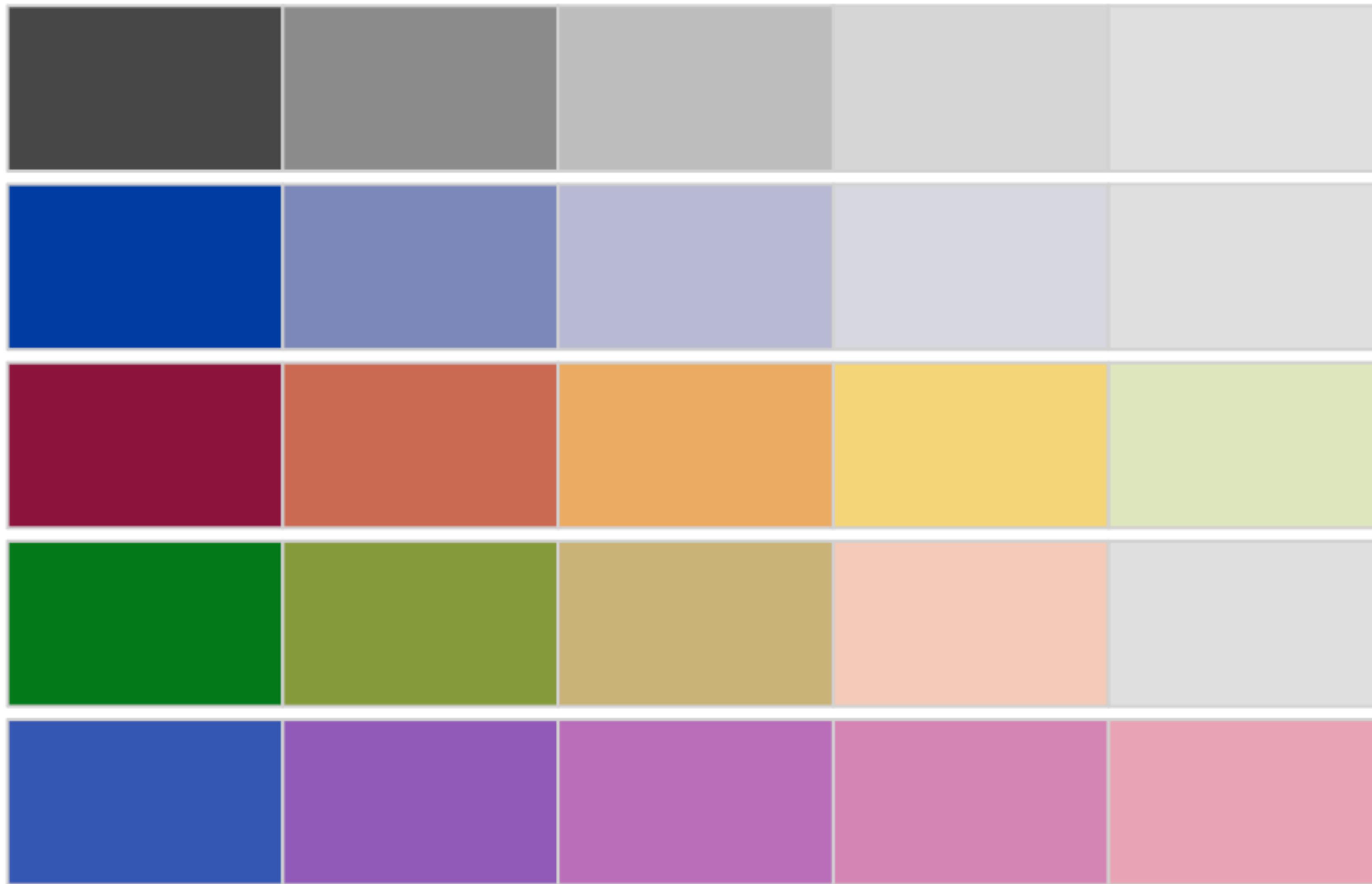


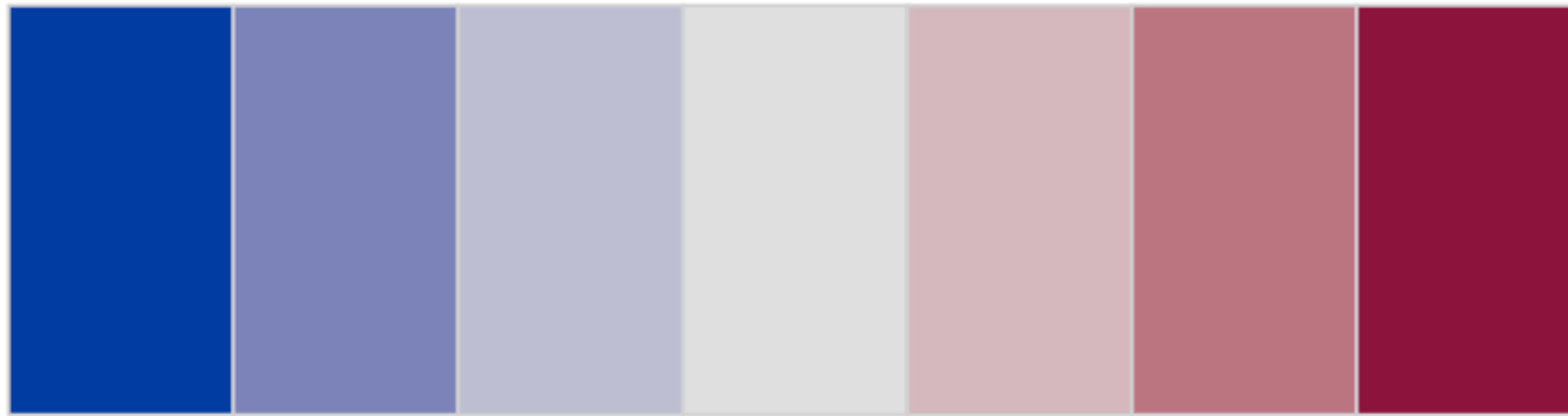








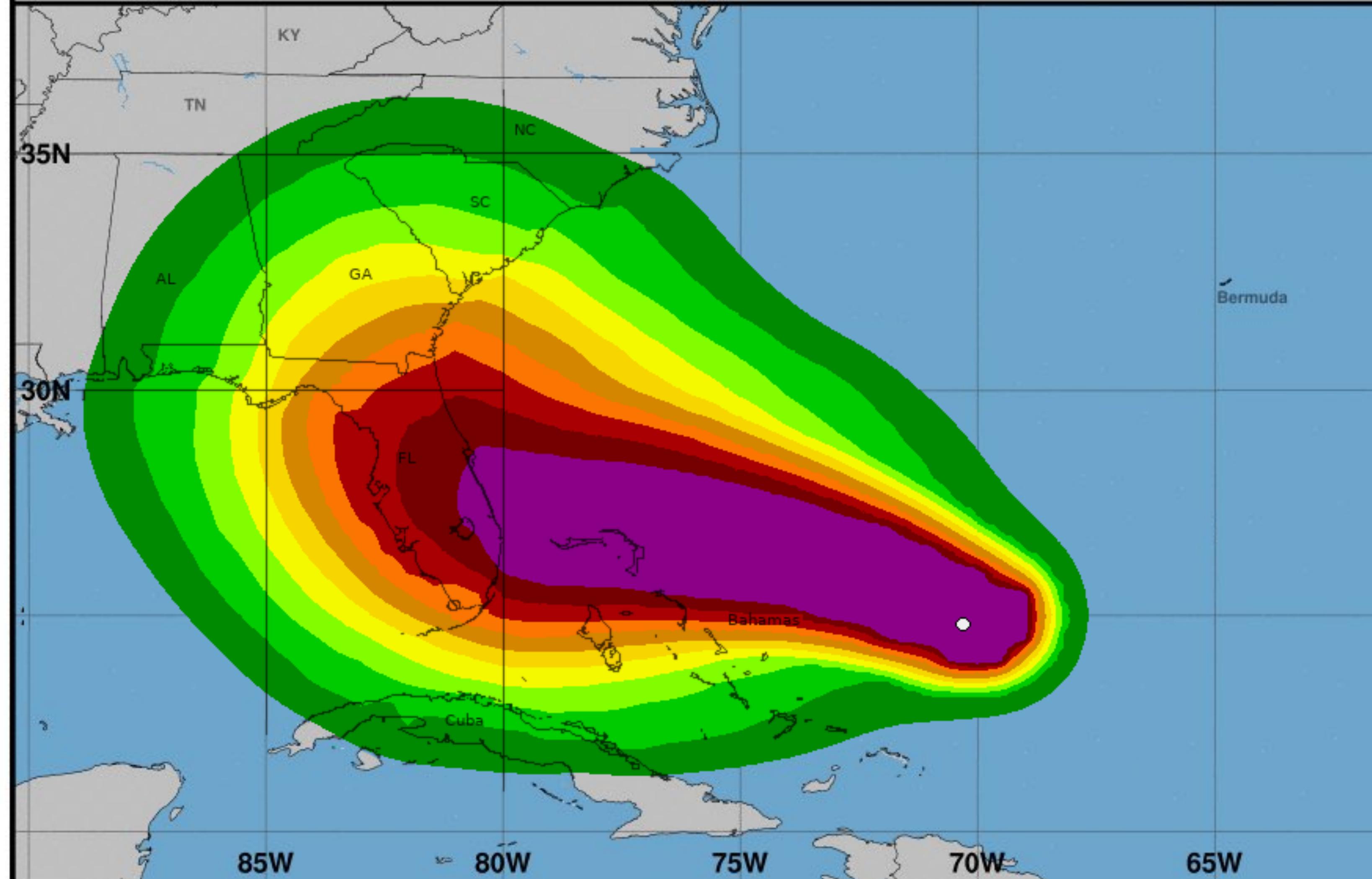






Tropical-Storm-Force Wind Speed Probabilities

For the 120 hours (5.00 days) from 2 PM EDT FRI AUG 30 to 2 PM EDT WED SEP 04



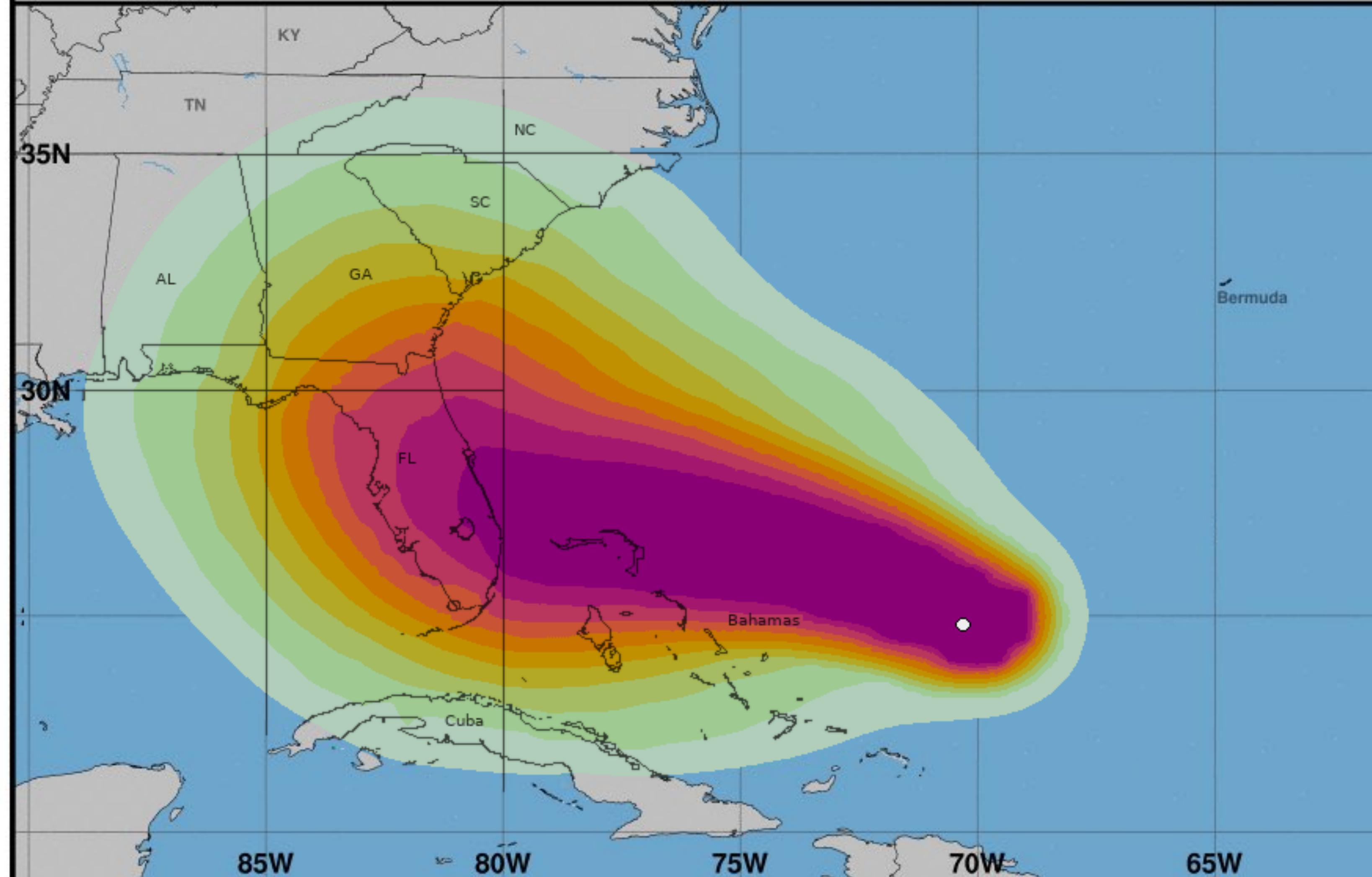
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





Tropical-Storm-Force Wind Speed Probabilities

For the 120 hours (5.00 days) from 2 PM EDT FRI AUG 30 to 2 PM EDT WED SEP 04



Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)

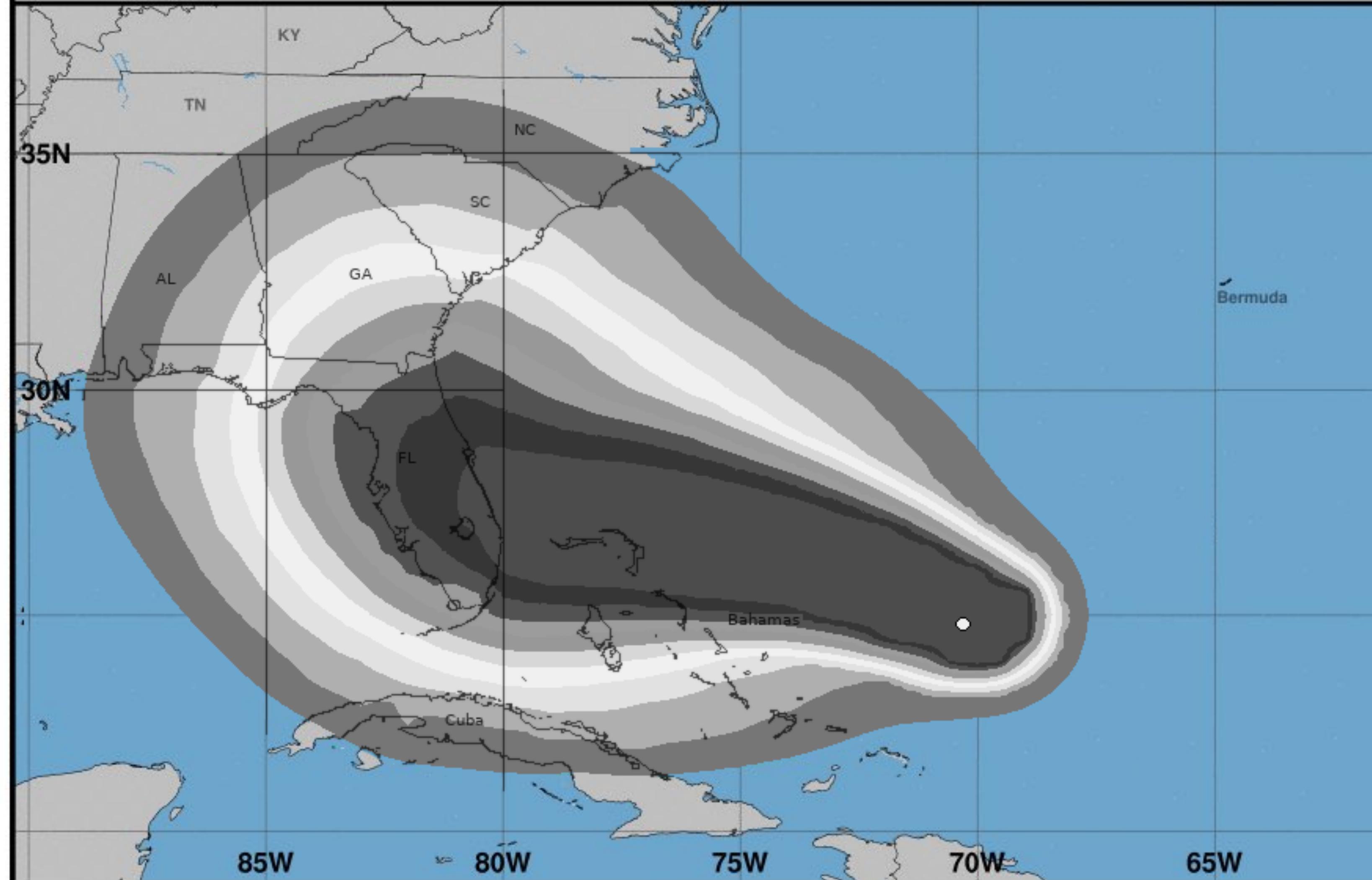


Achim Zeileis
<https://bit.ly/35zBIYL>



Tropical-Storm-Force Wind Speed Probabilities

For the 120 hours (5.00 days) from 2 PM EDT FRI AUG 30 to 2 PM EDT WED SEP 04



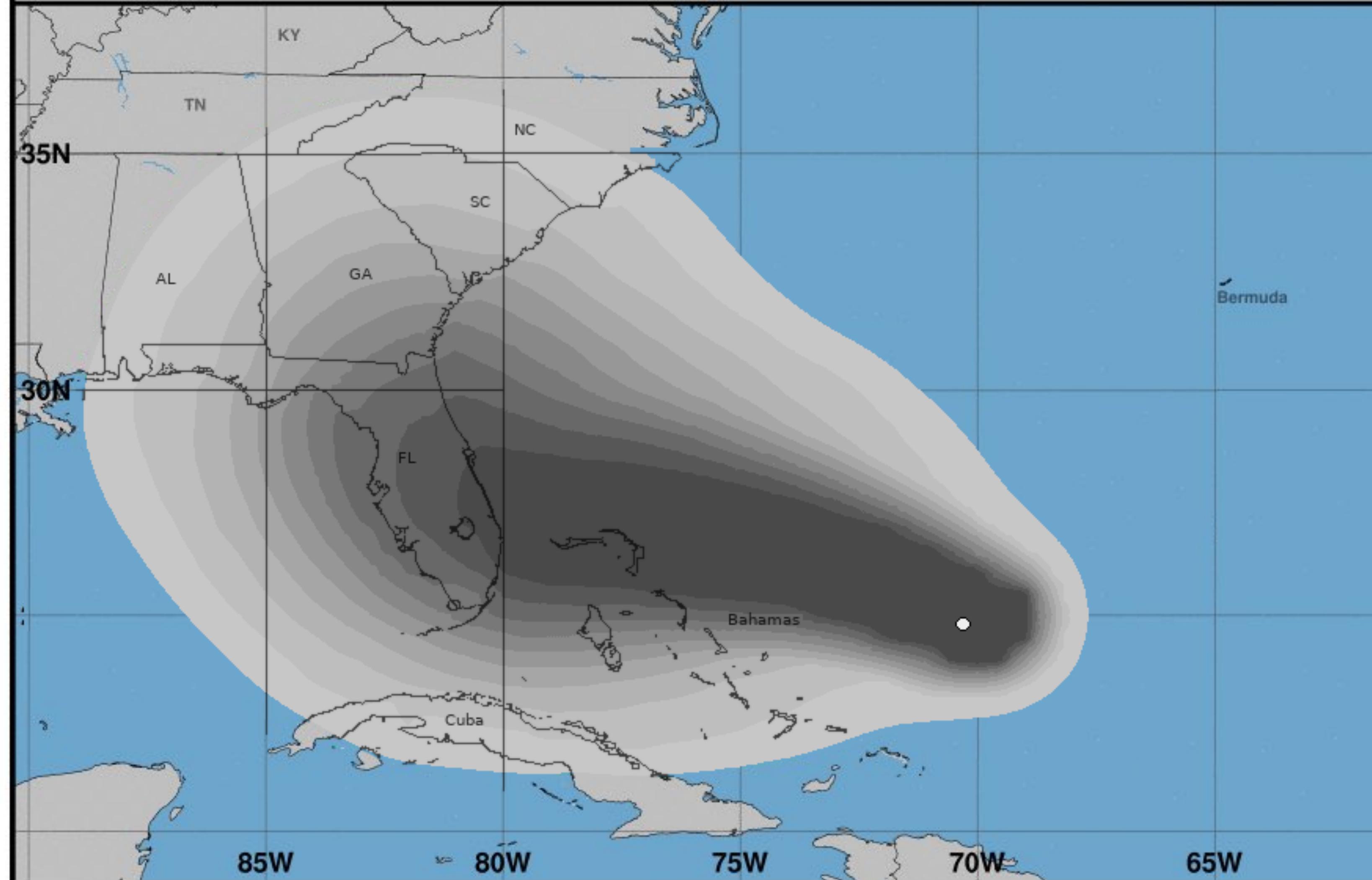
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





Tropical-Storm-Force Wind Speed Probabilities

For the 120 hours (5.00 days) from 2 PM EDT FRI AUG 30 to 2 PM EDT WED SEP 04



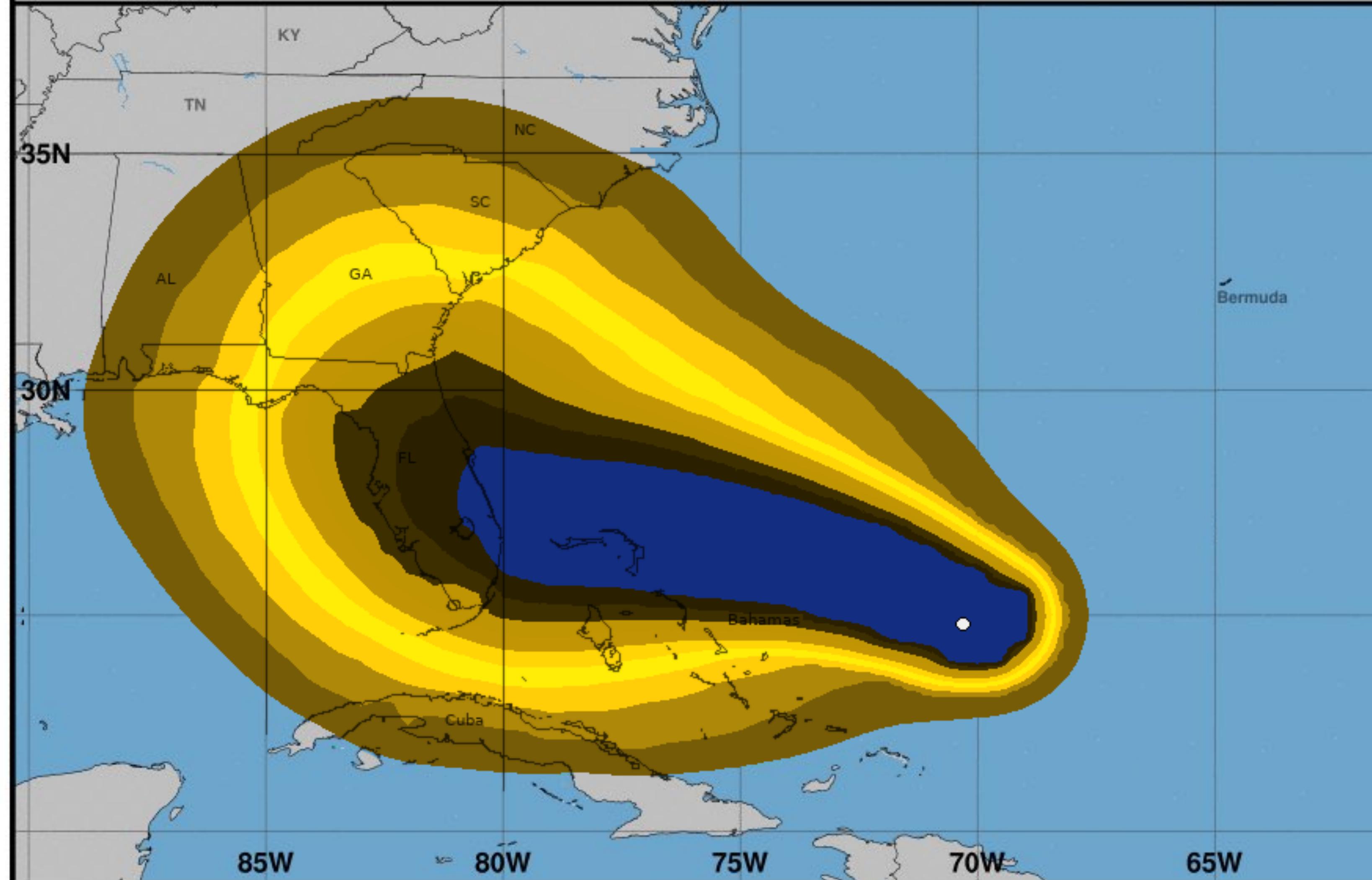
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





Tropical-Storm-Force Wind Speed Probabilities

For the 120 hours (5.00 days) from 2 PM EDT FRI AUG 30 to 2 PM EDT WED SEP 04



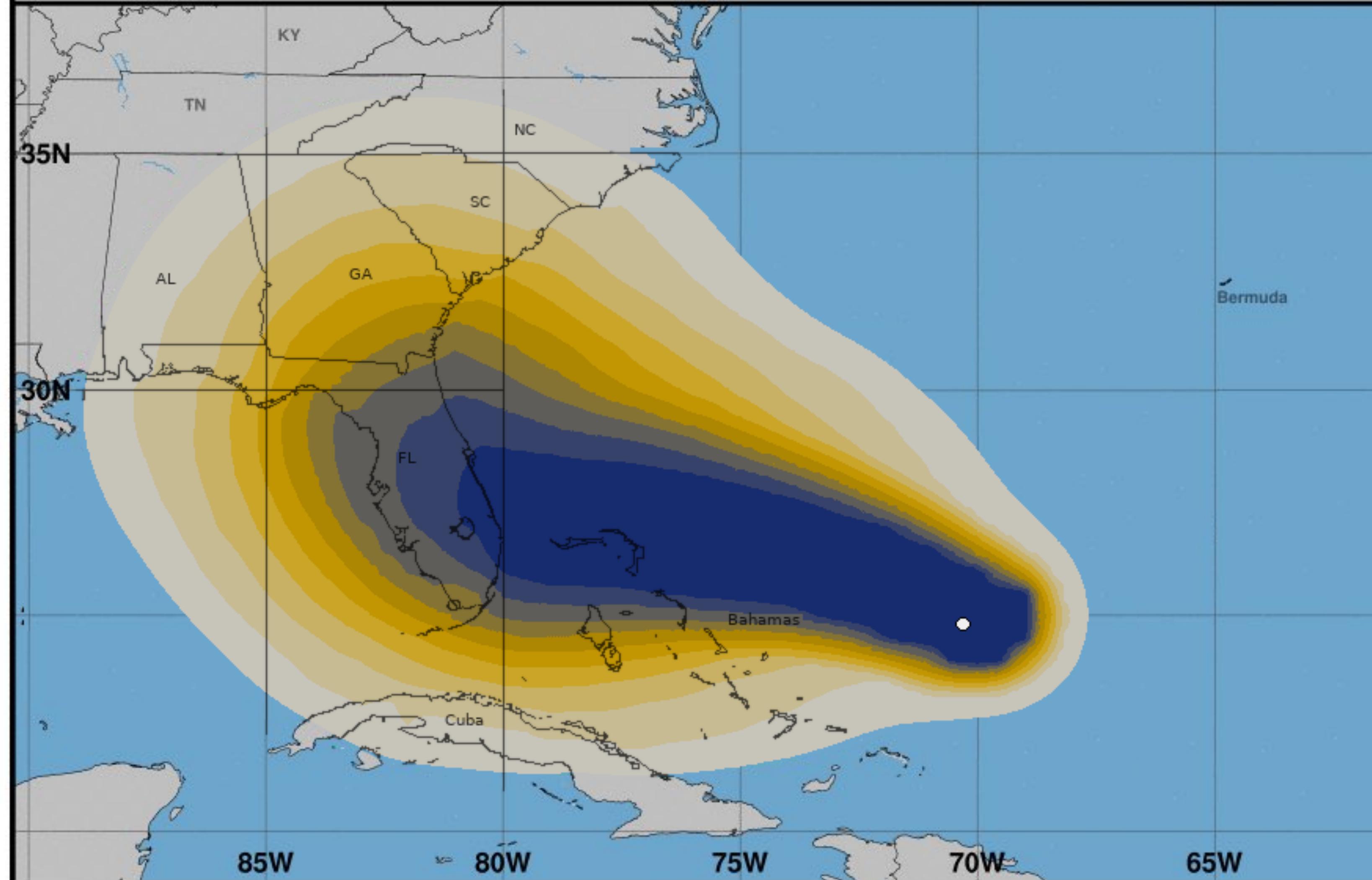
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





Tropical-Storm-Force Wind Speed Probabilities

For the 120 hours (5.00 days) from 2 PM EDT FRI AUG 30 to 2 PM EDT WED SEP 04

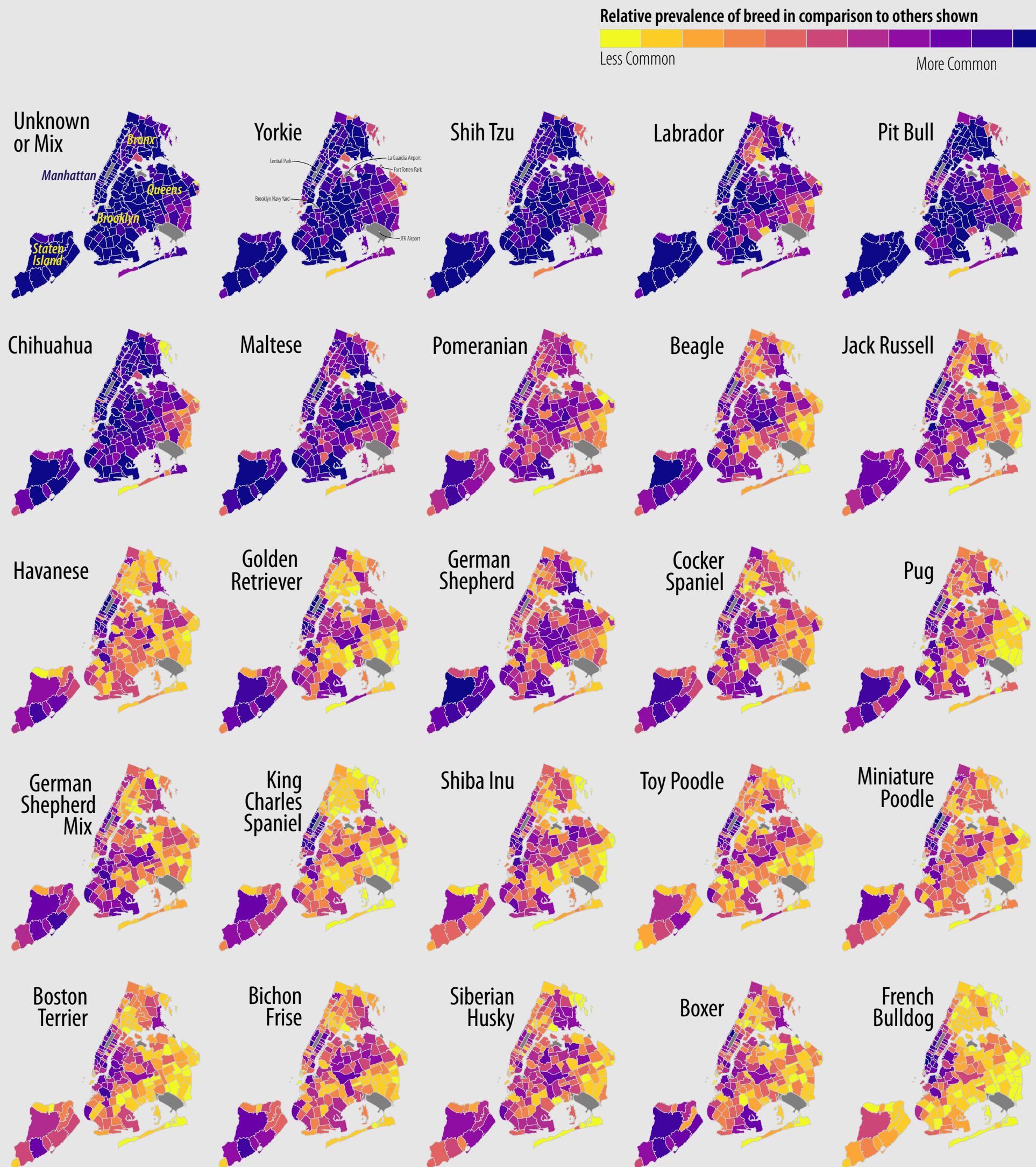


Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)



Dogs of New York

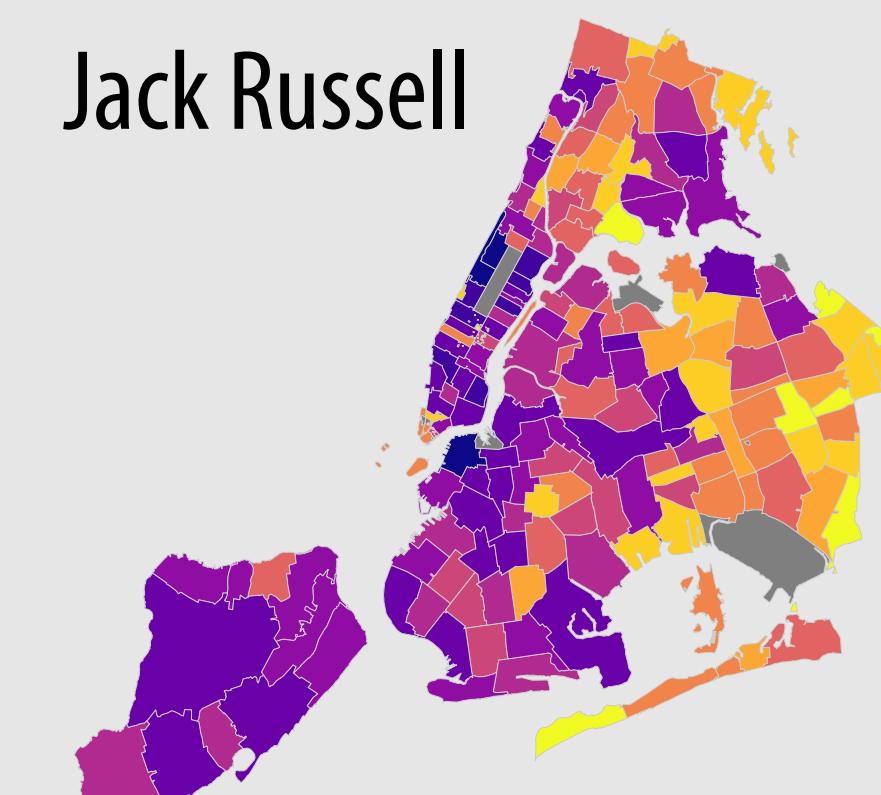
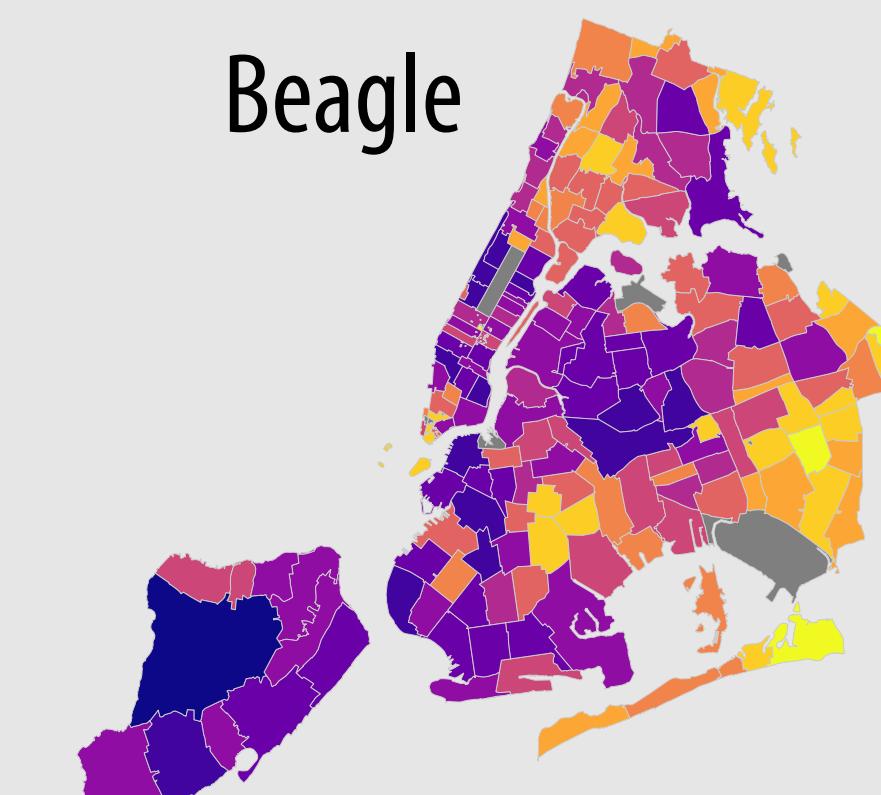
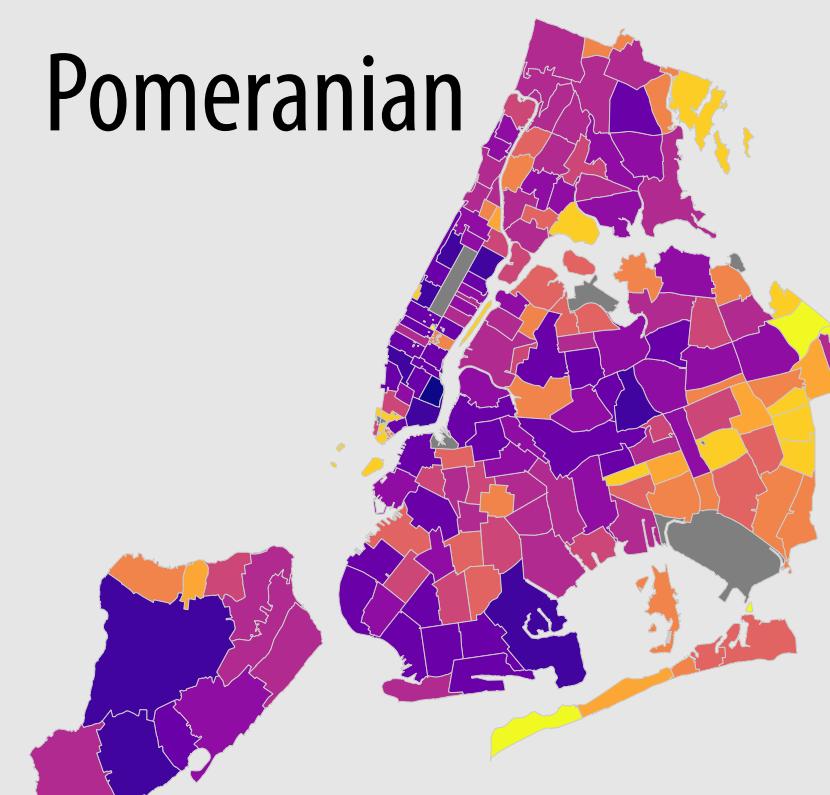
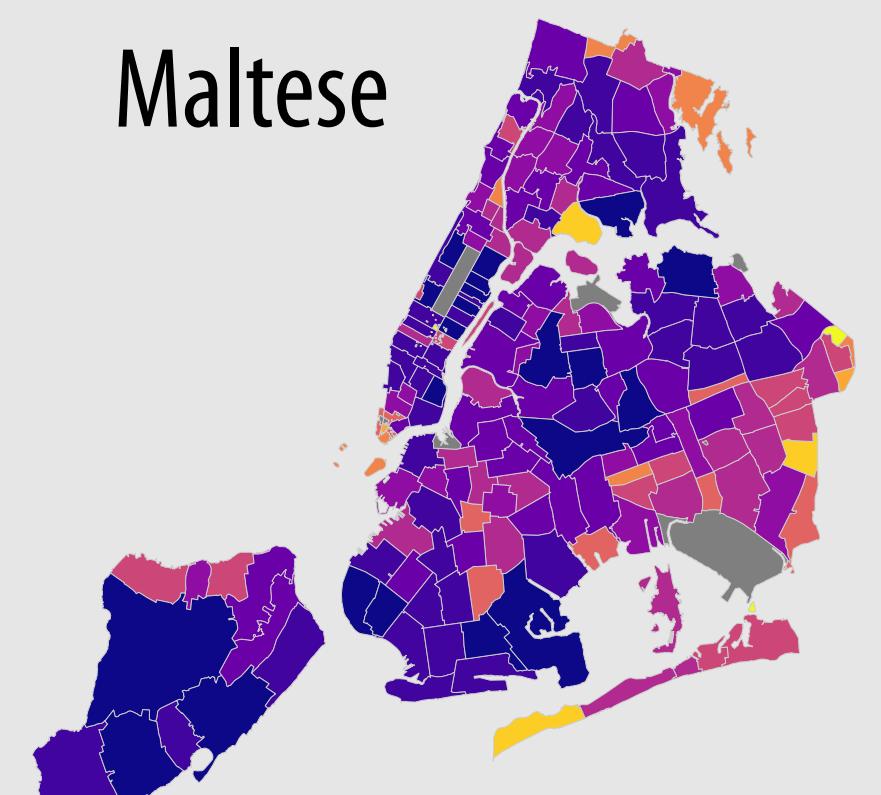
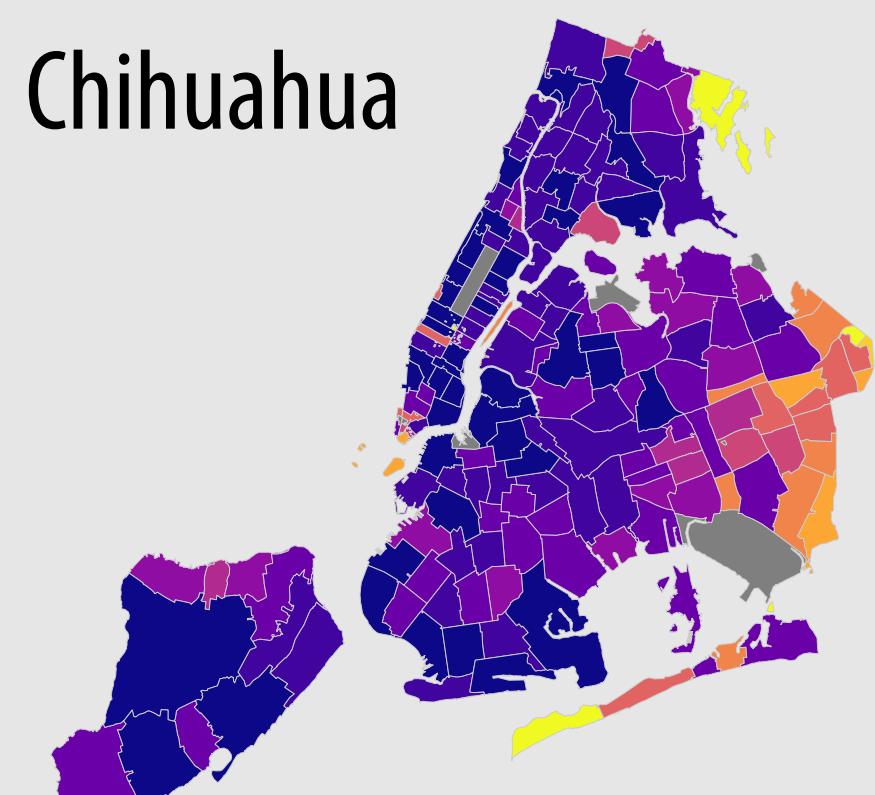
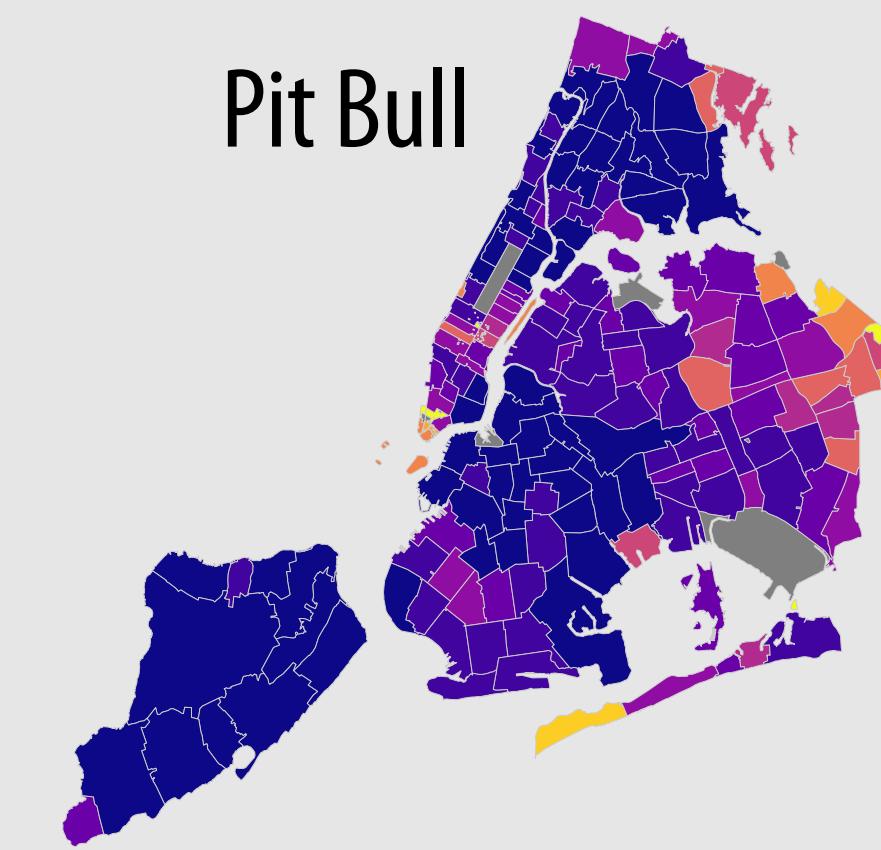
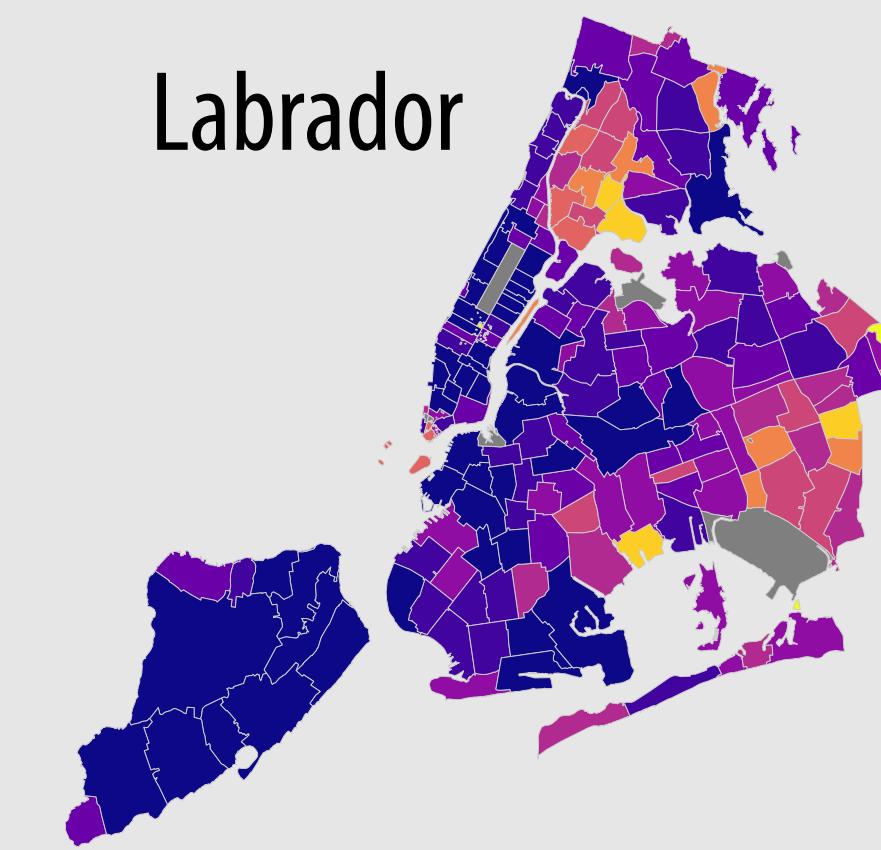
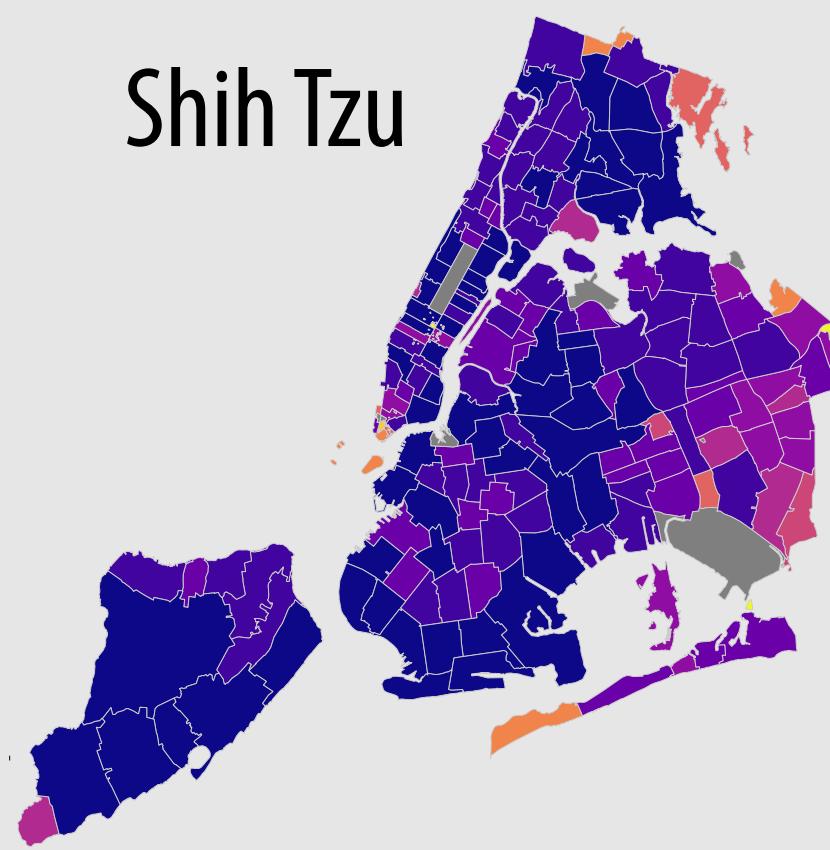
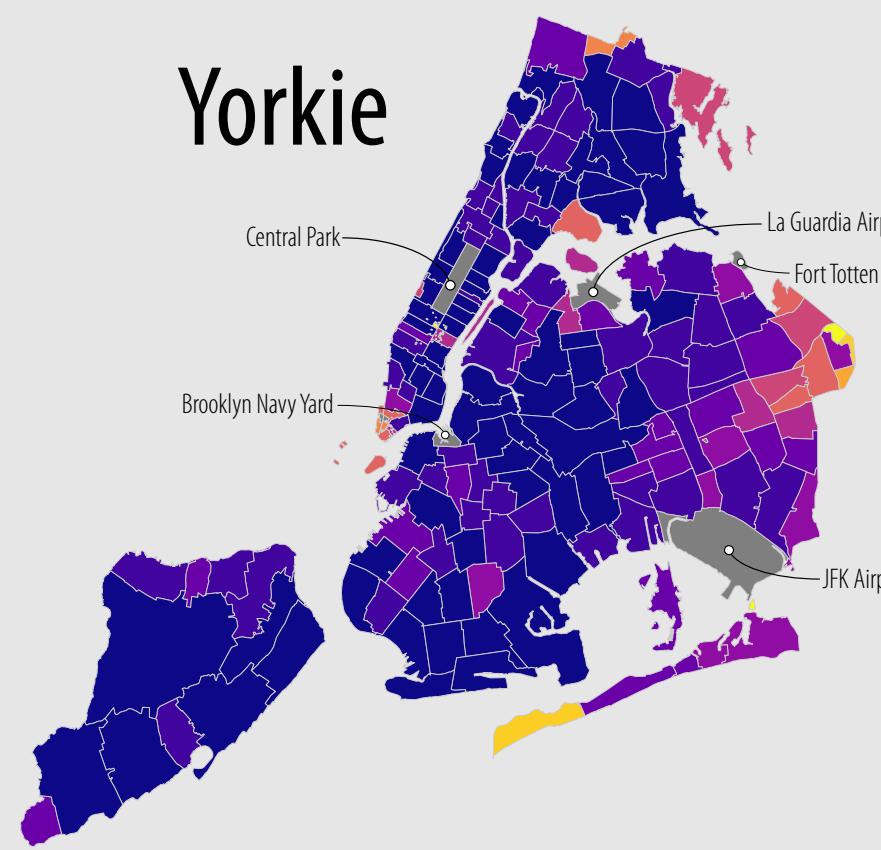
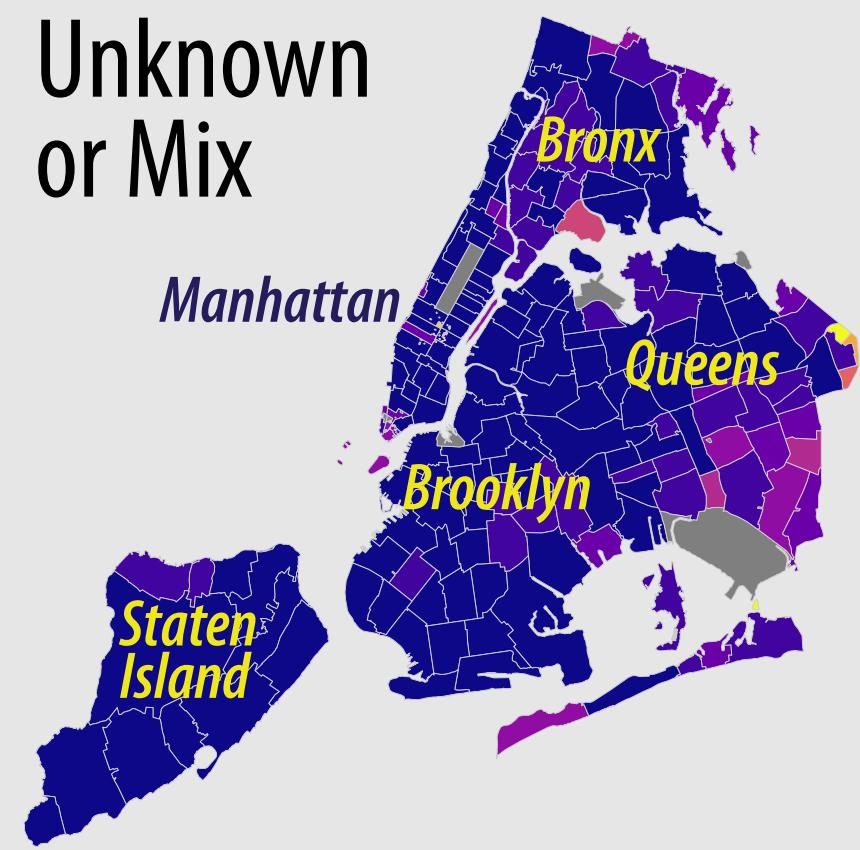
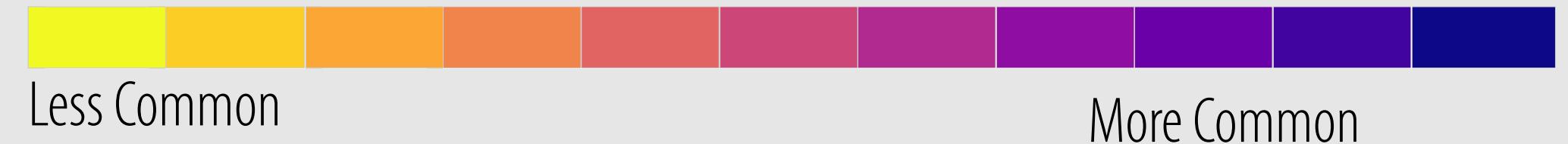
Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.



Dog Breeds in New York

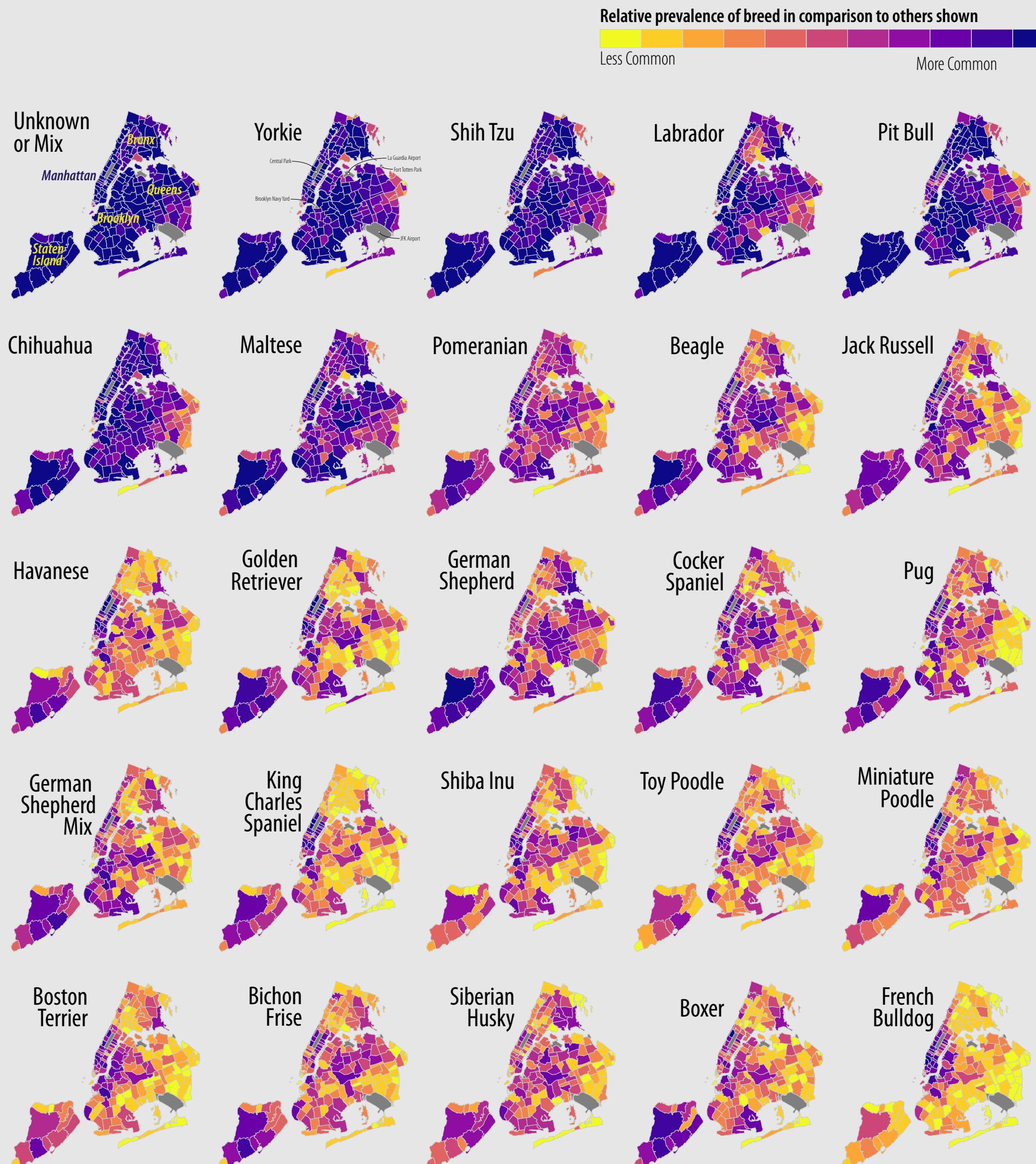
Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.

Relative prevalence of breed in comparison to others shown



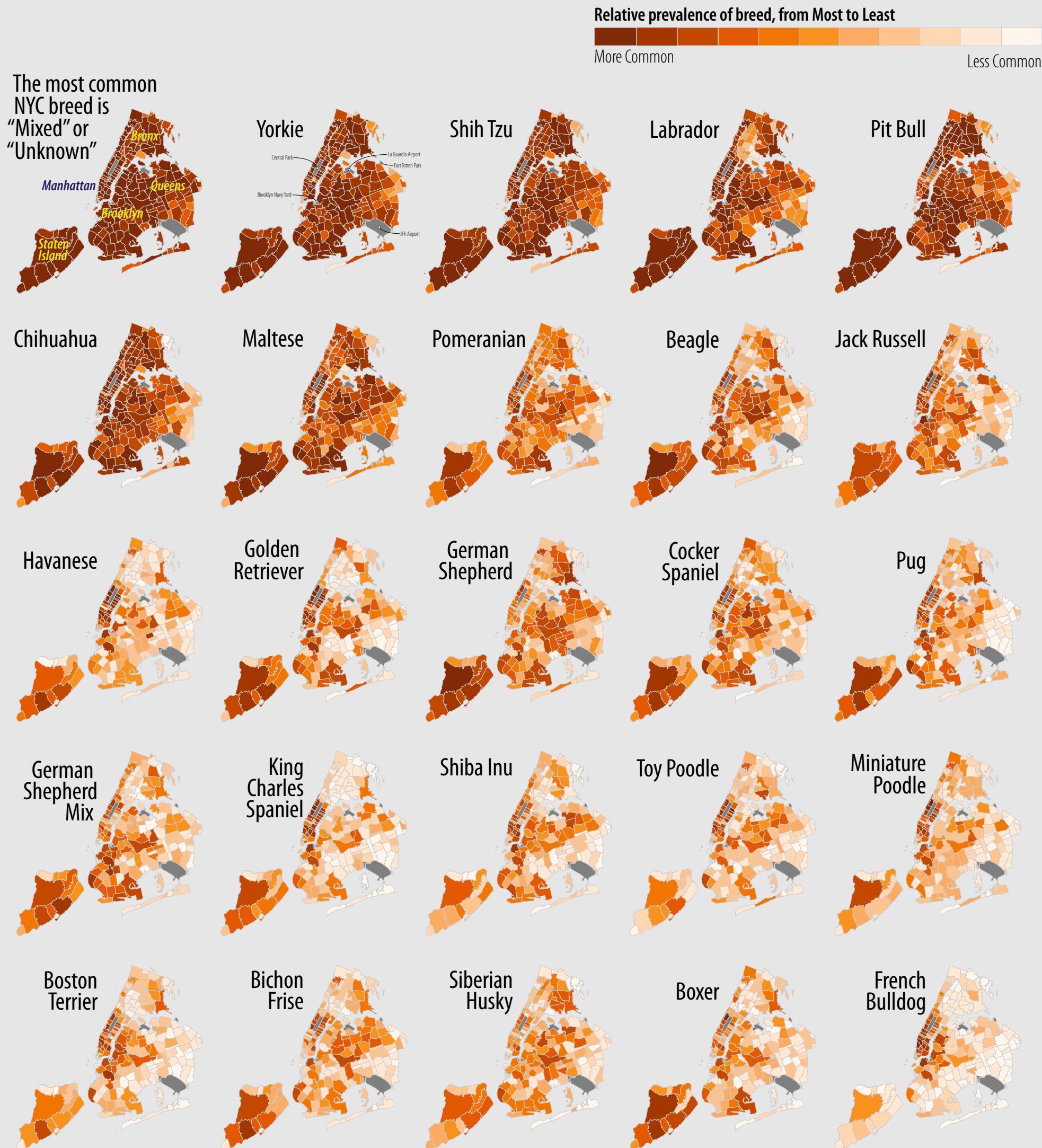
Dogs of New York

Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.



Dogs of New York

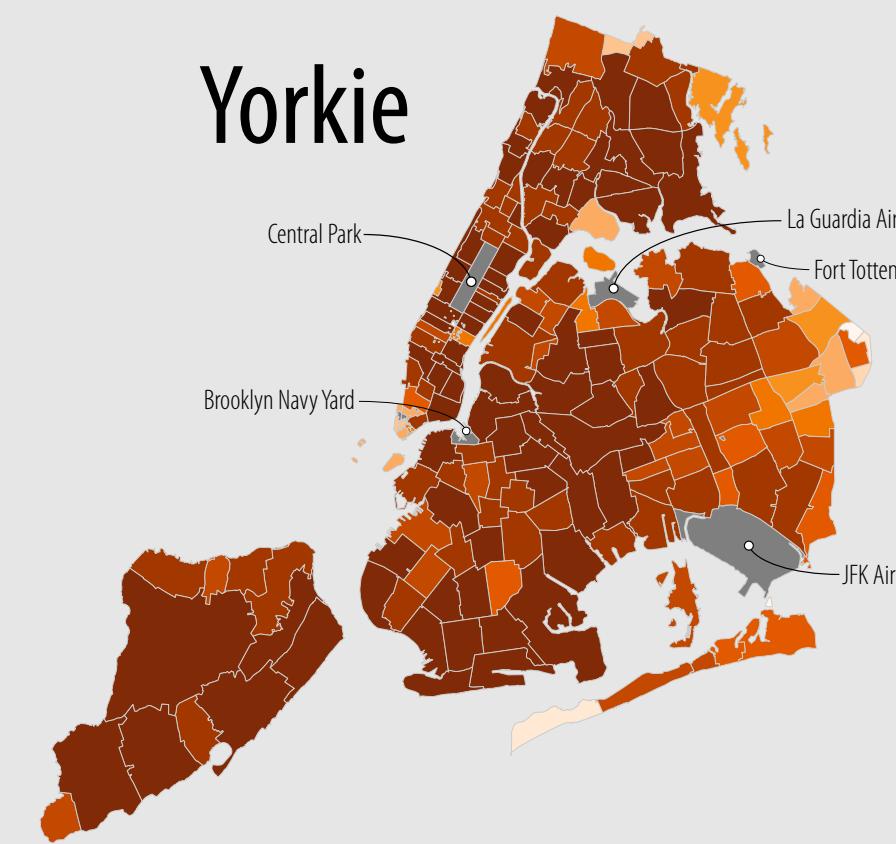
Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.



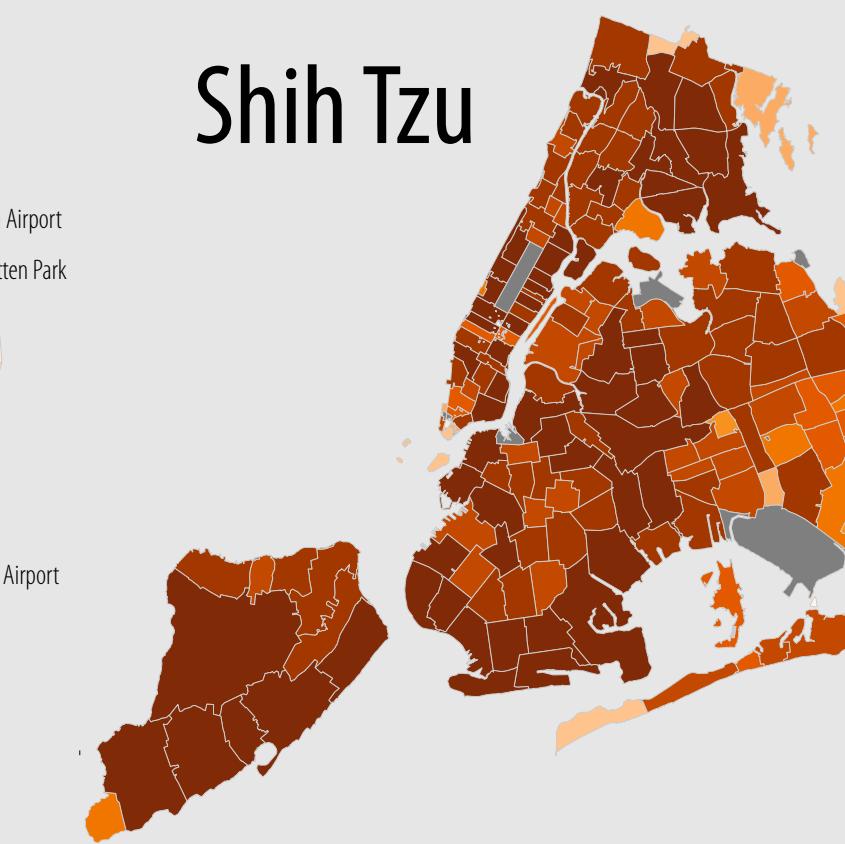
The most common
NYC breed is
“Mixed” or
“Unknown”



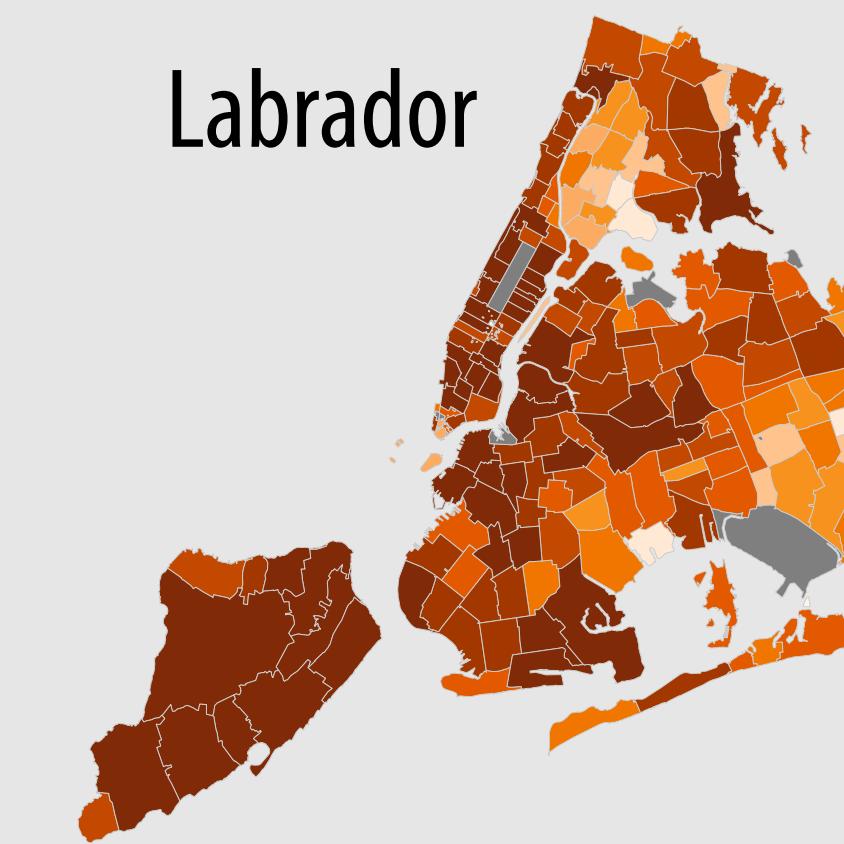
Yorkie



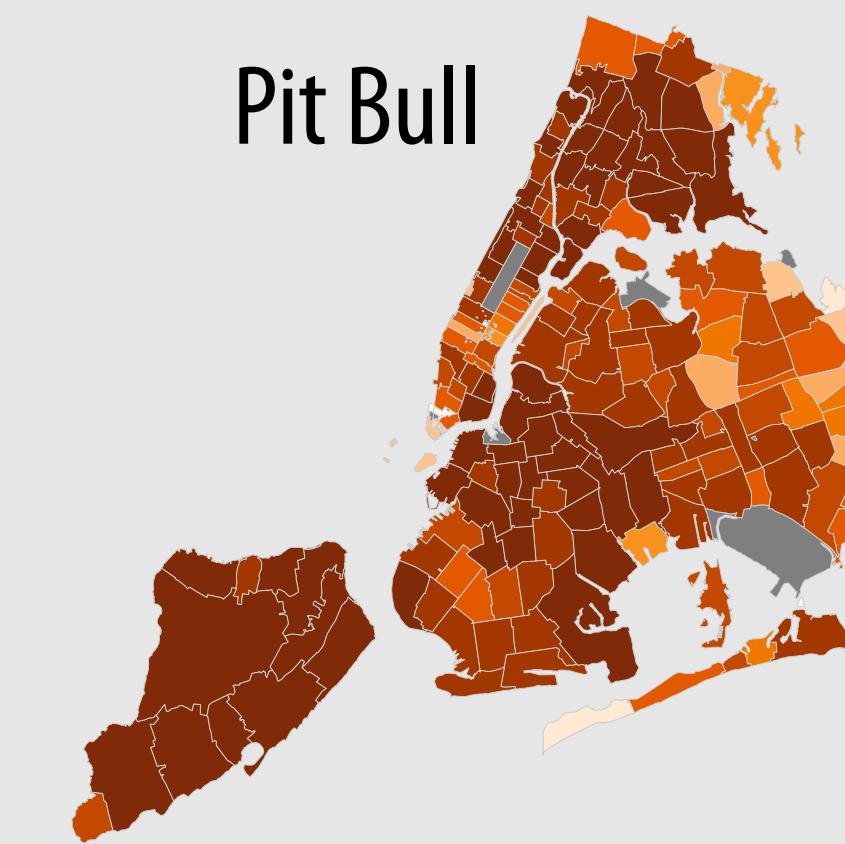
Shih Tzu



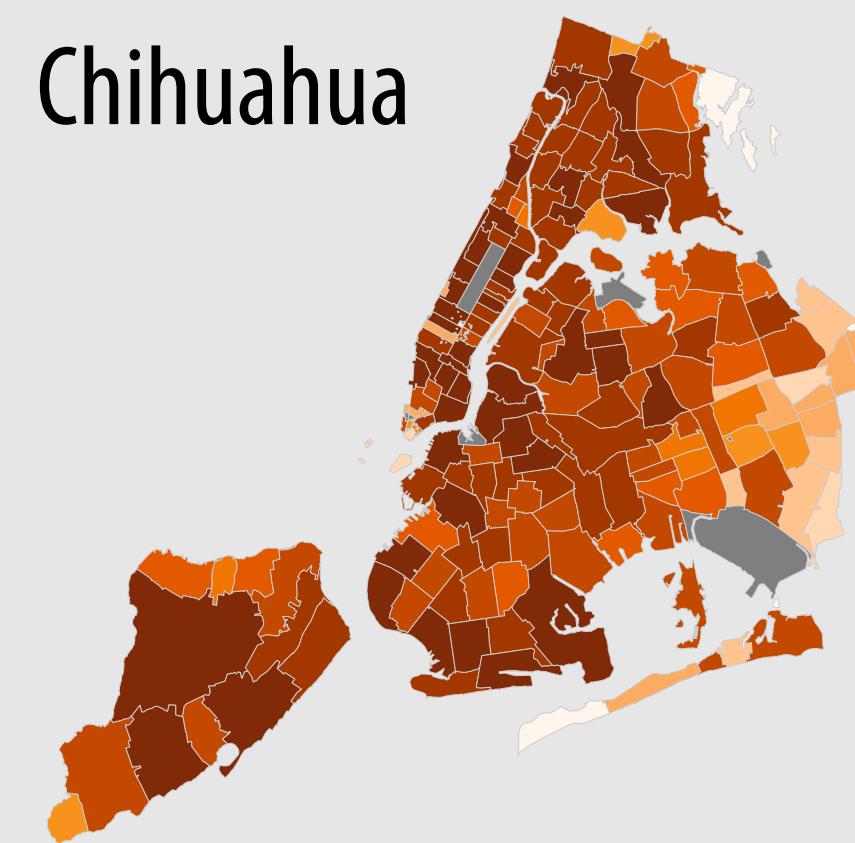
Labrador



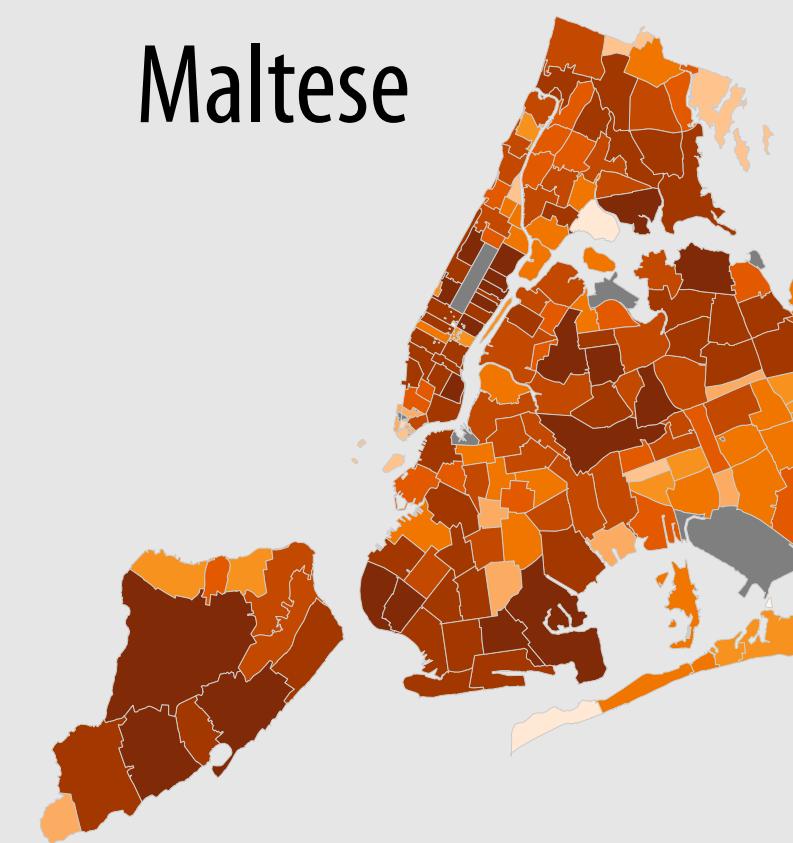
Pit Bull



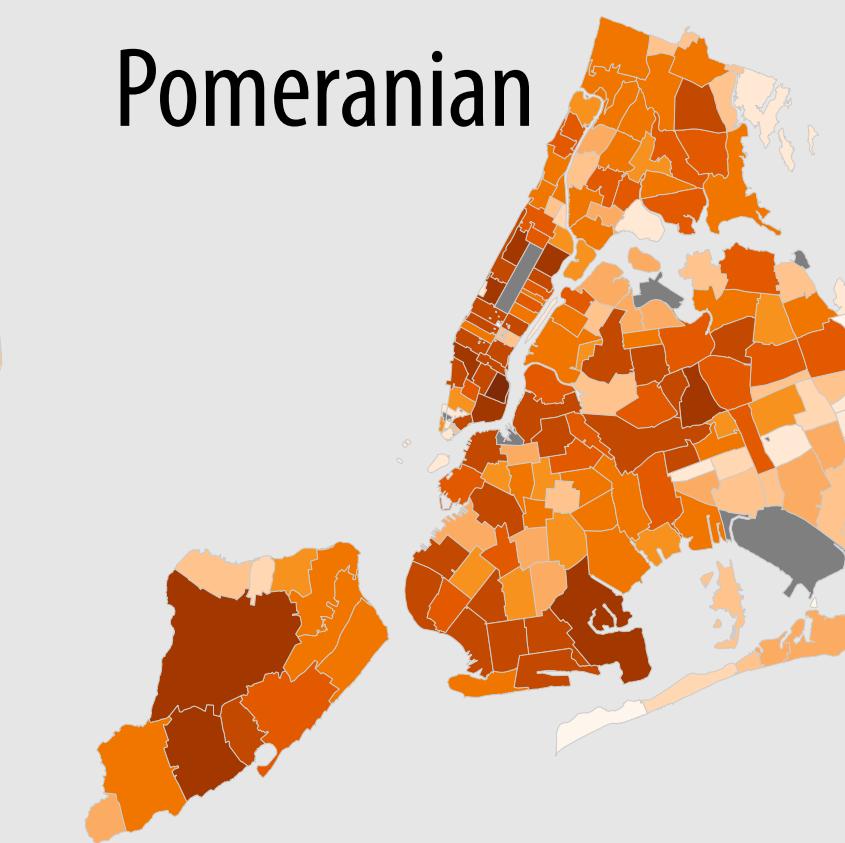
Chihuahua



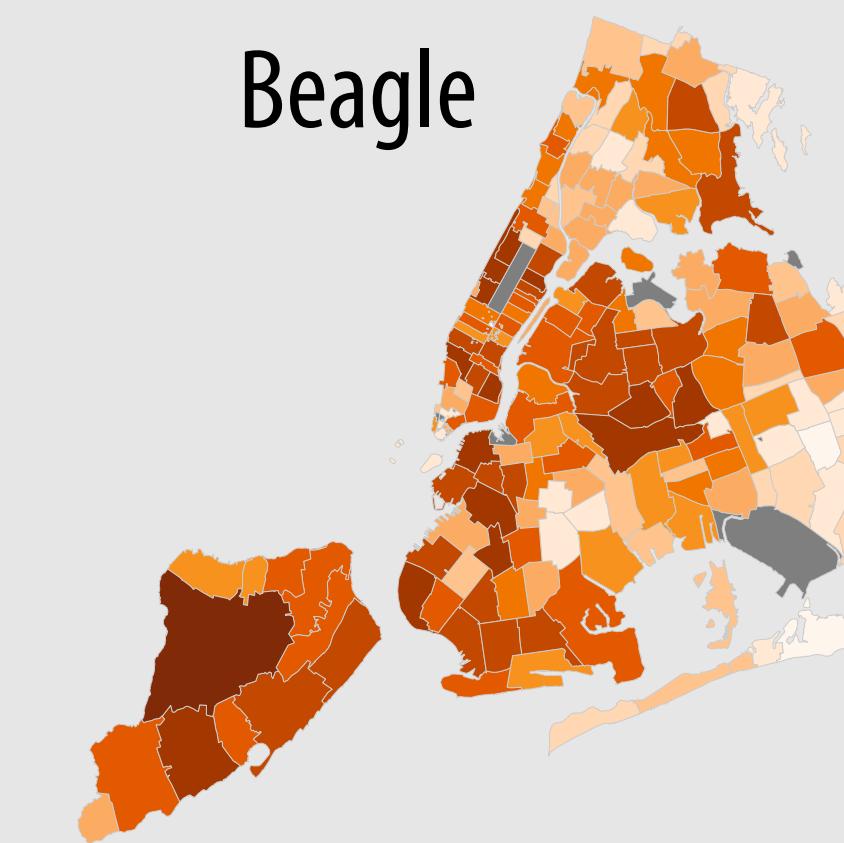
Maltese



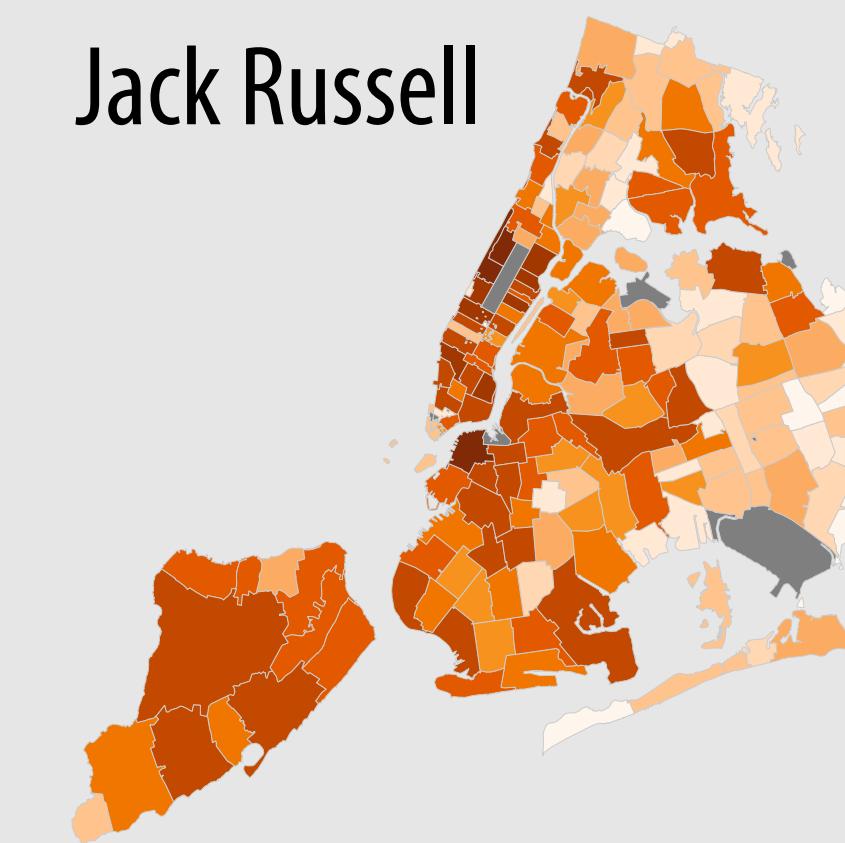
Pomeranian



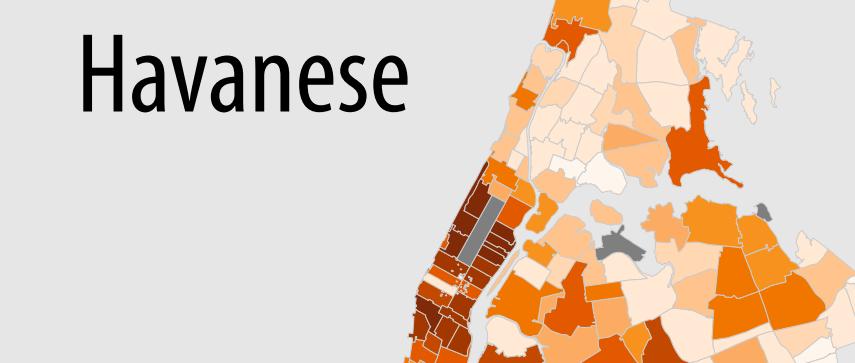
Beagle



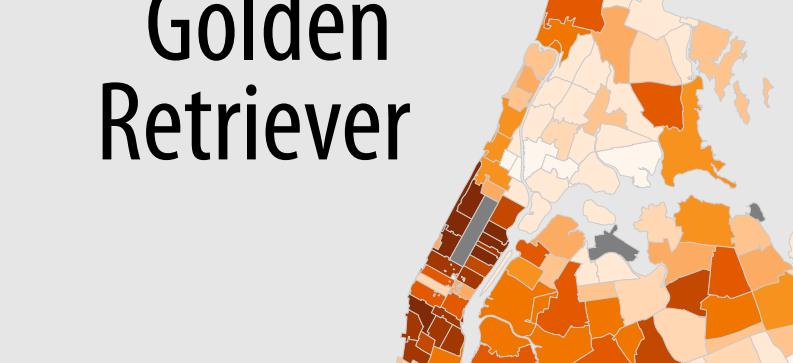
Jack Russell



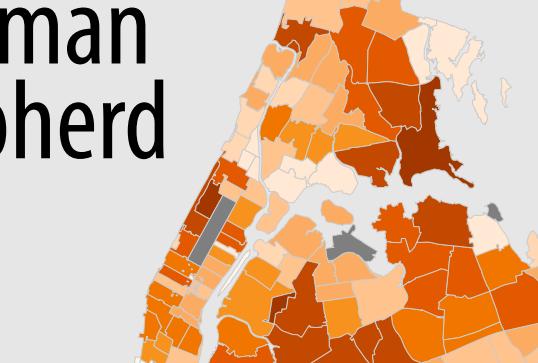
Havanese



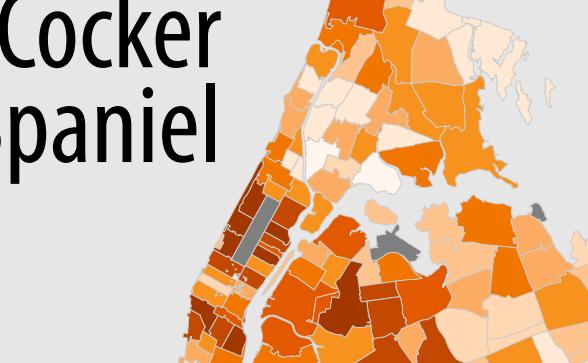
Golden Retriever



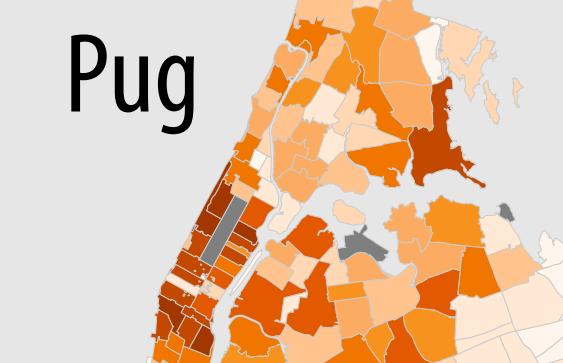
German Shepherd



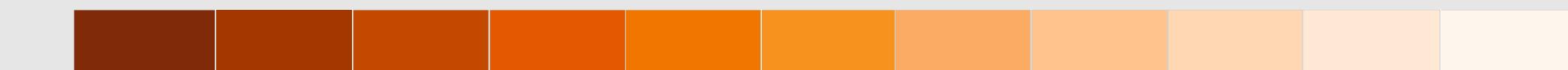
Cocker Spaniel



Pug



Relative prevalence of breed, from Most to Least

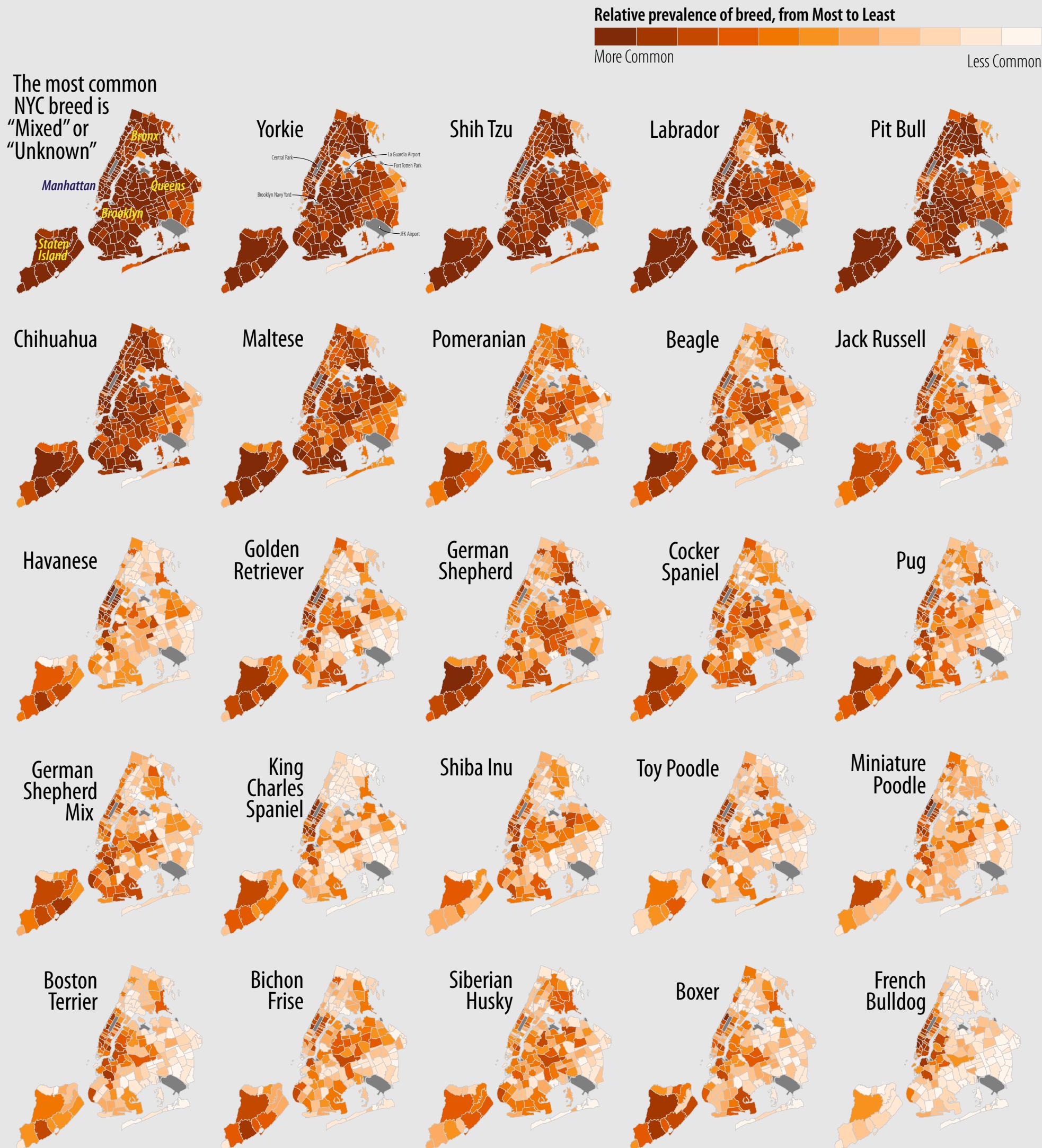


More Common

Less Common

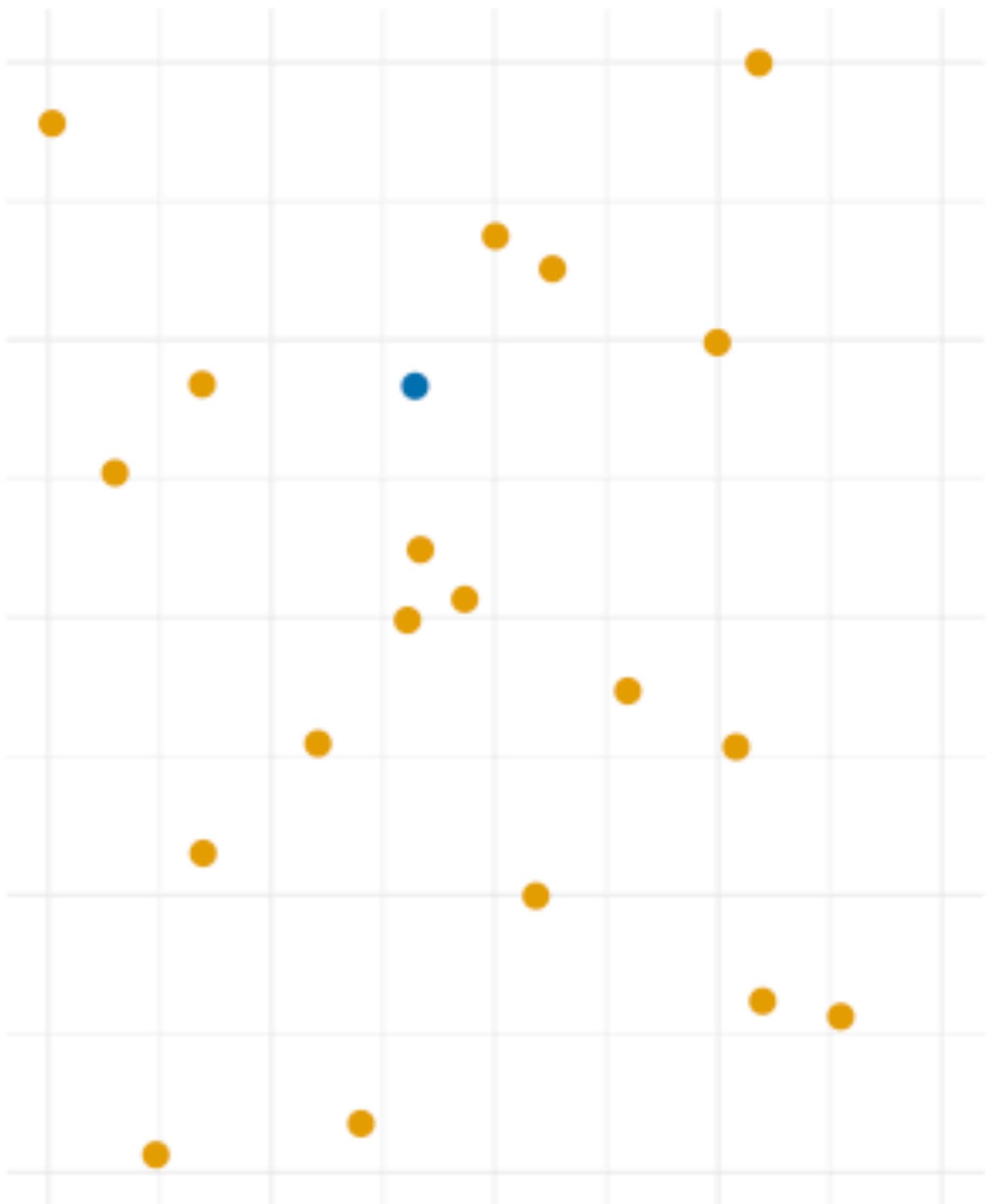
Dogs of New York

Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.

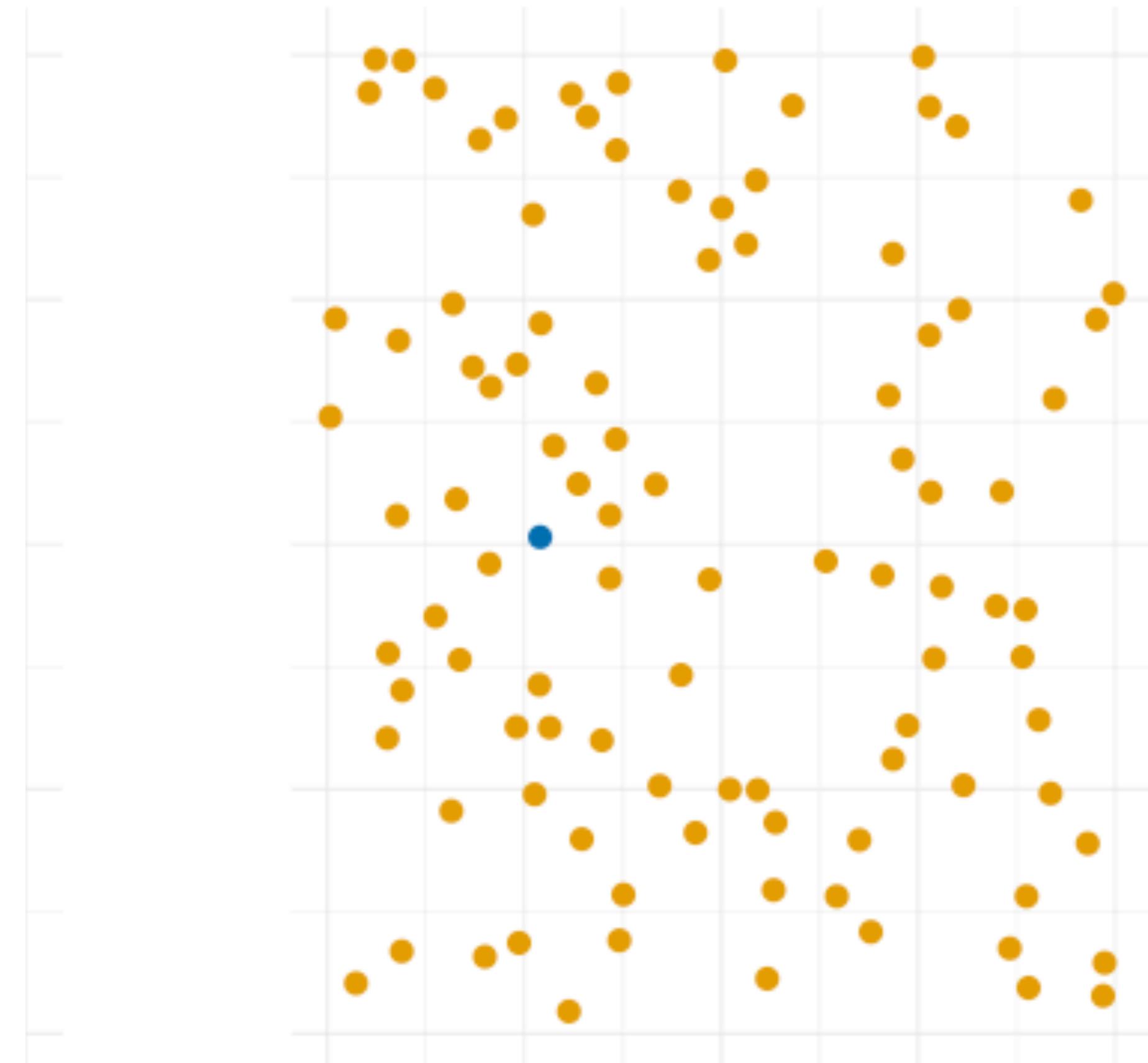


PRE-ATTENTIVE PROCESSING

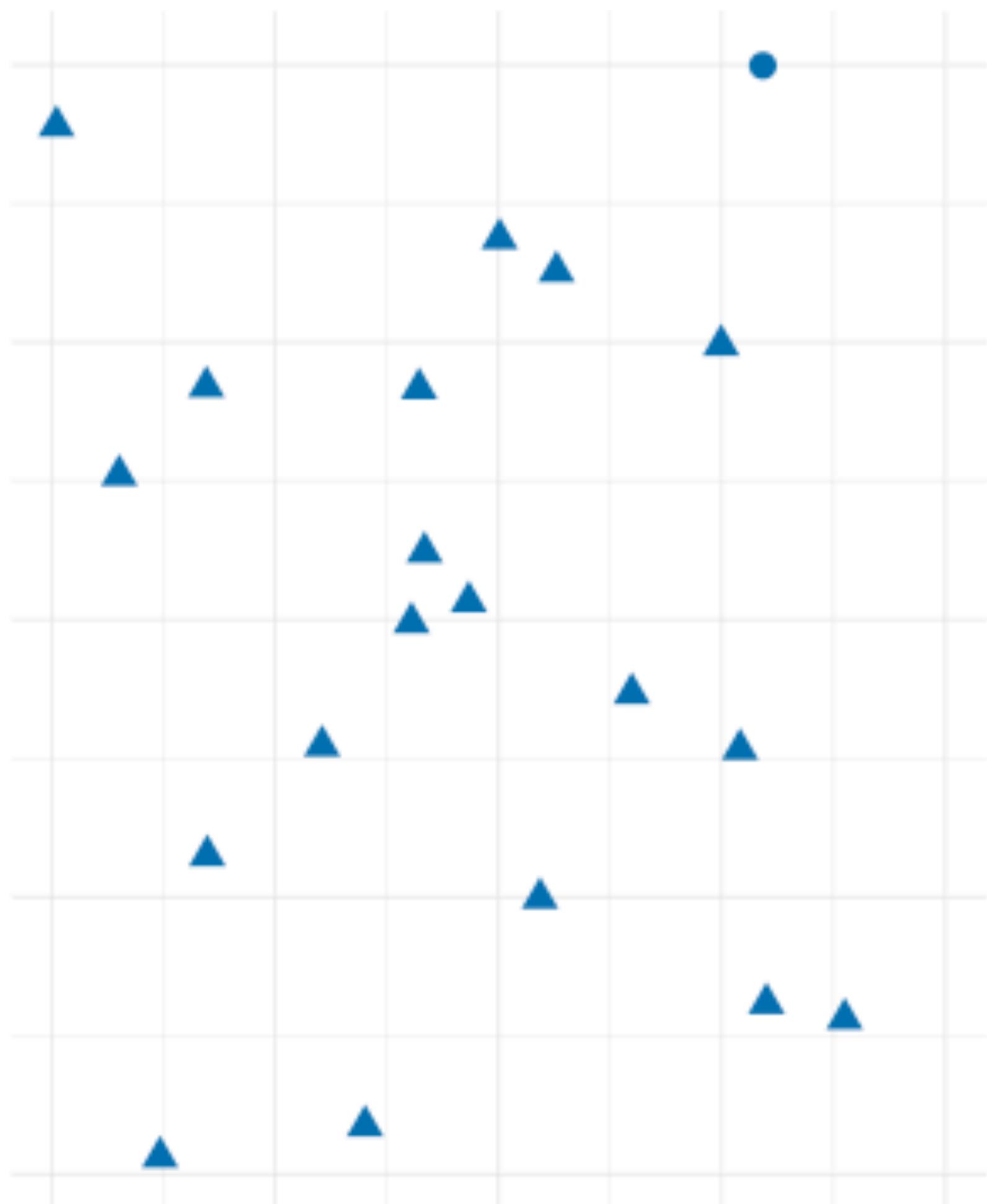
Color Only, N=20



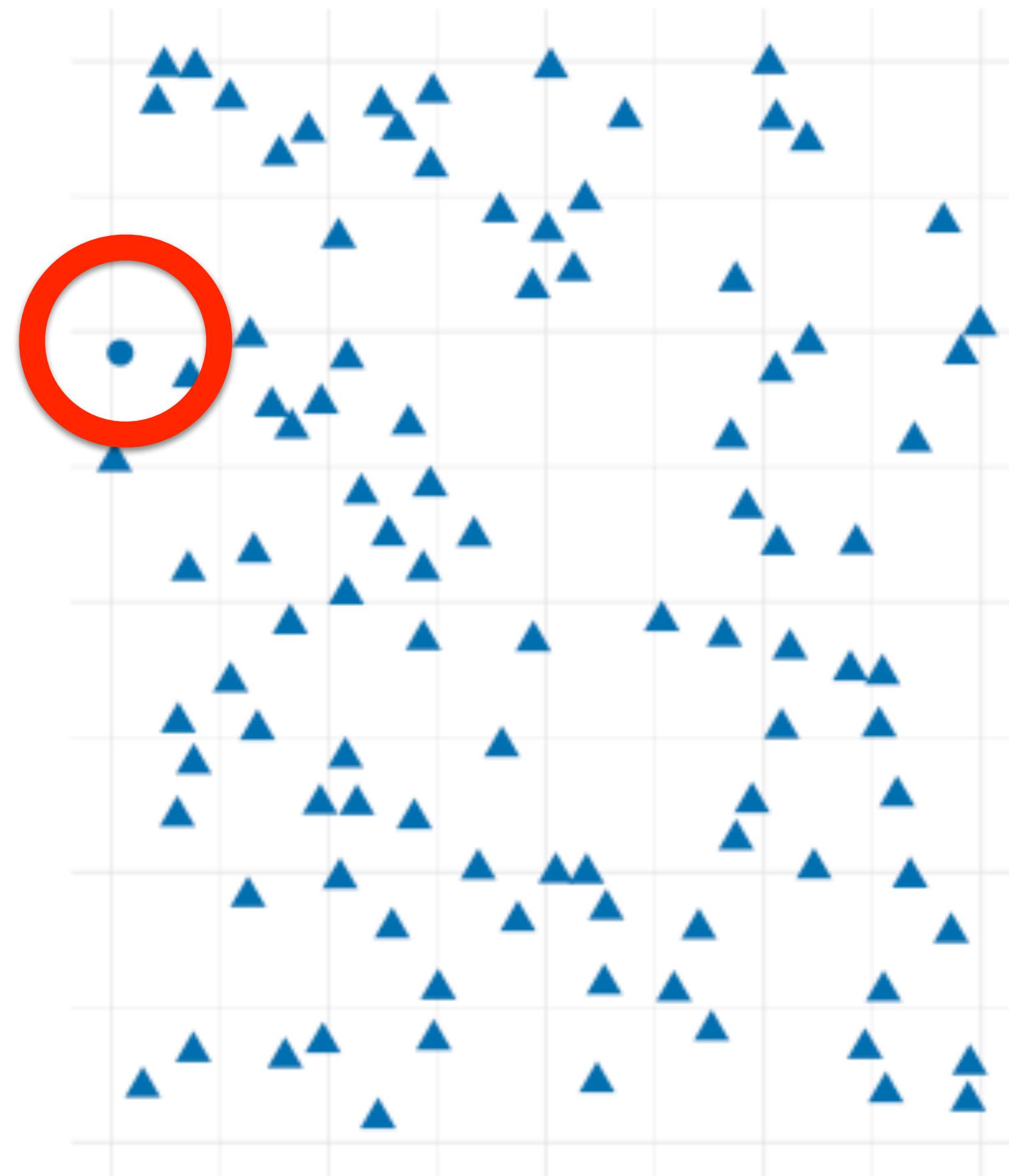
Color Only, N=100



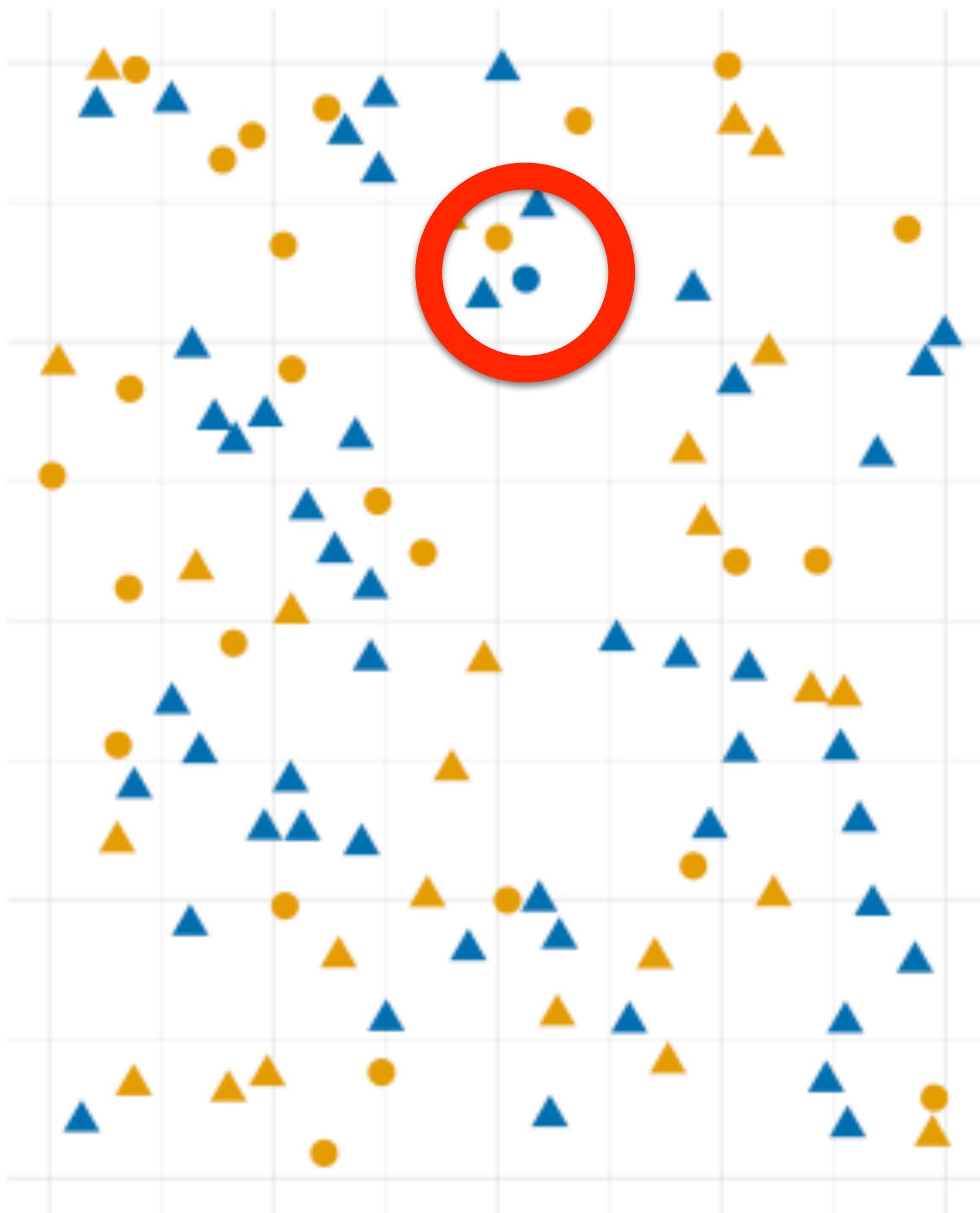
Shape Only, N=20

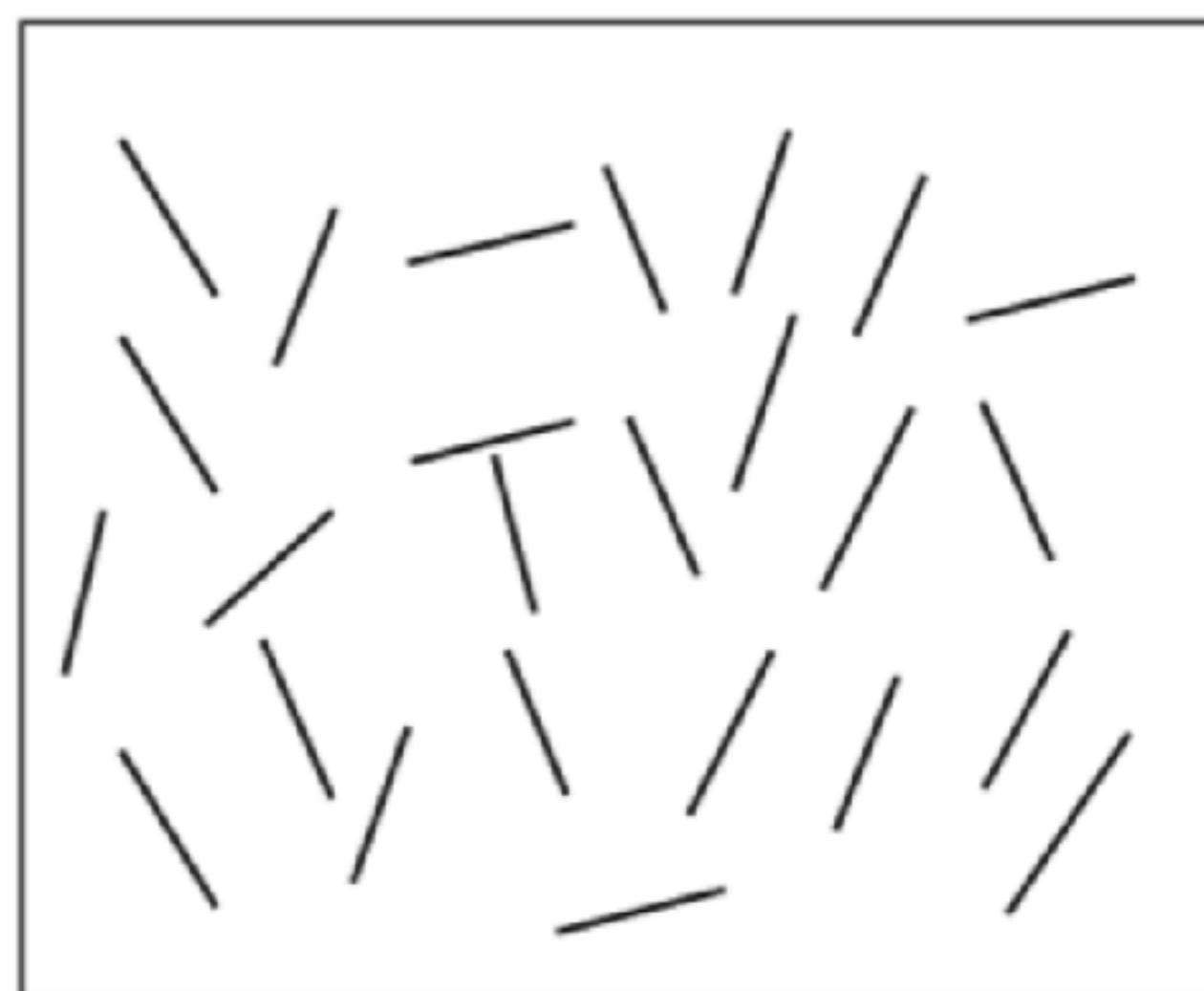
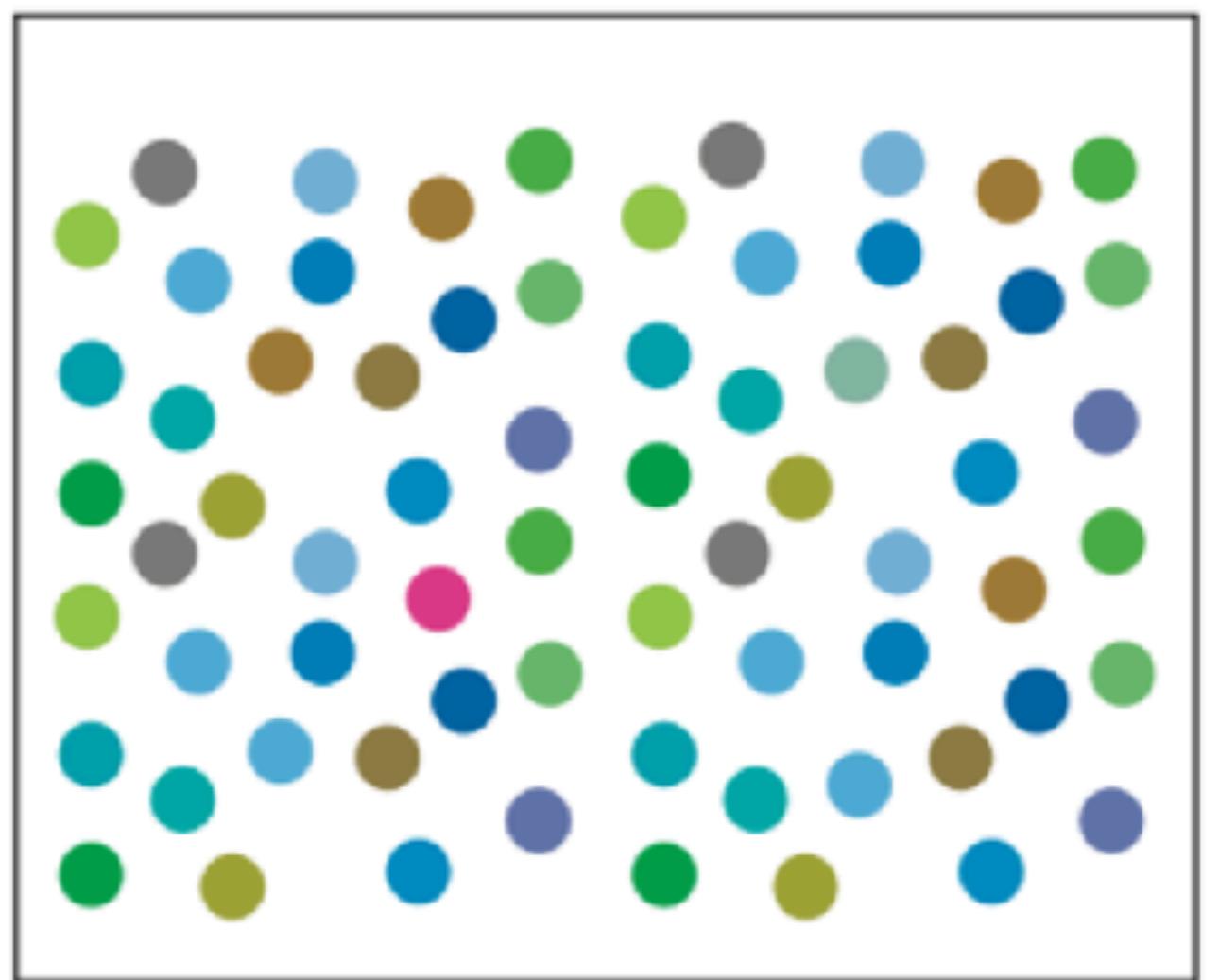
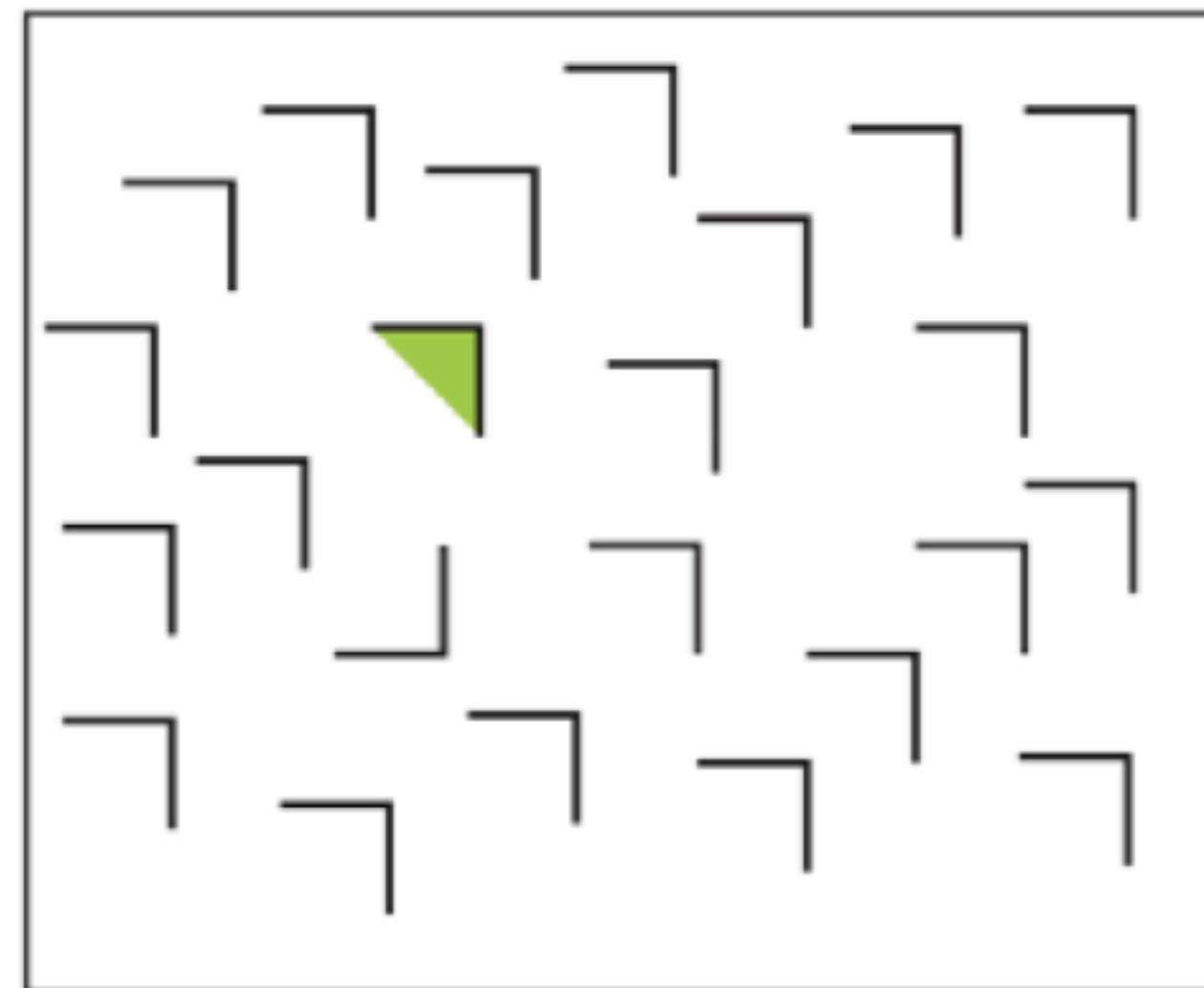
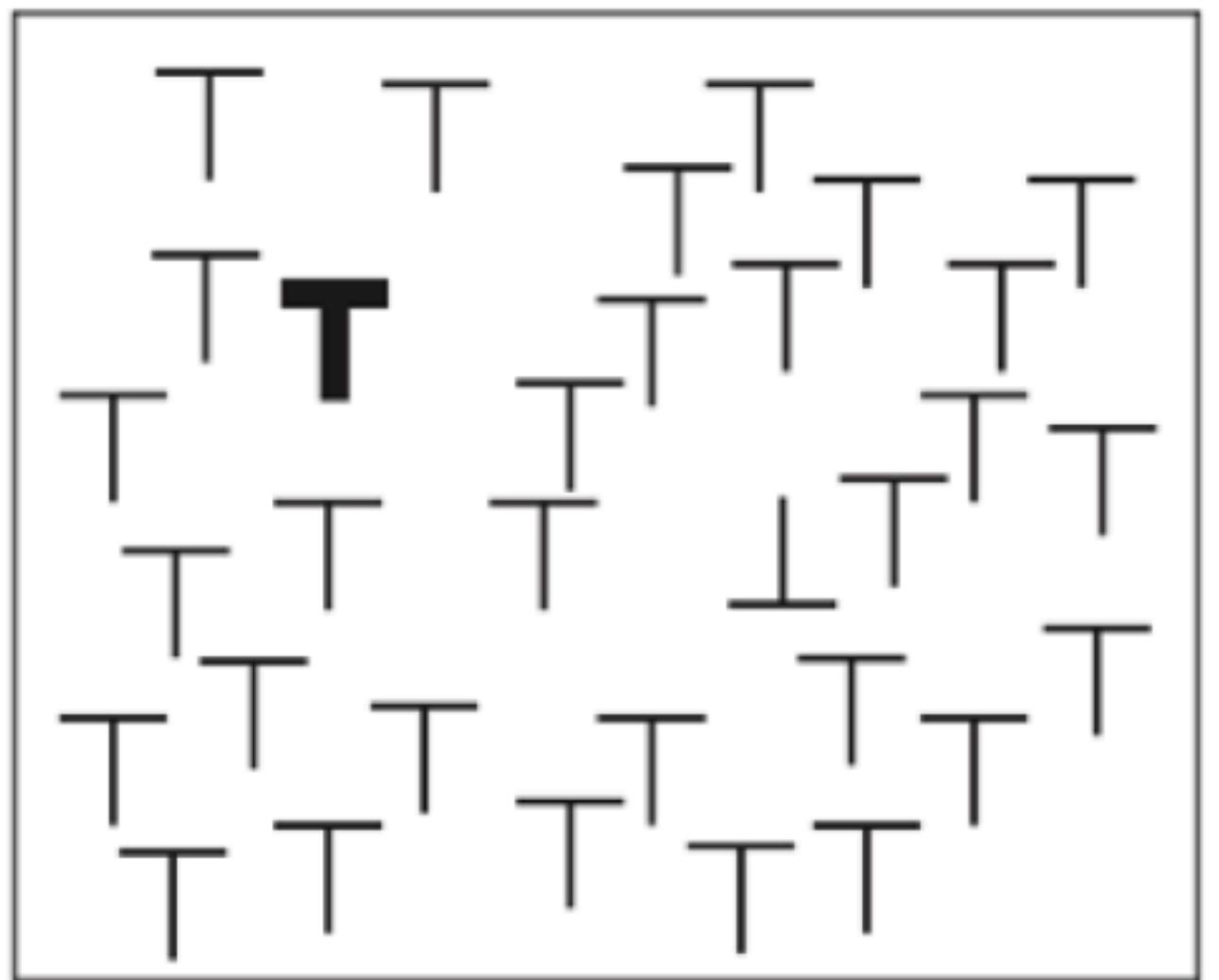


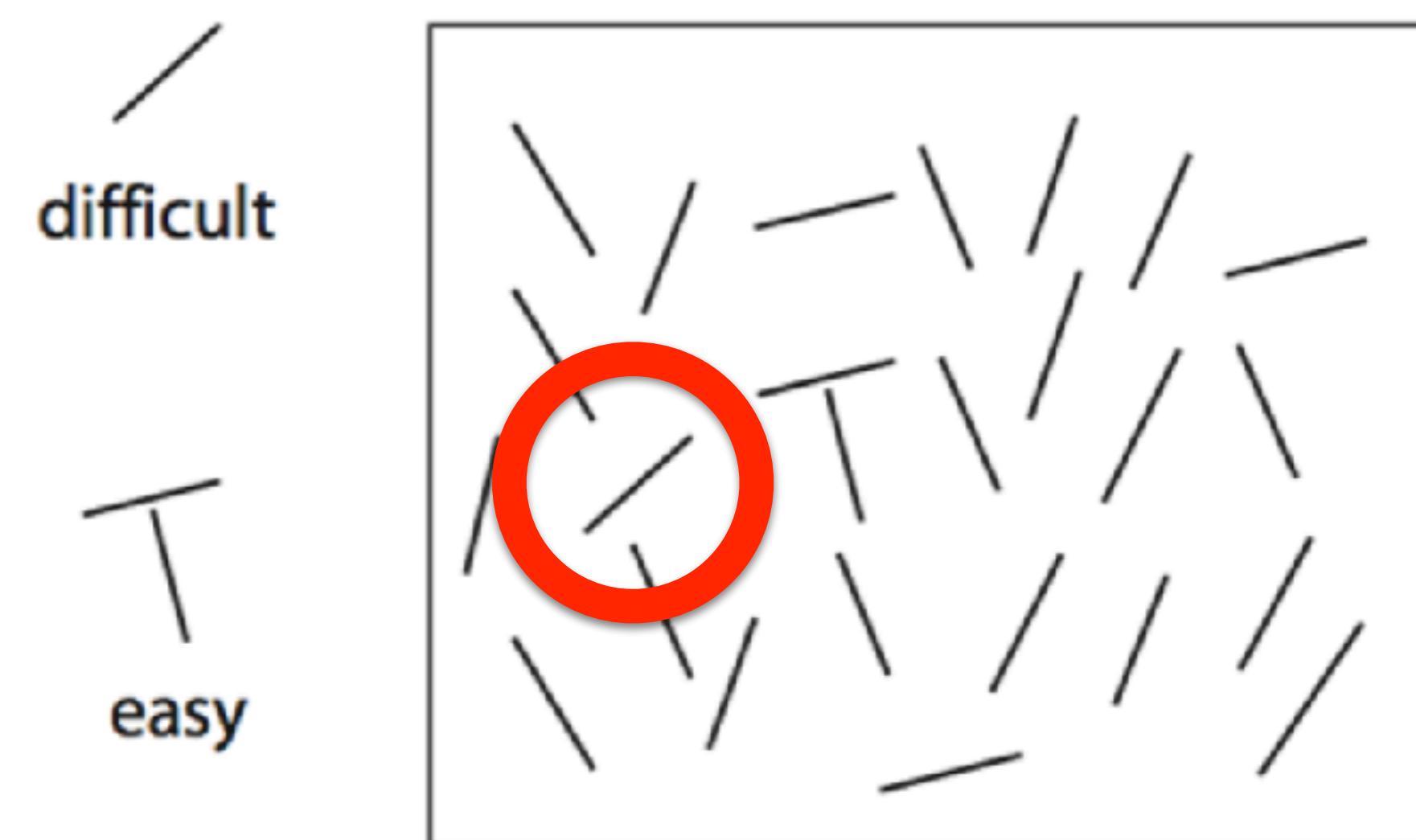
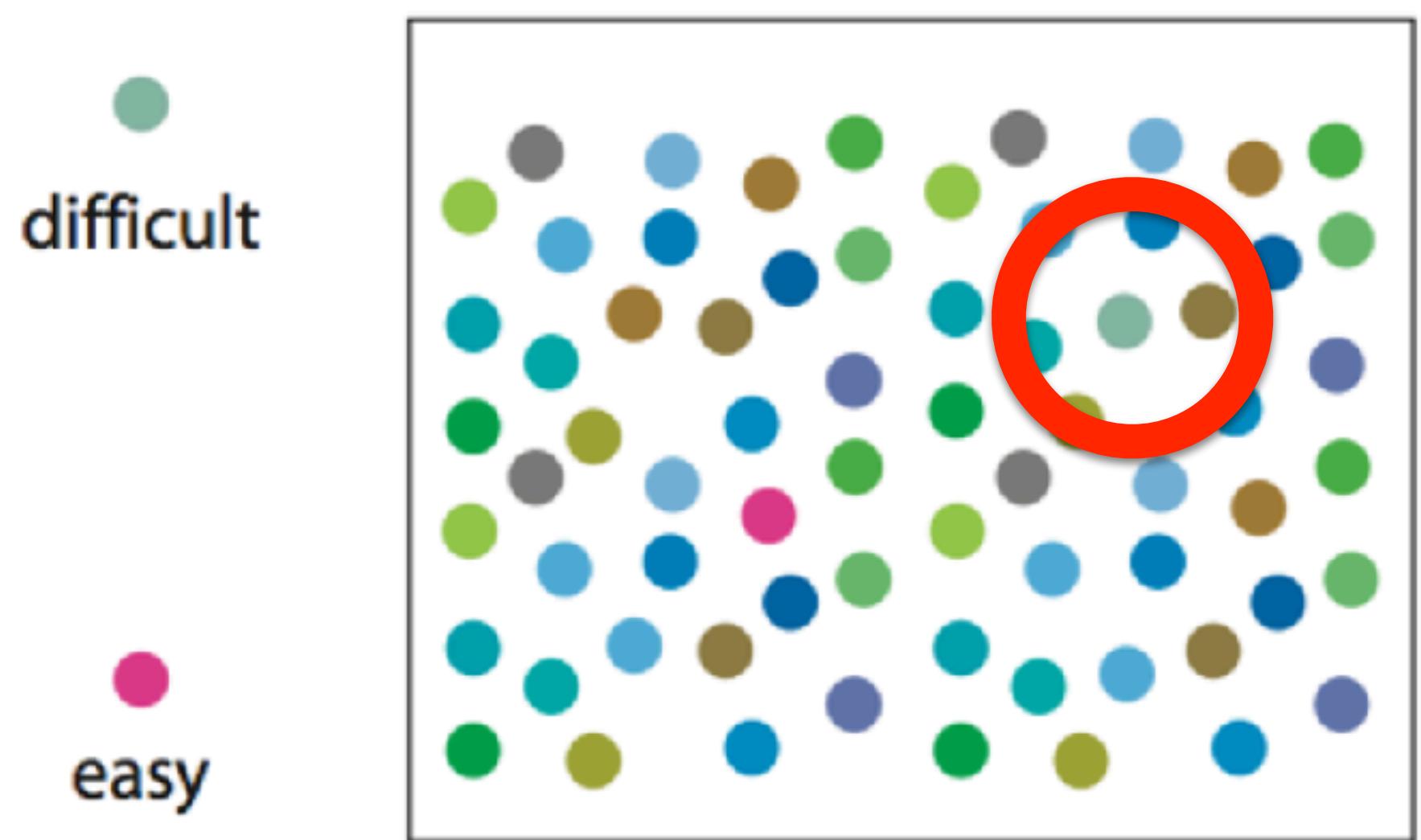
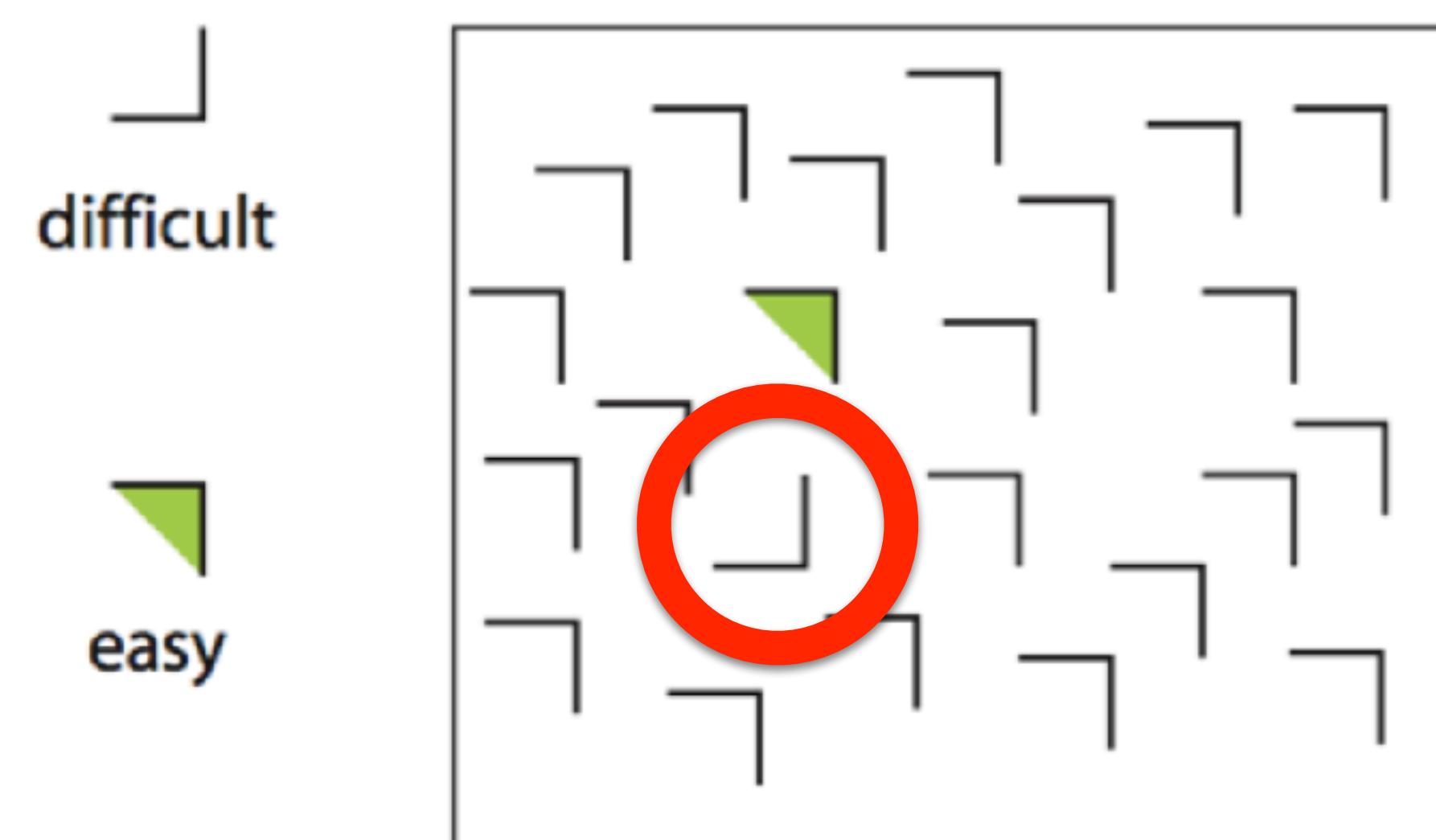
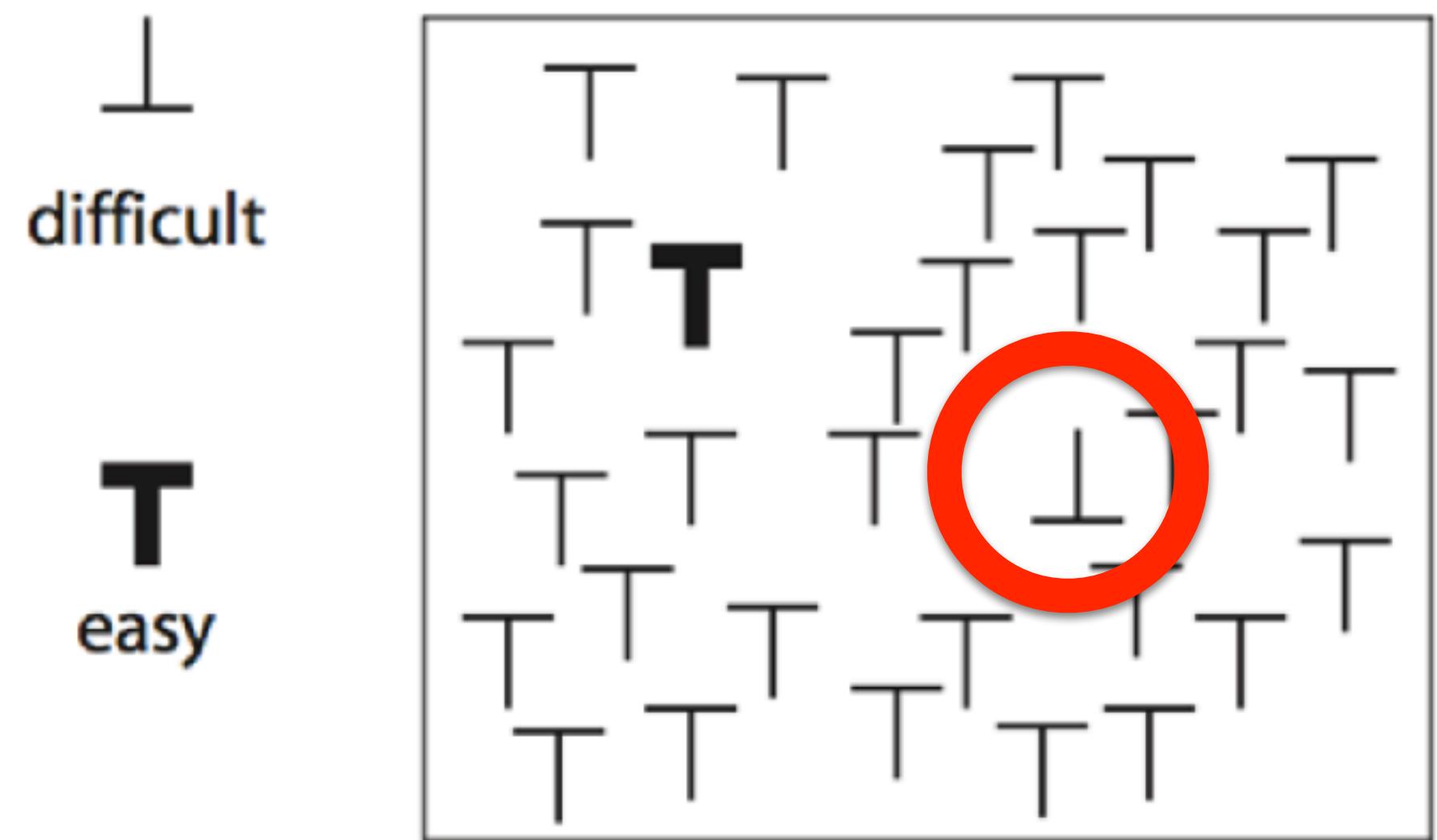
Shape Only, N=100

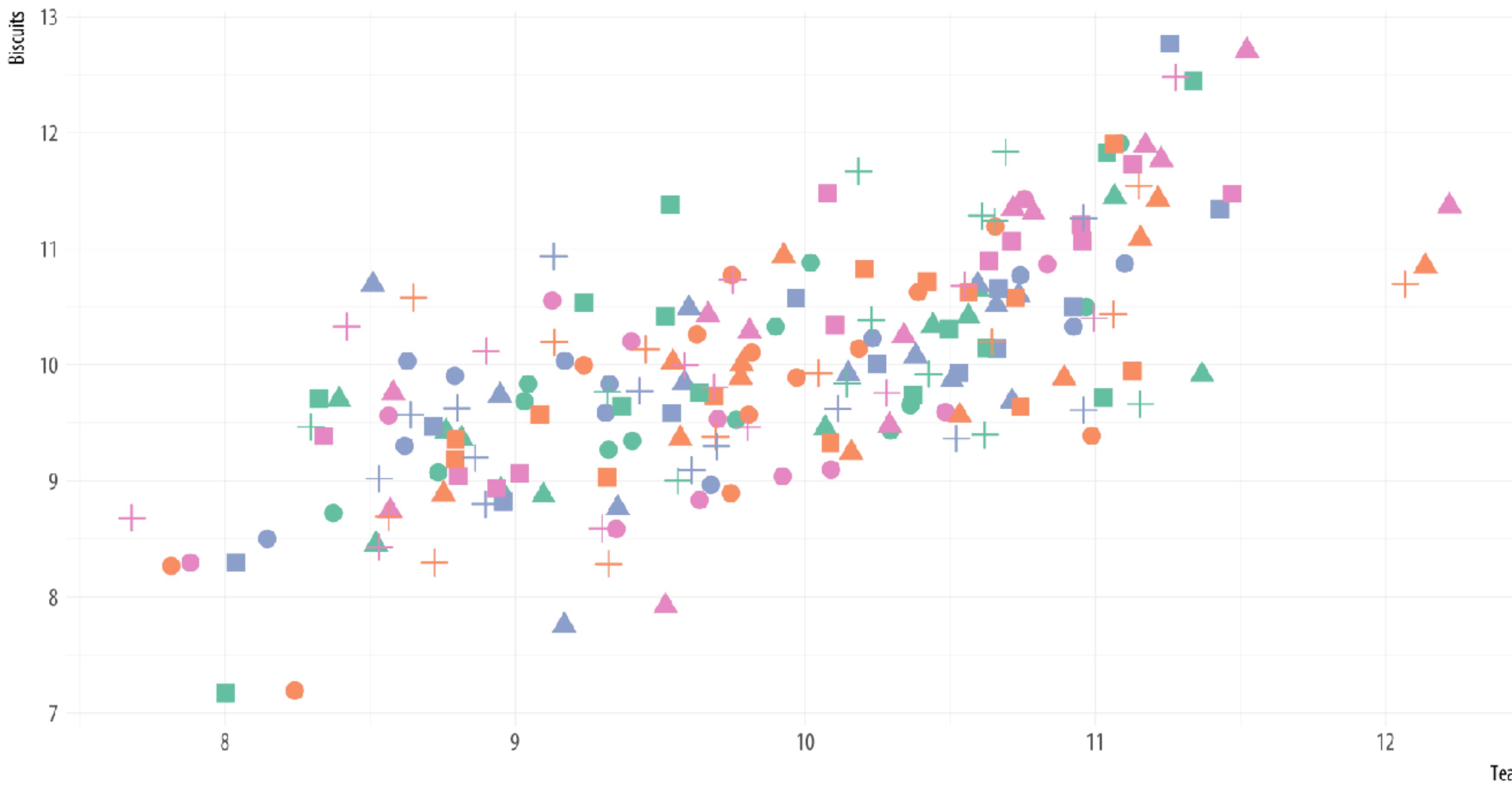


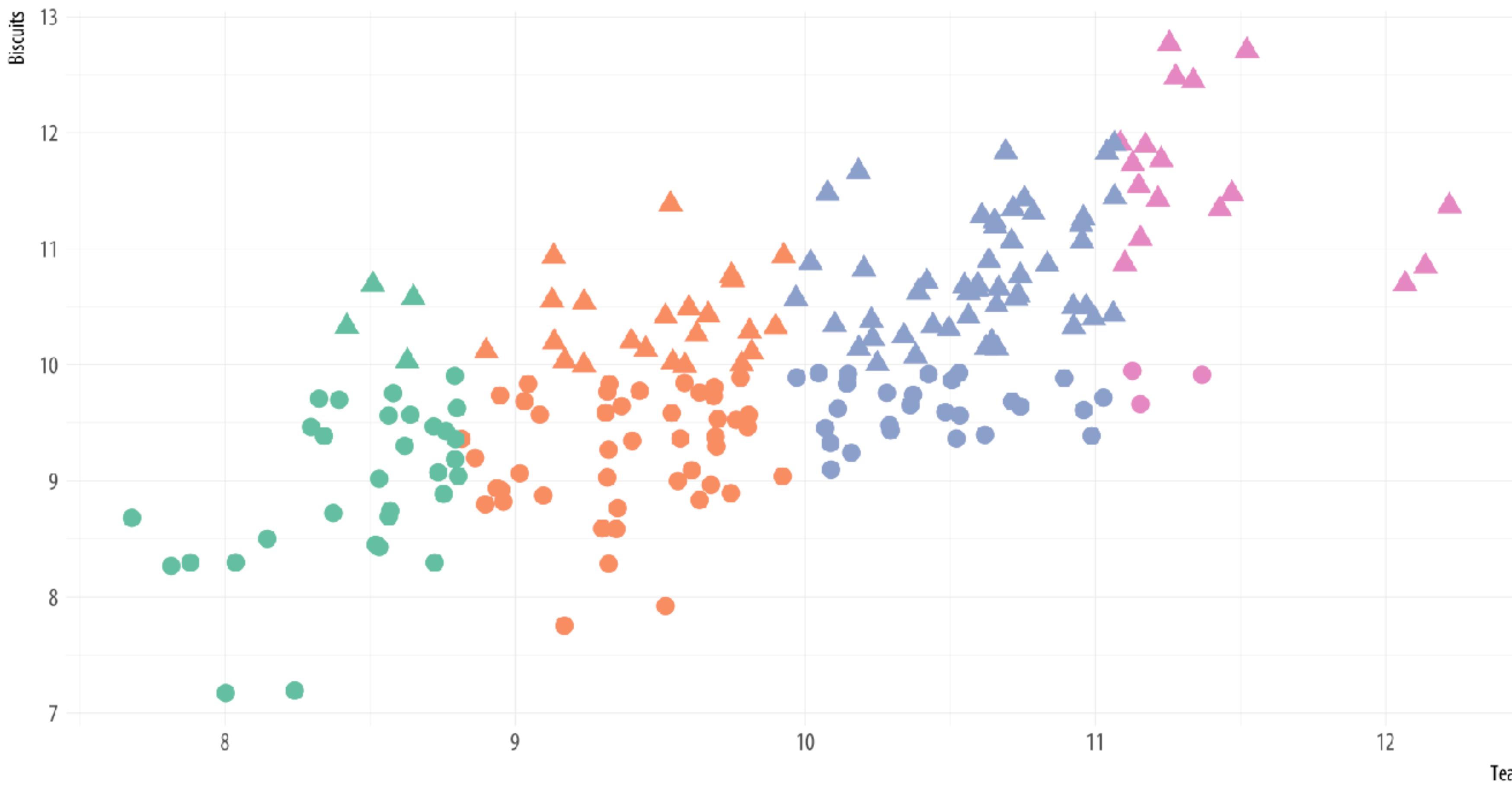
Color & Shape, N=100



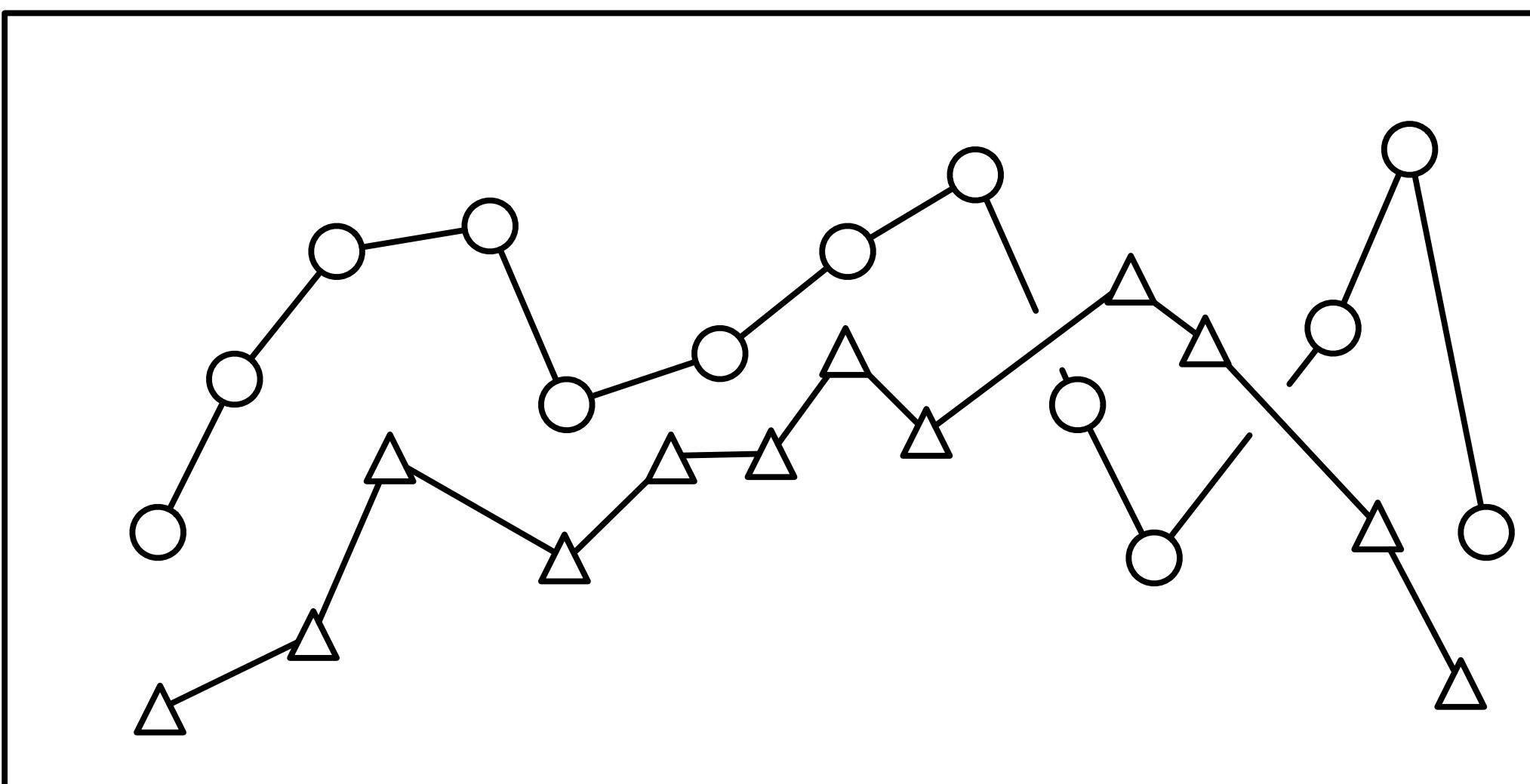
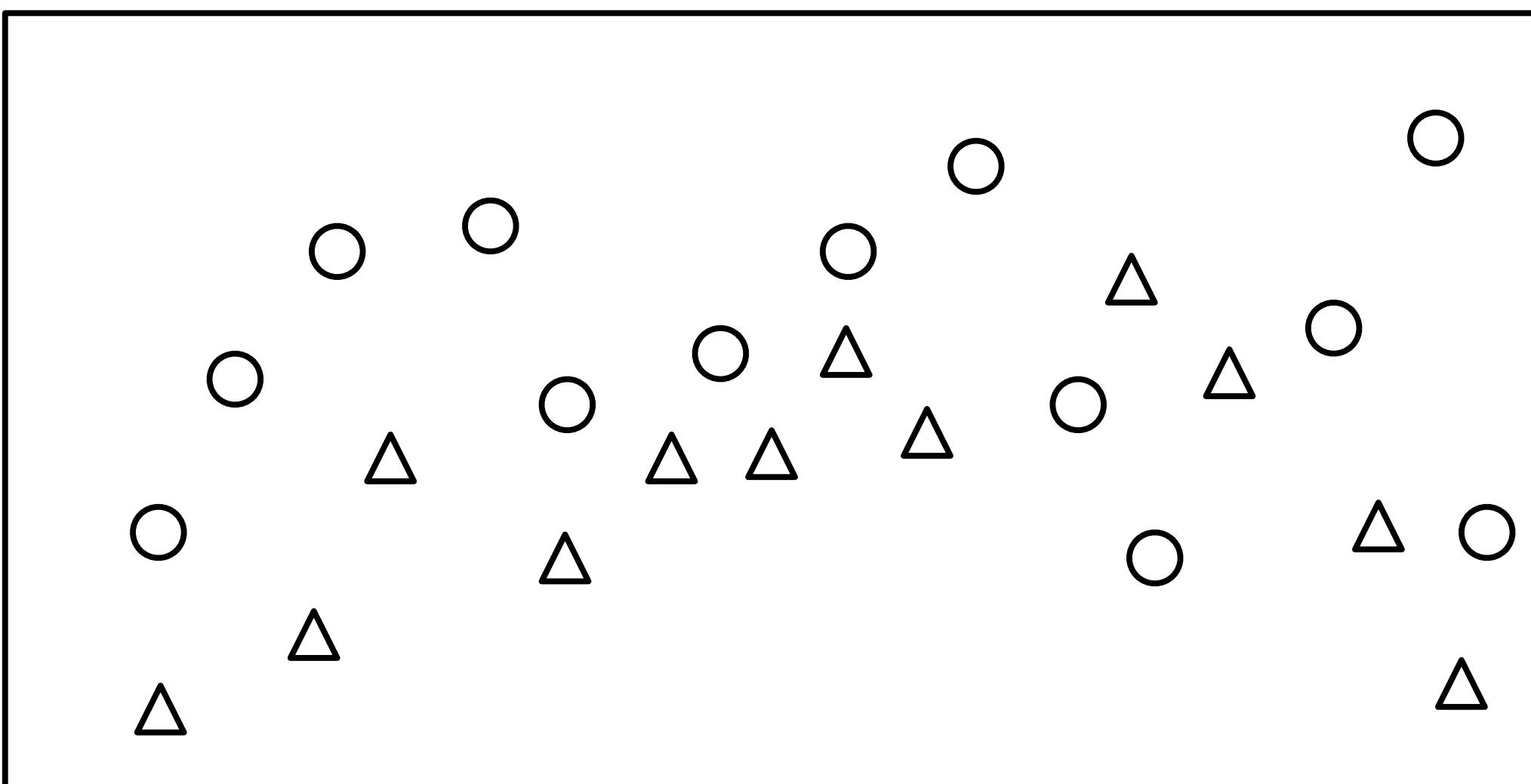
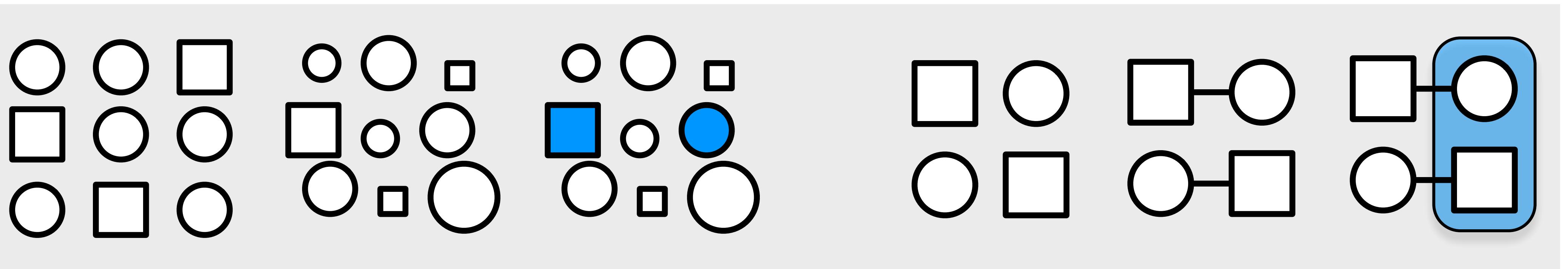
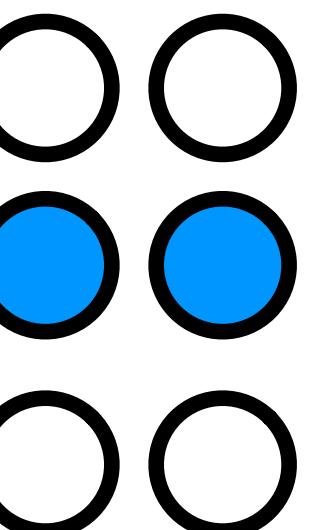
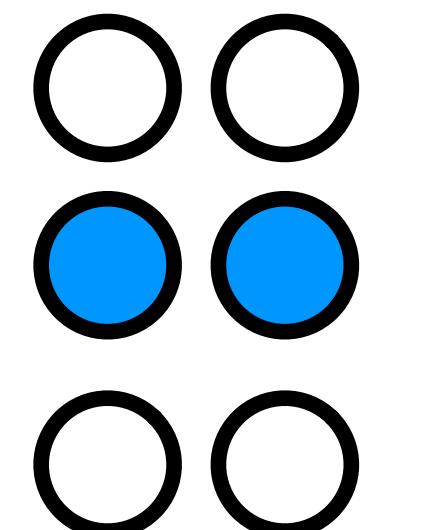
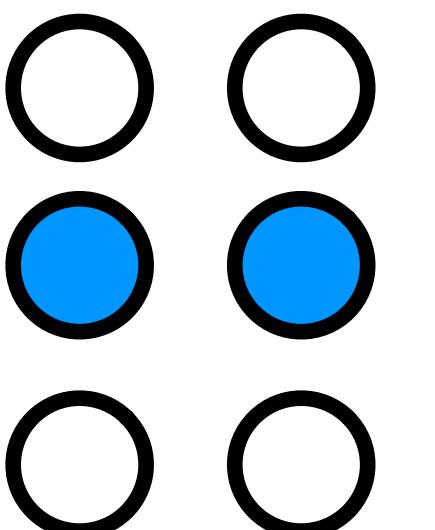
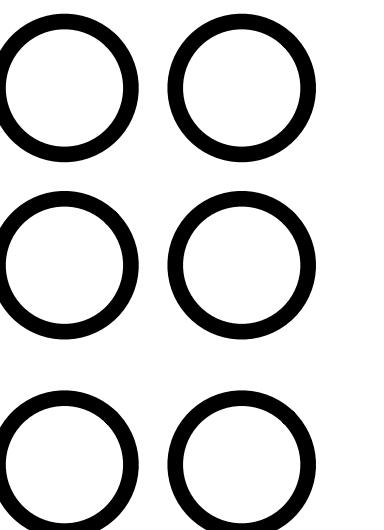
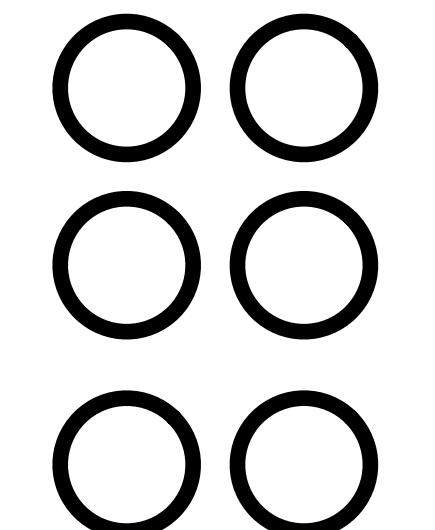
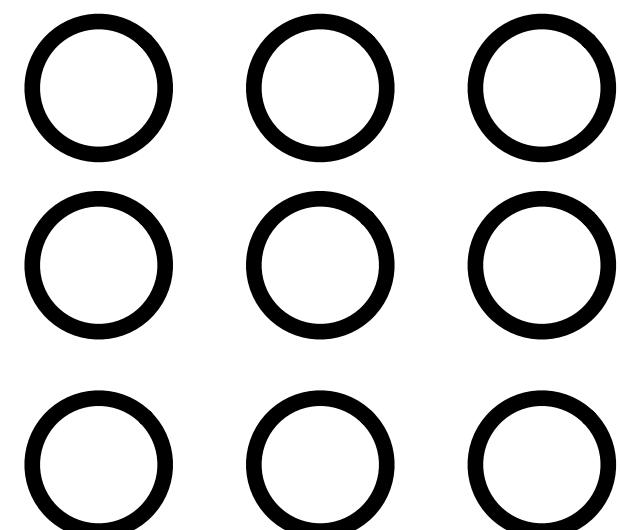




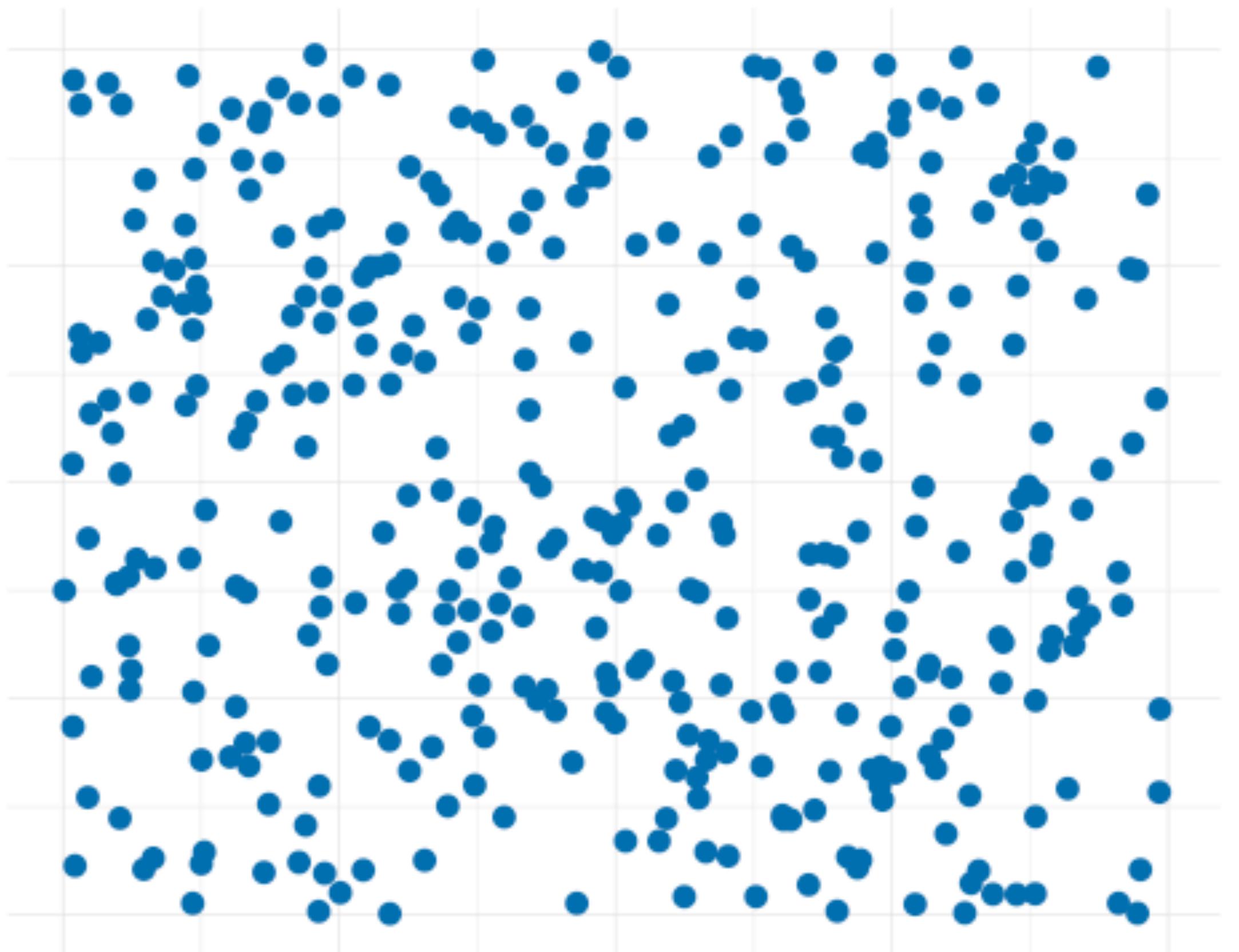




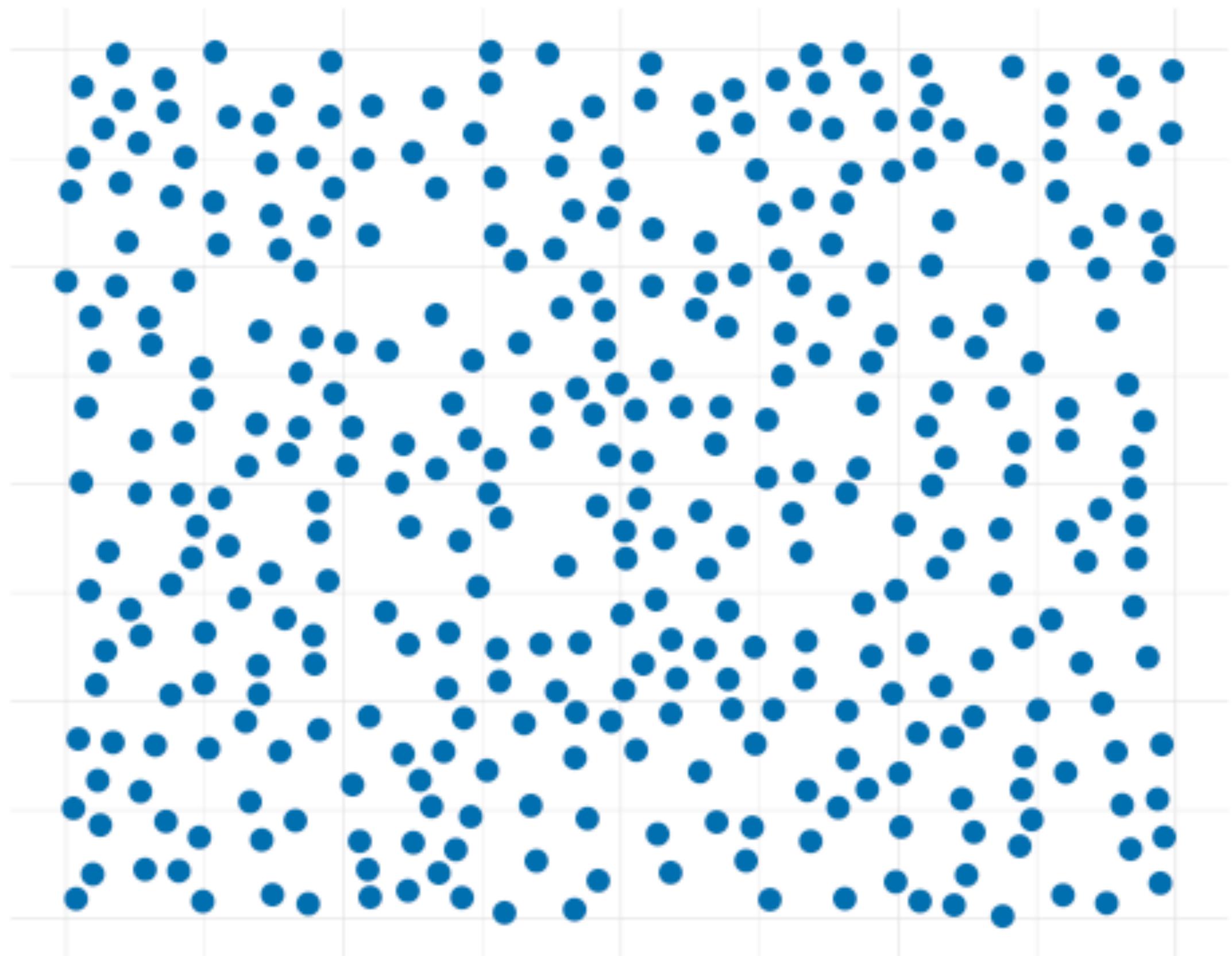
GESTALT INFERENCES



Poisson



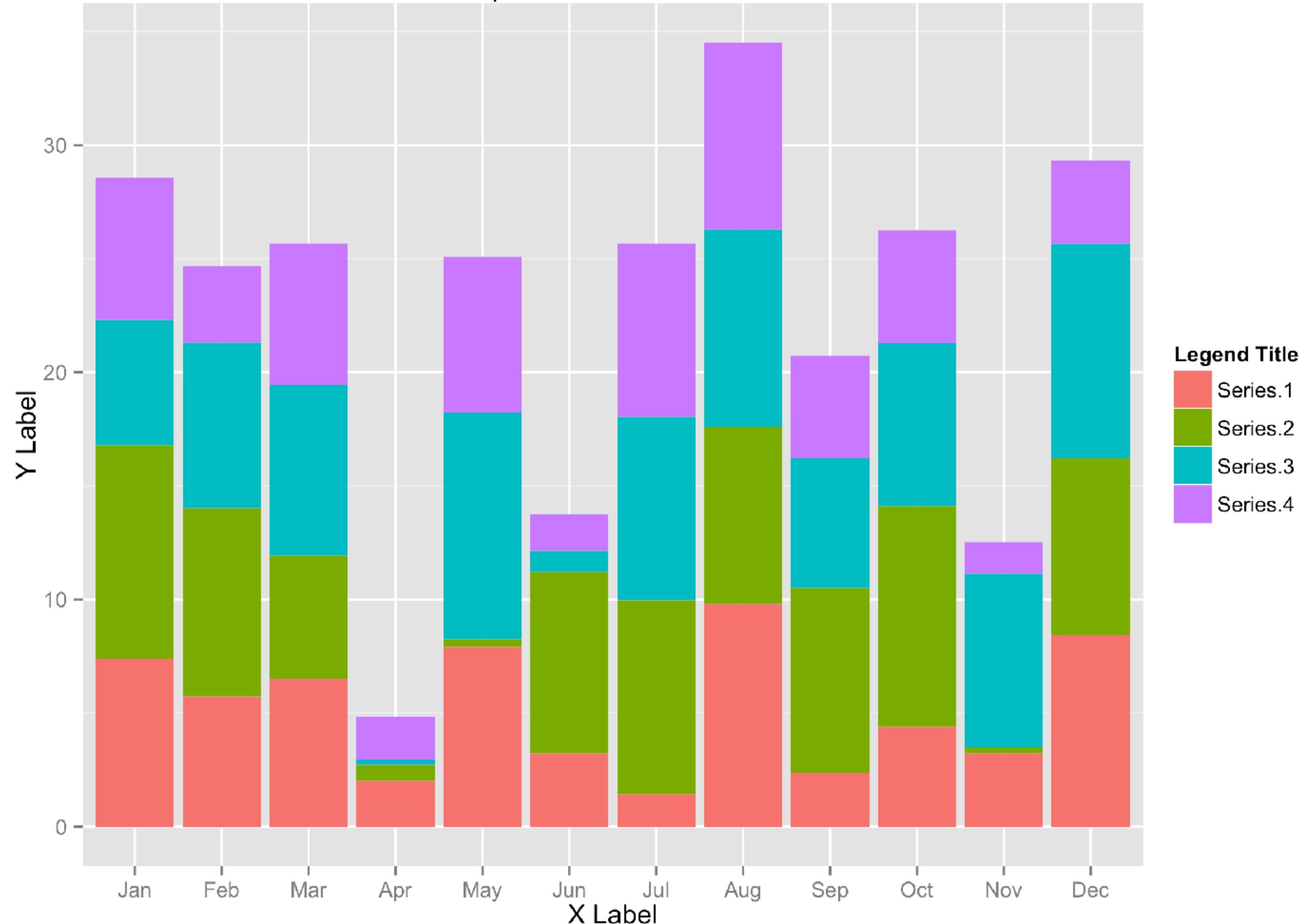
Matérn

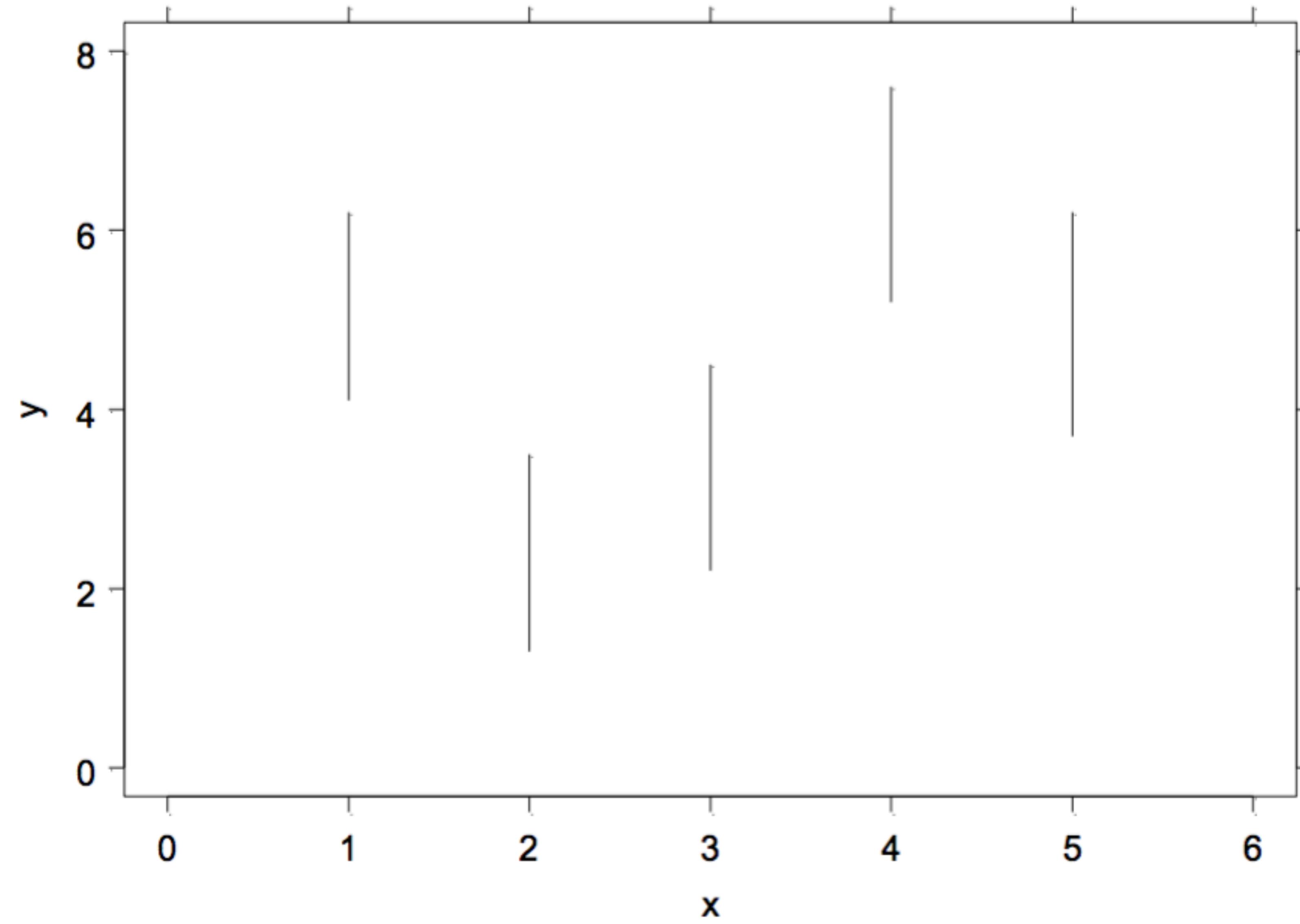


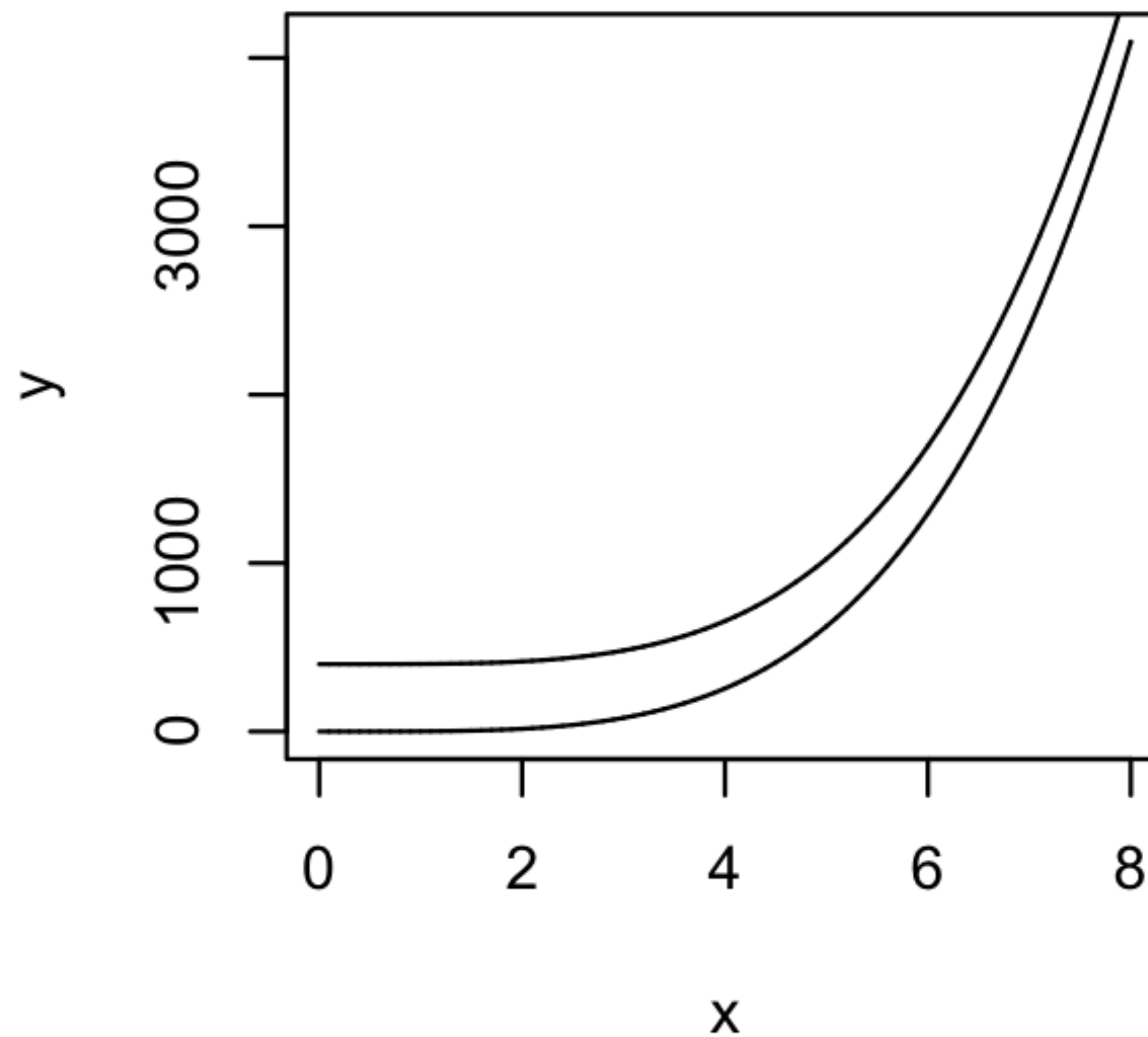
MAPPINGS FOR DATA

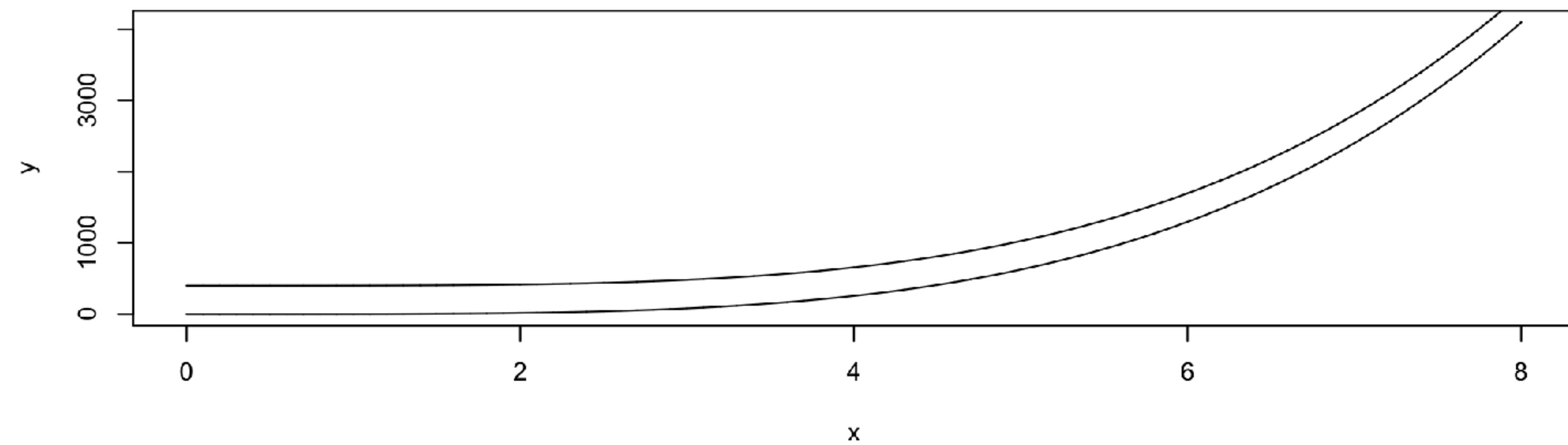
Visual Tasks in Decoding Graphs

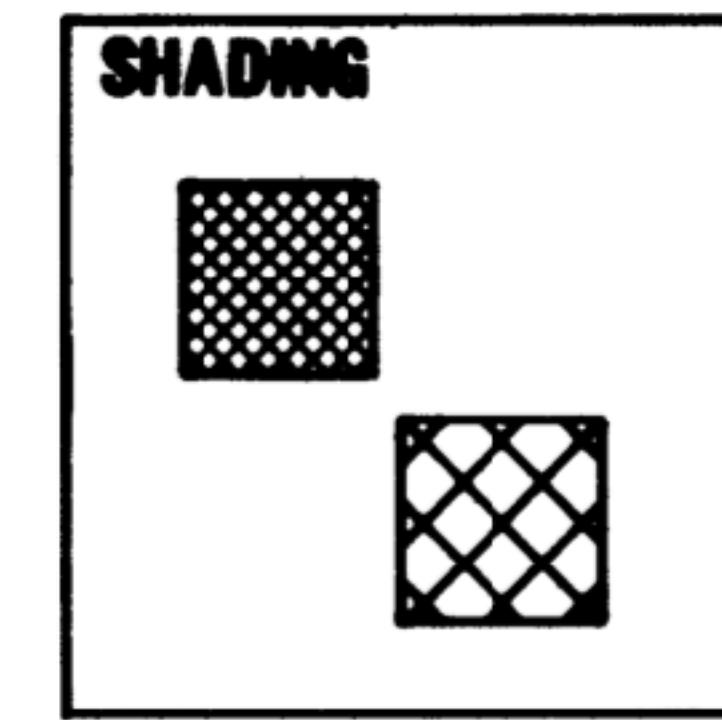
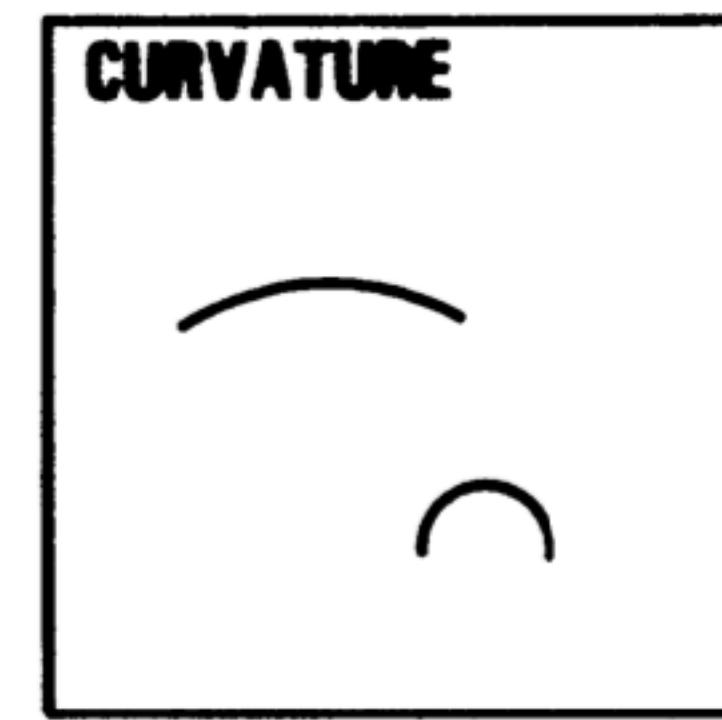
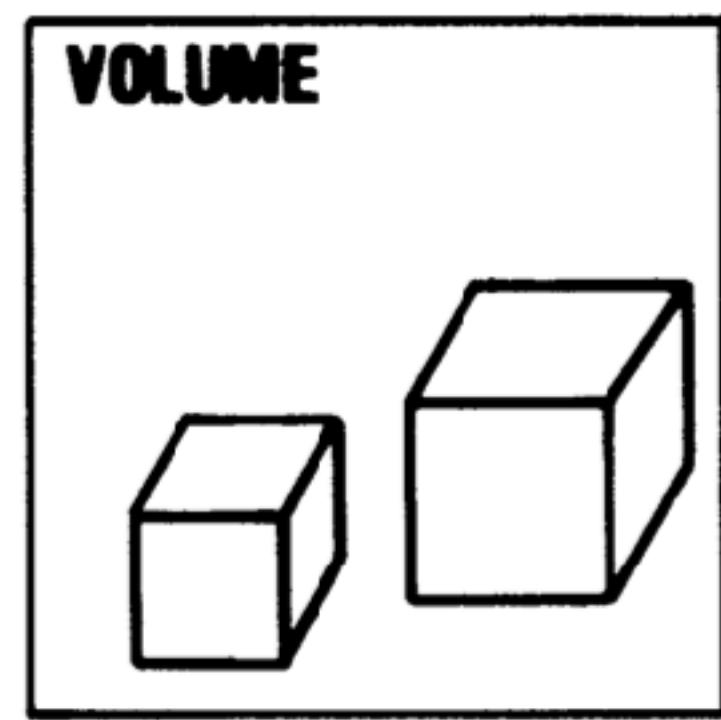
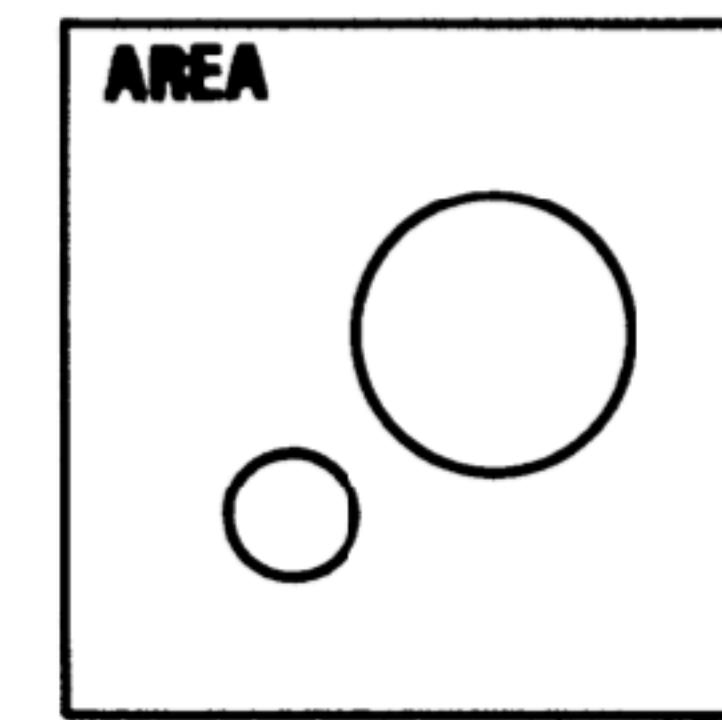
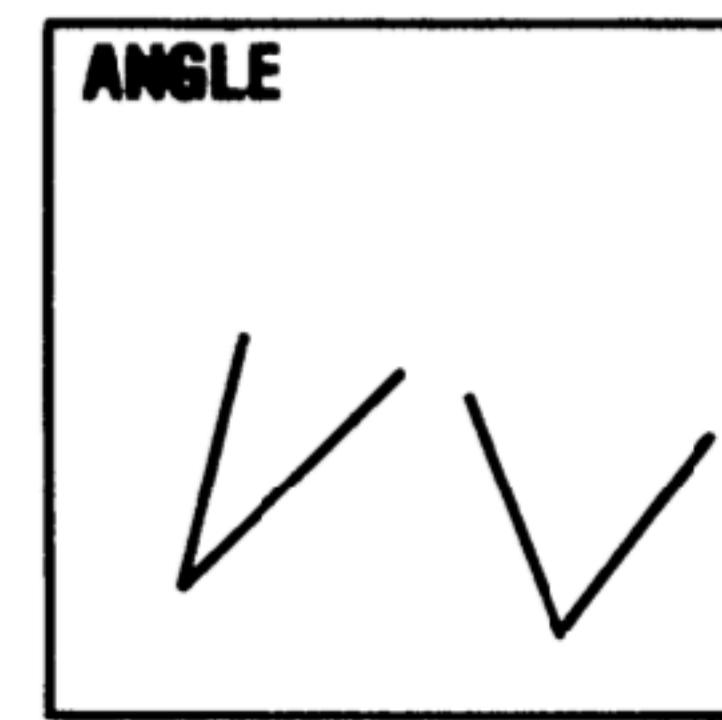
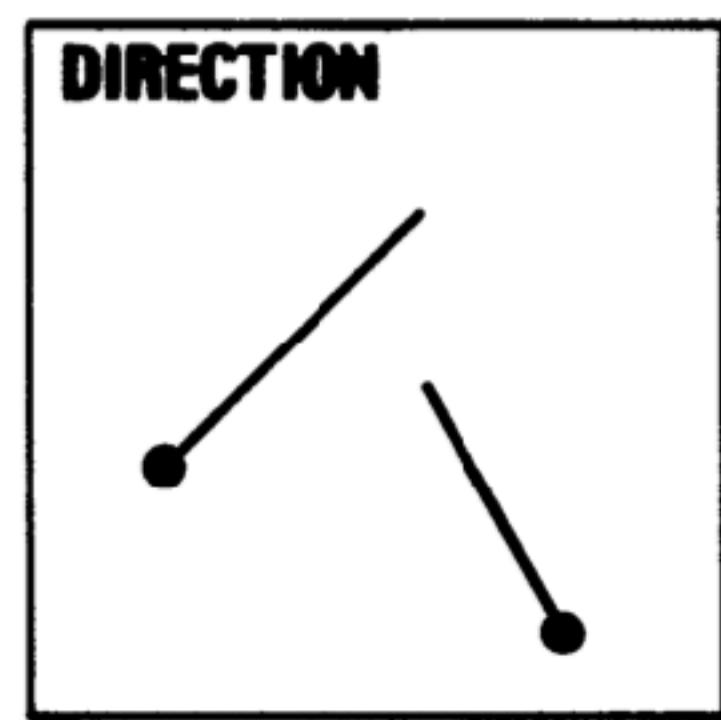
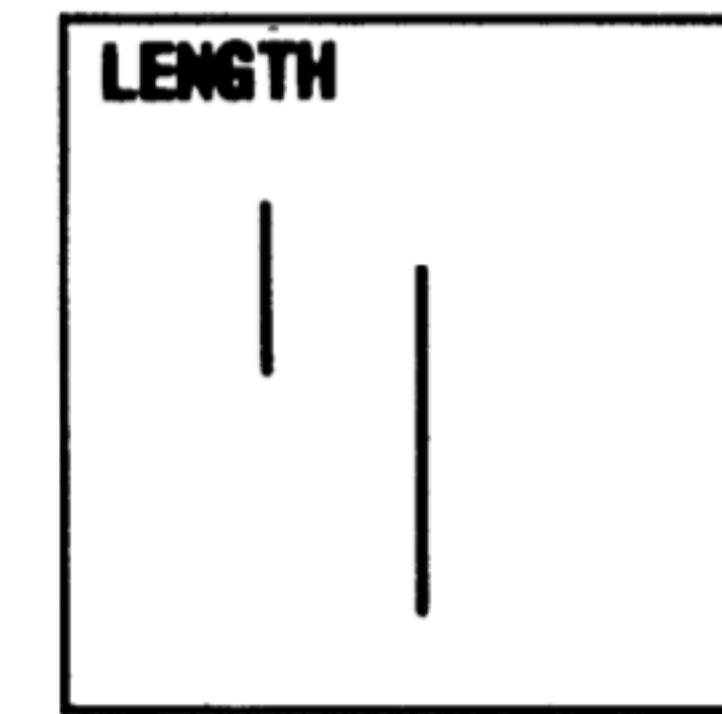
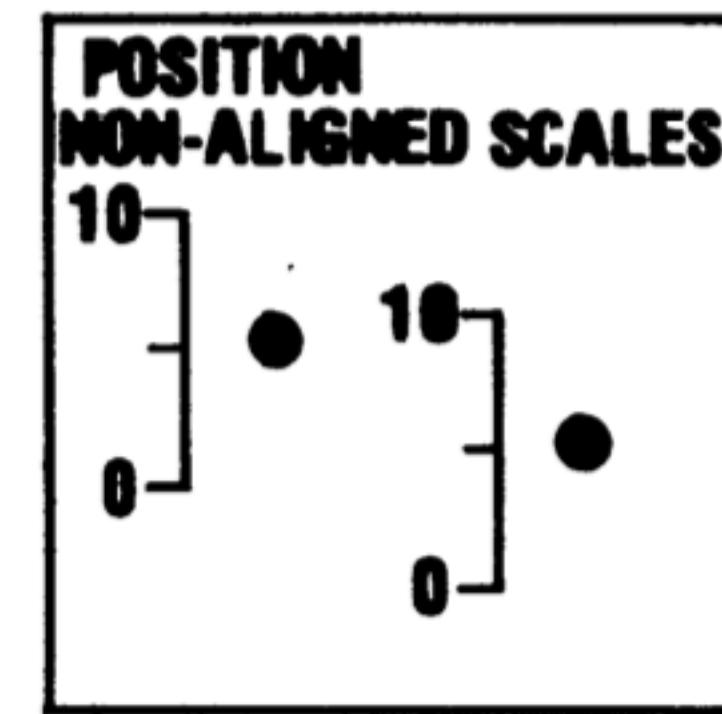
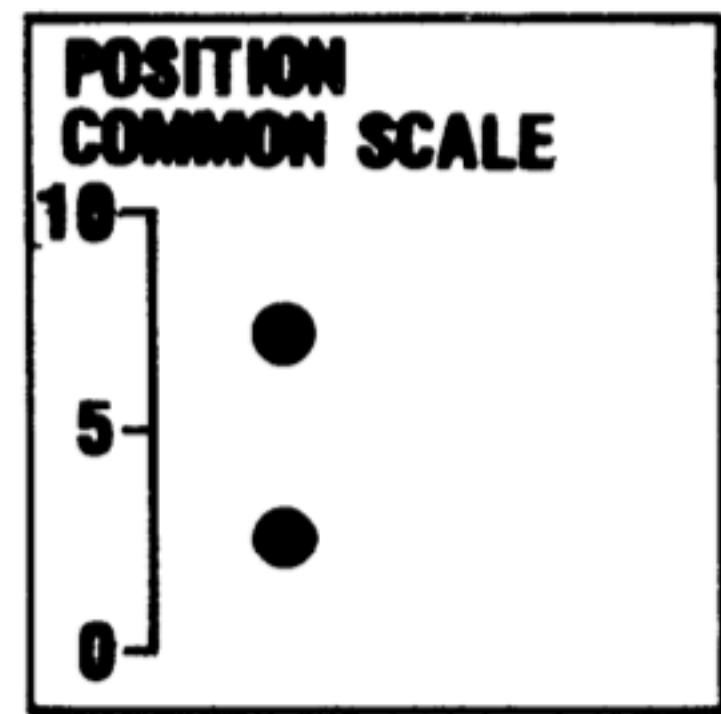
An Example Stacked Column Chart











COLOR SATURATION

Figure 1. Elementary perceptual tasks.

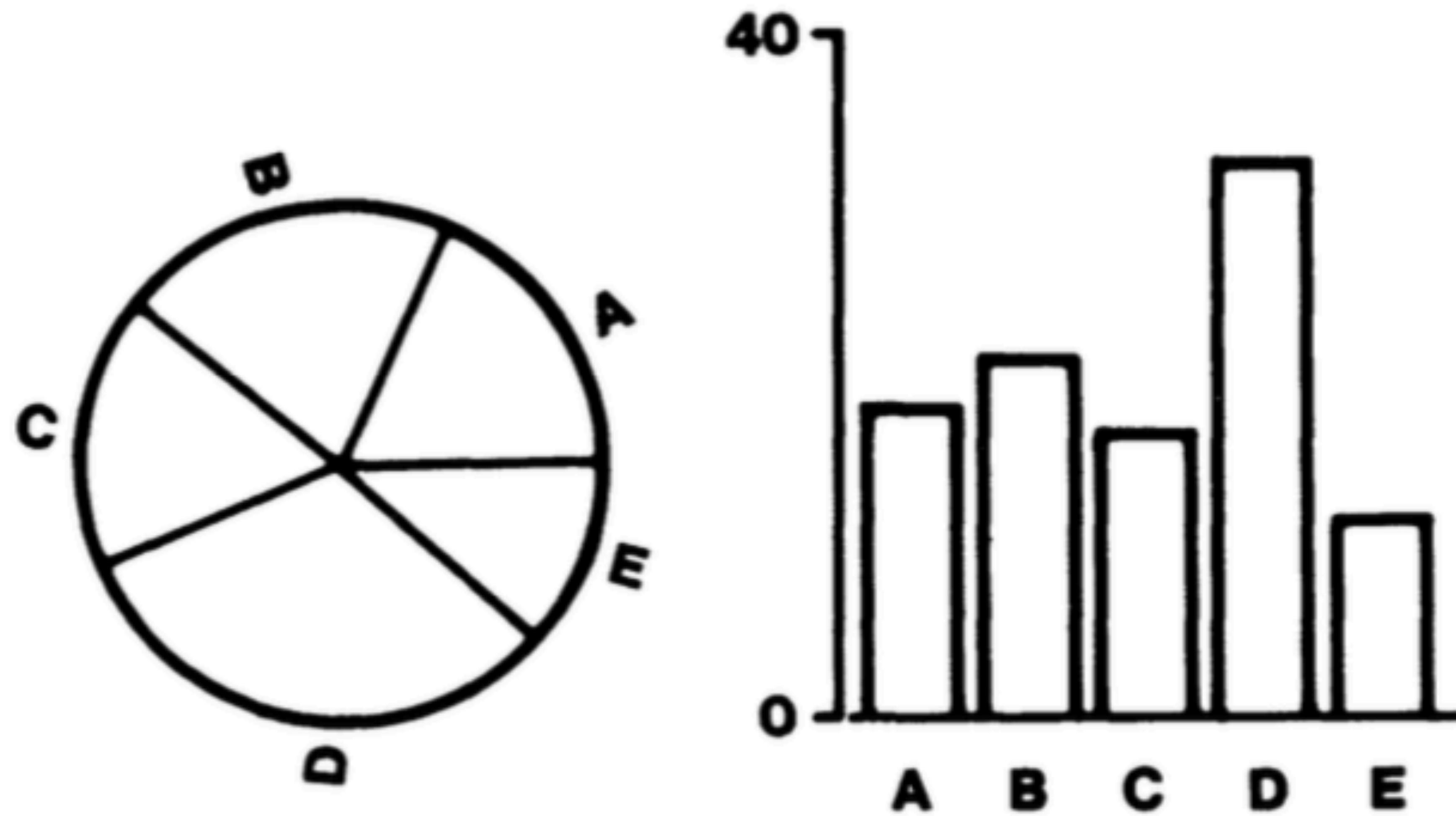


Figure 3. Graphs from position-angle experiment.

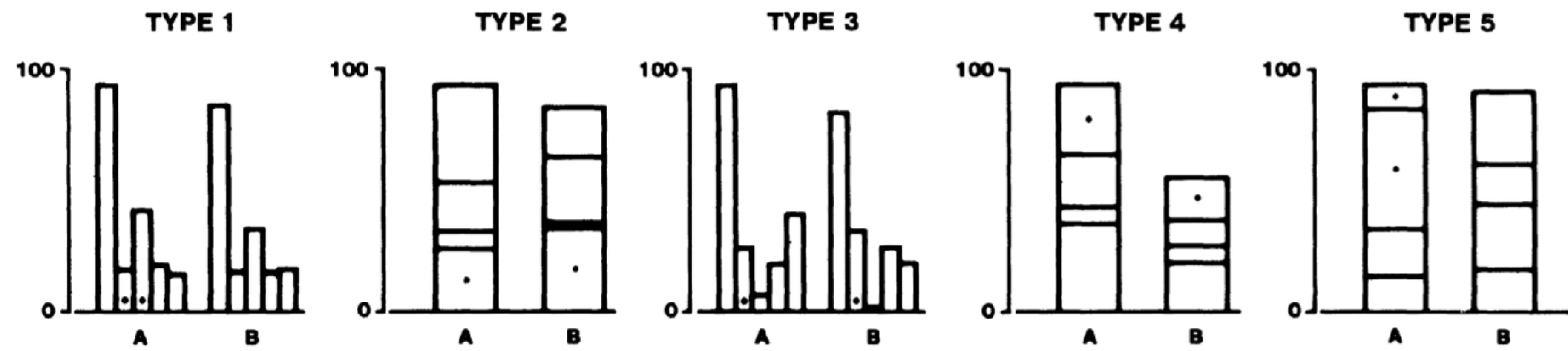
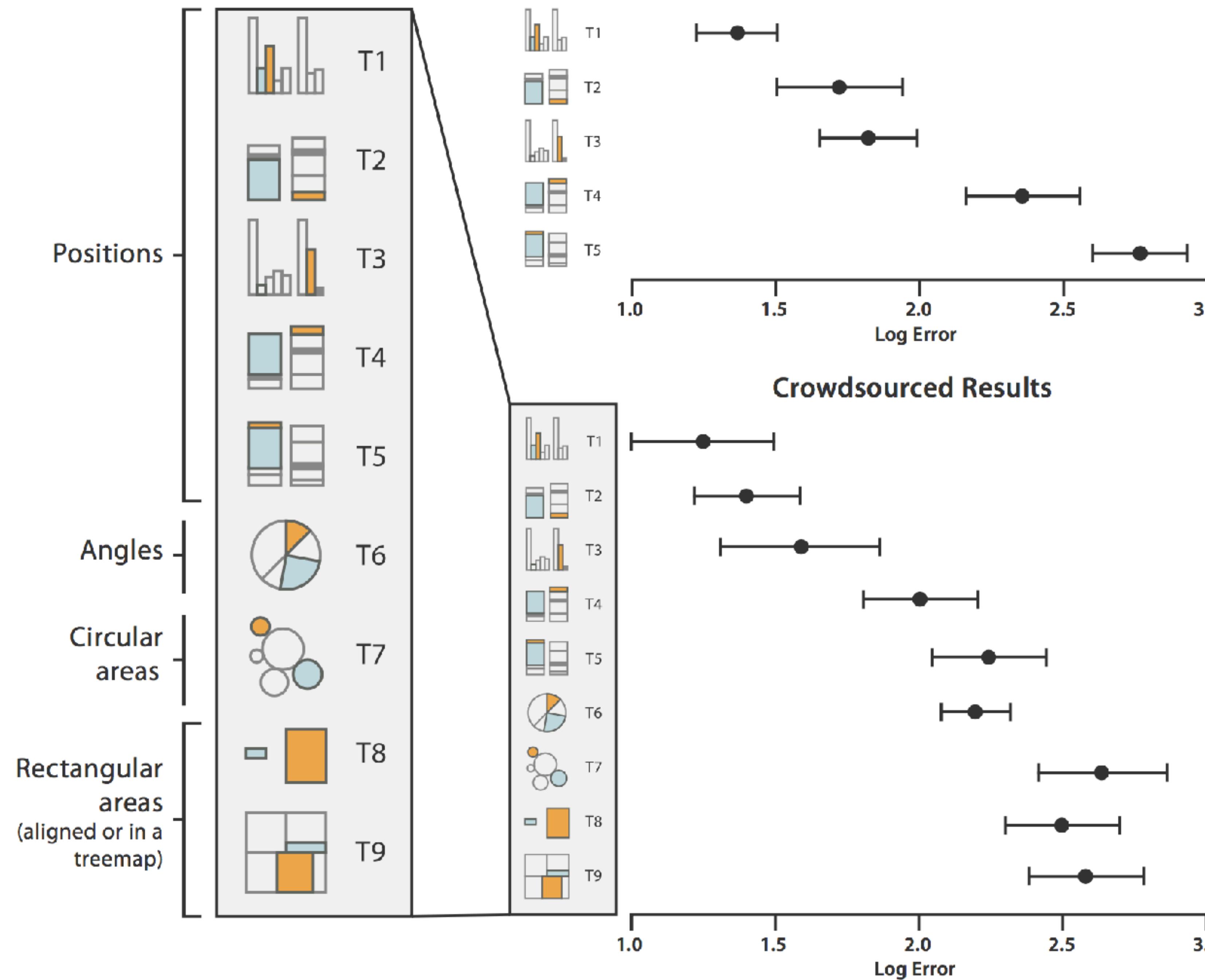


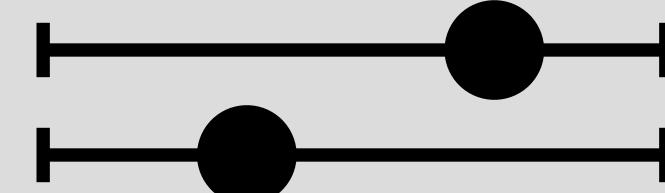
Figure 4. Graphs from position-length experiment.

Cleveland & McGill's Results



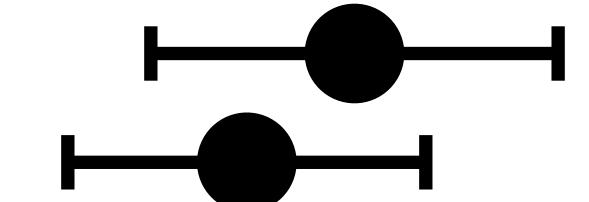
A Rough Hierarchy of Mappings

Position on a common scale



Better

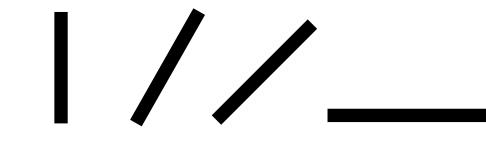
Position on unaligned scale



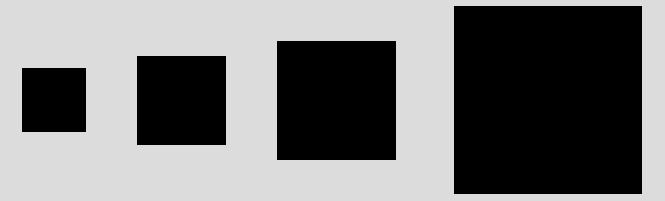
Length (1D as size)



Tilt or Angle



Area (2D as size)



Effectiveness

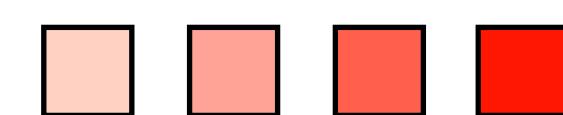
Depth (3D as Position)



Color luminance [brightness]

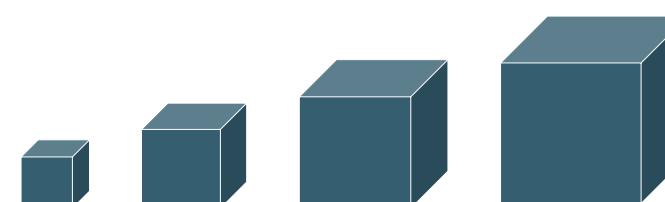


Color saturation [intensity]

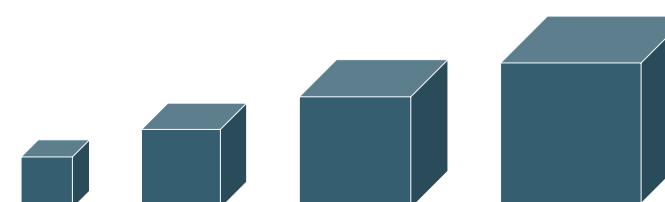


Worse

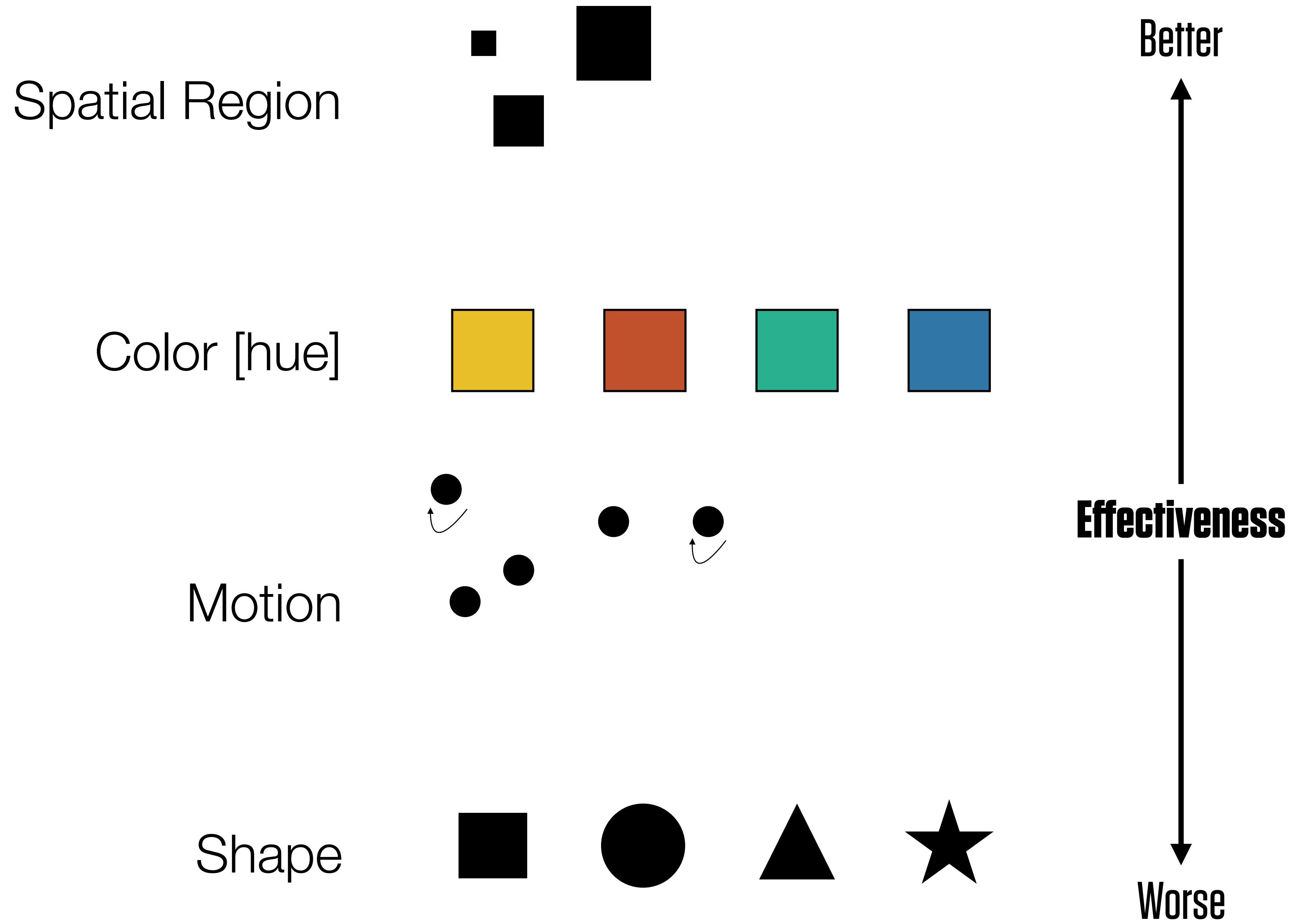
Curvature



Volume (3D as size)



See e.g. Munzer (2014)

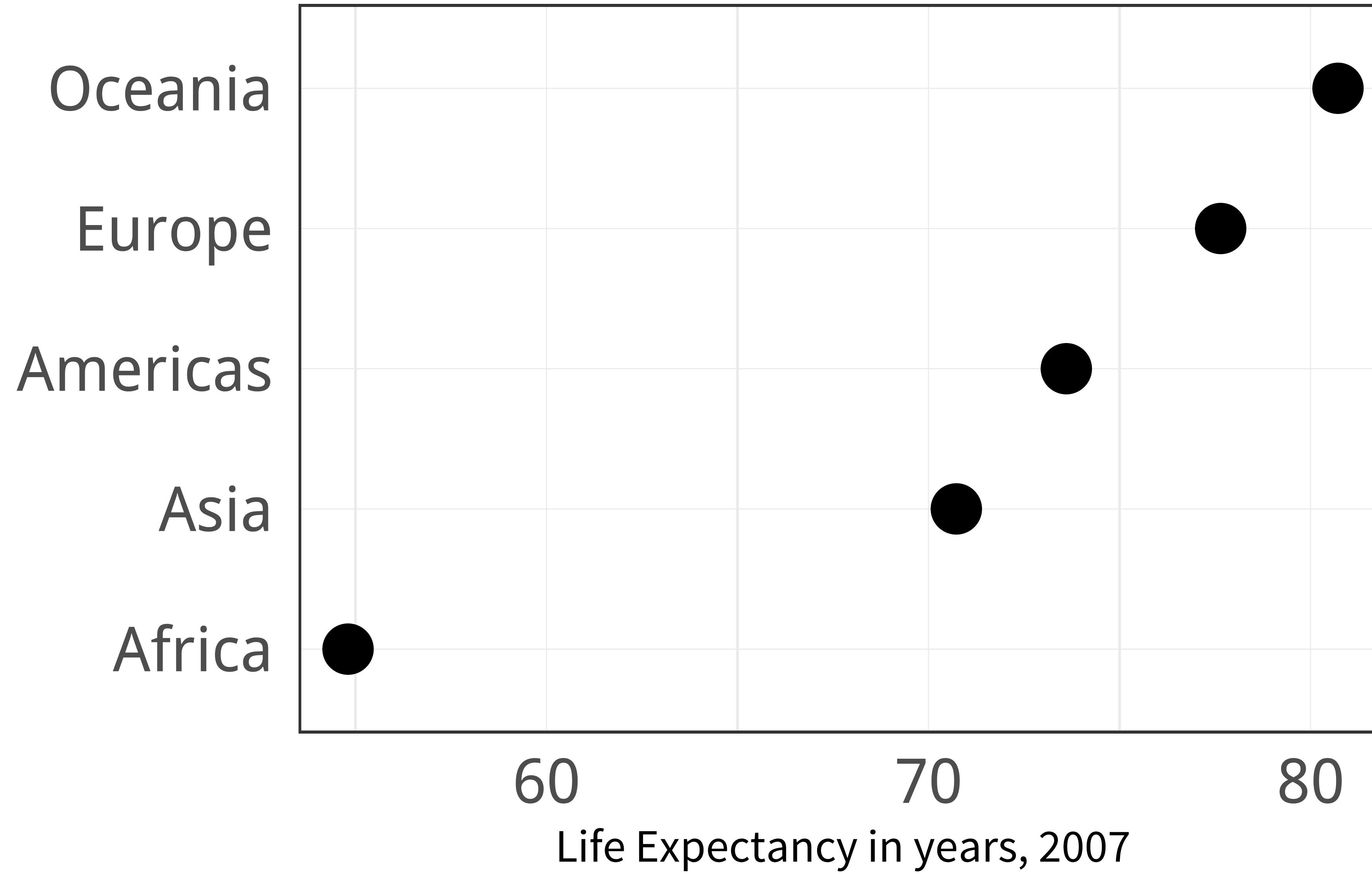


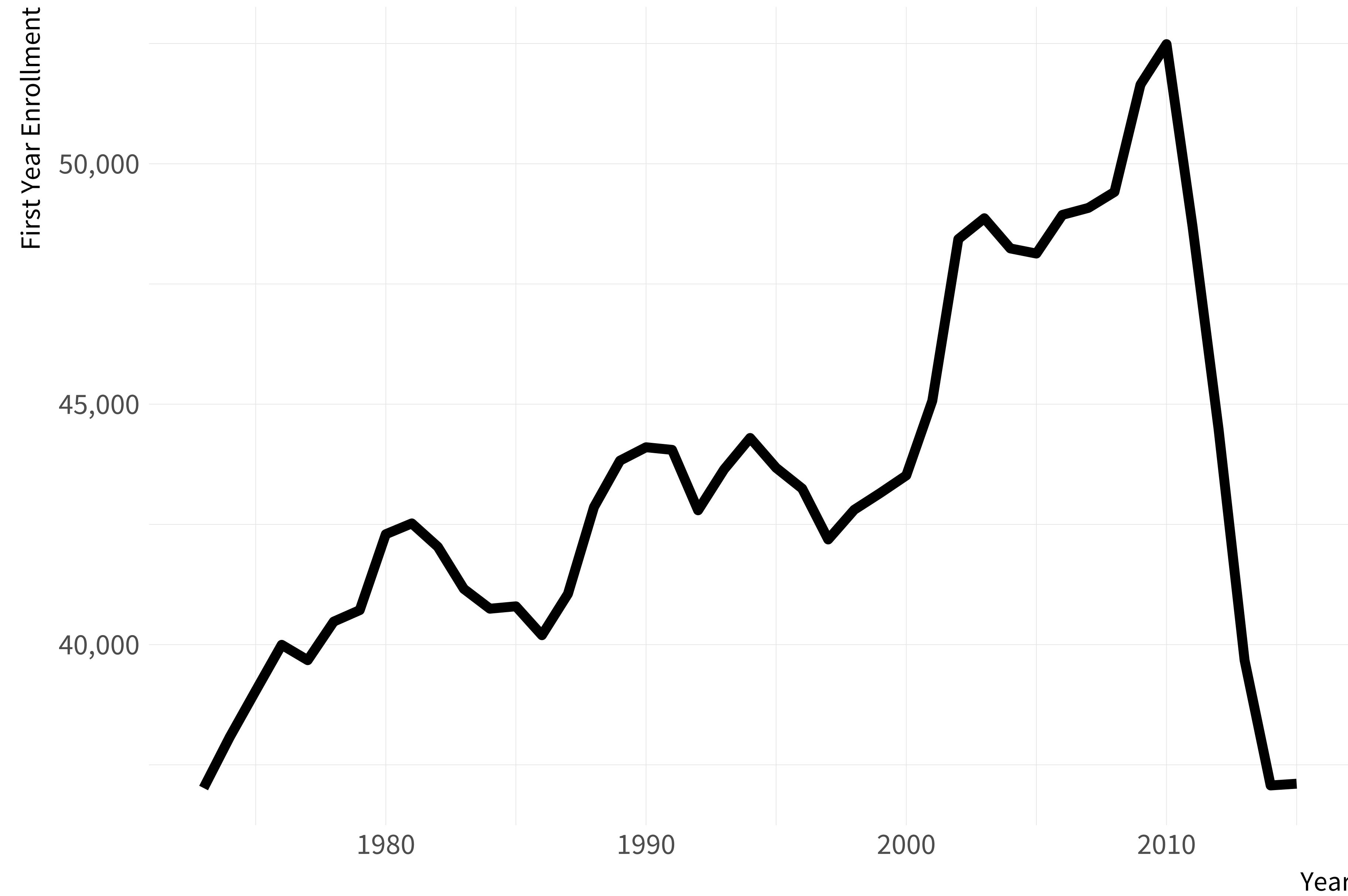
See e.g. Munzer (2014)

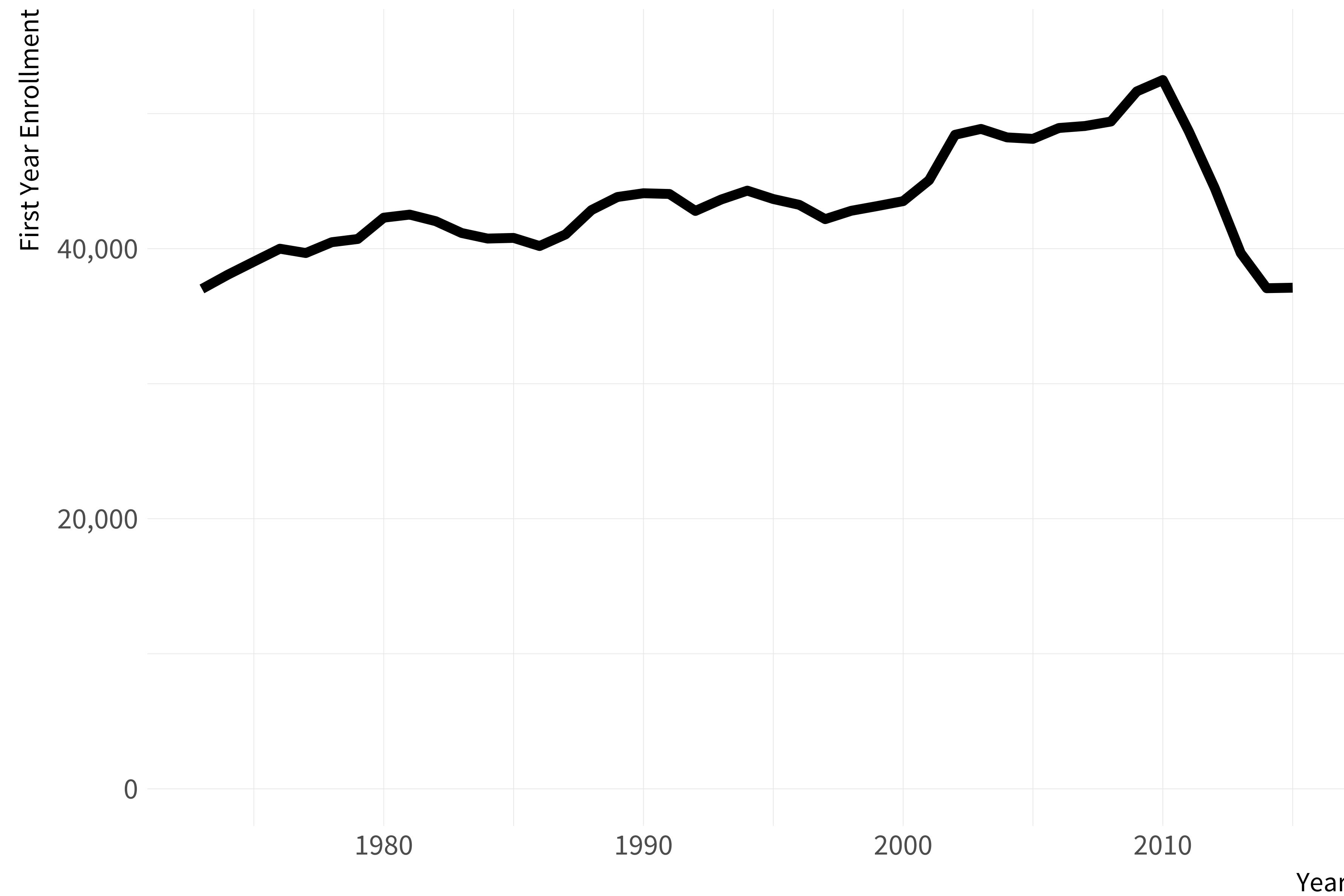
**HONESTY AND
GOOD JUDGMENT**

**Example:
Baselines**



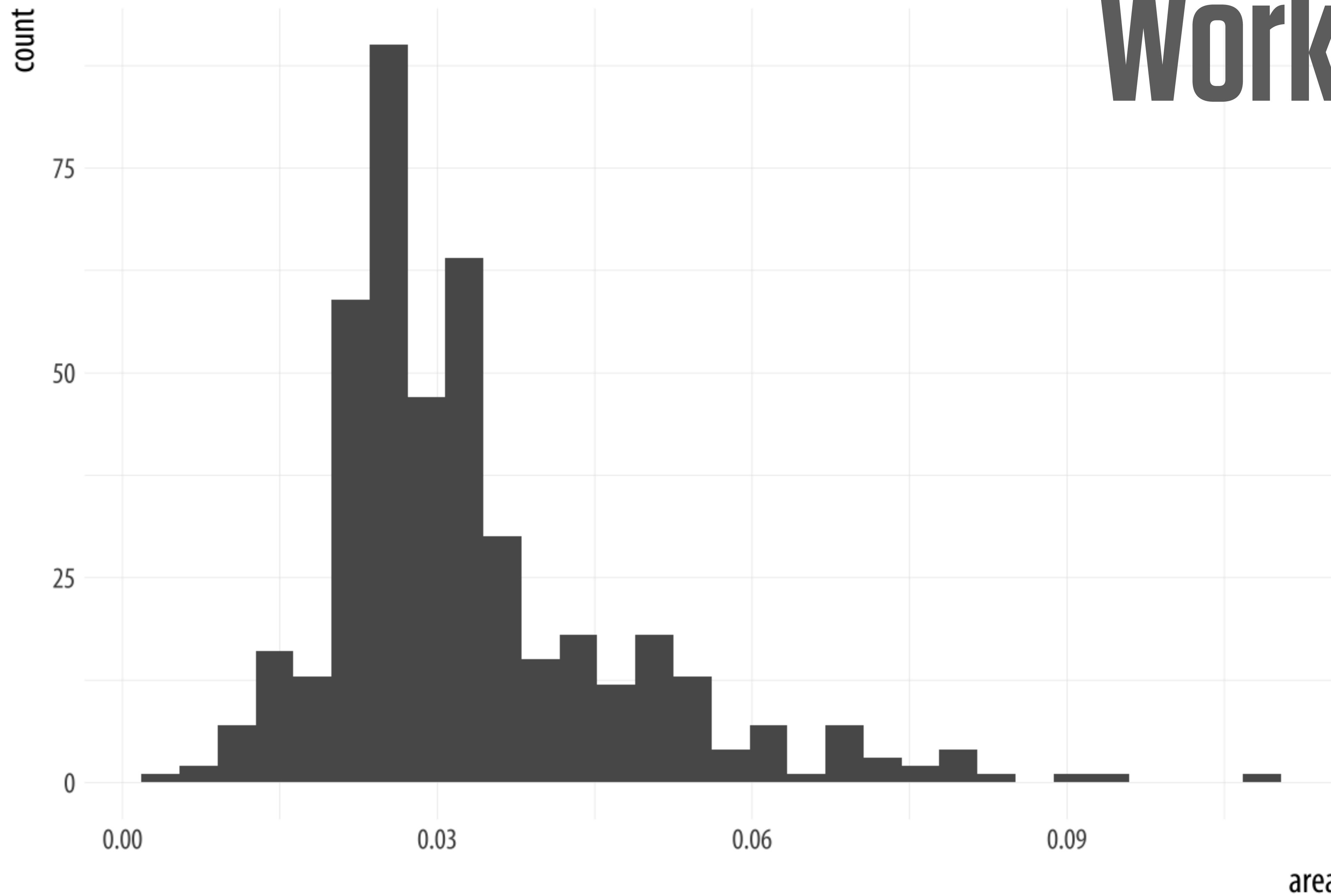






In Practice

Workhorses



lifeExp

80

60

40

30000

60000

90000

gdpPerCap

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

60

40

0

80

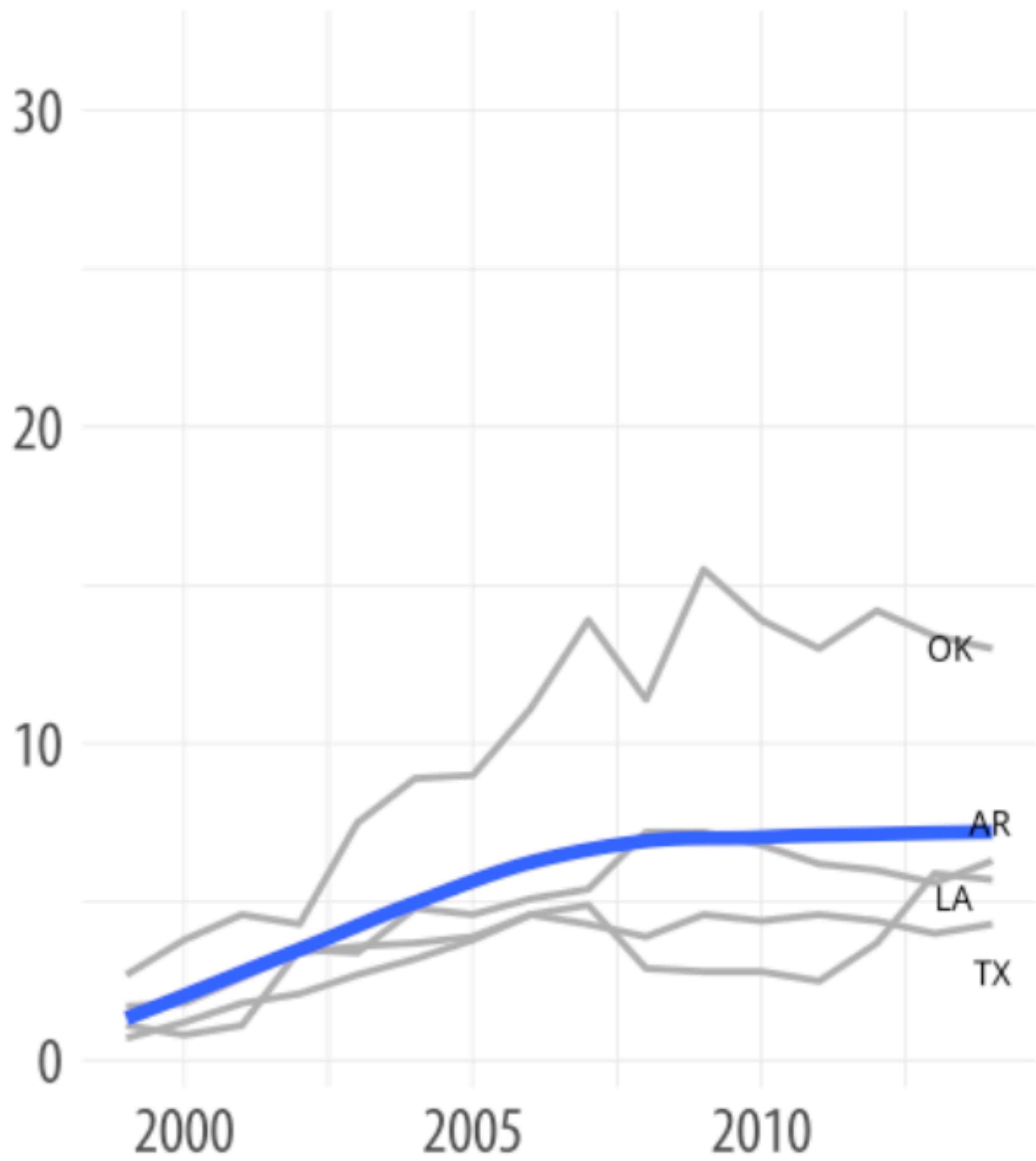
60

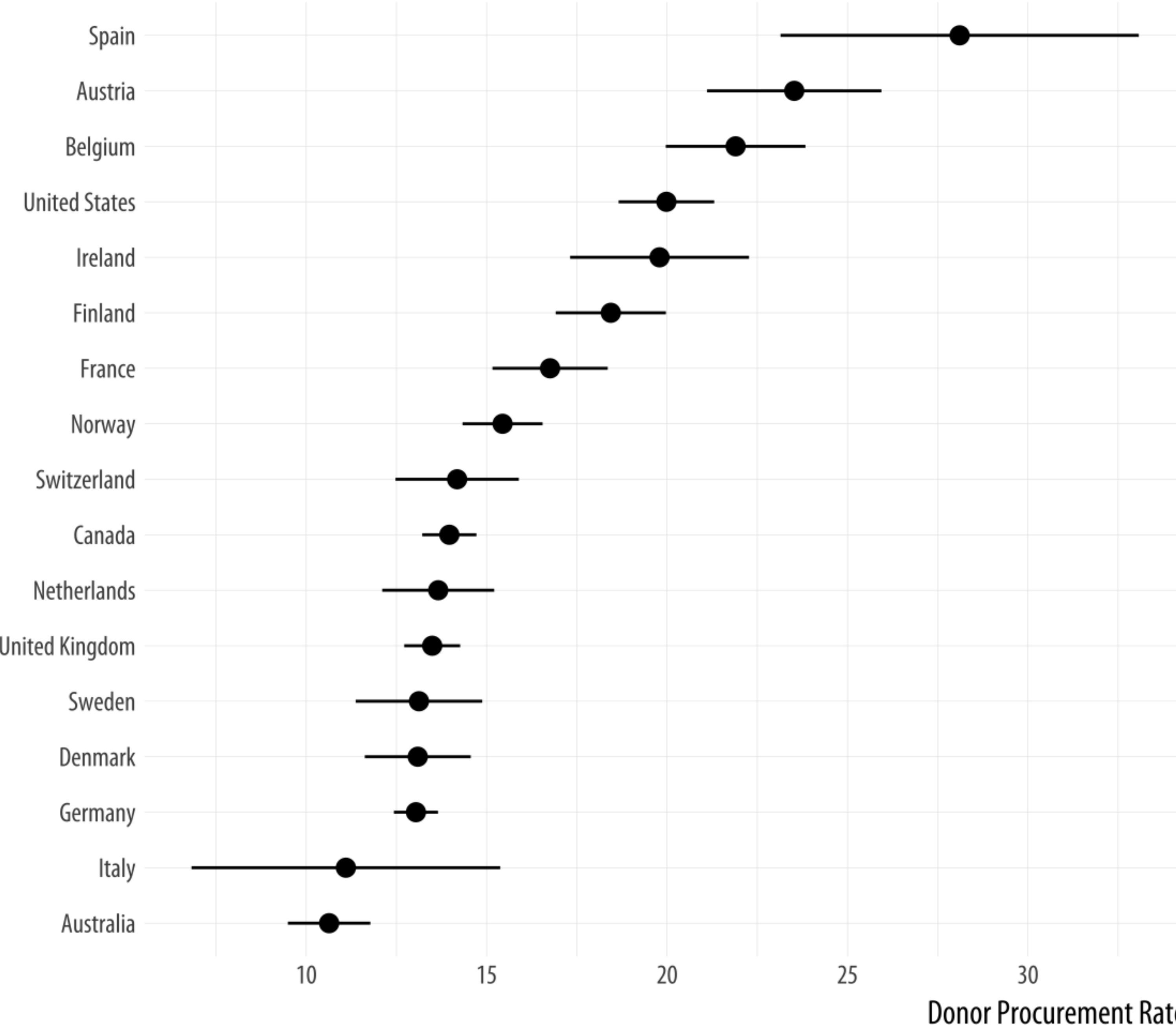
40

0

80

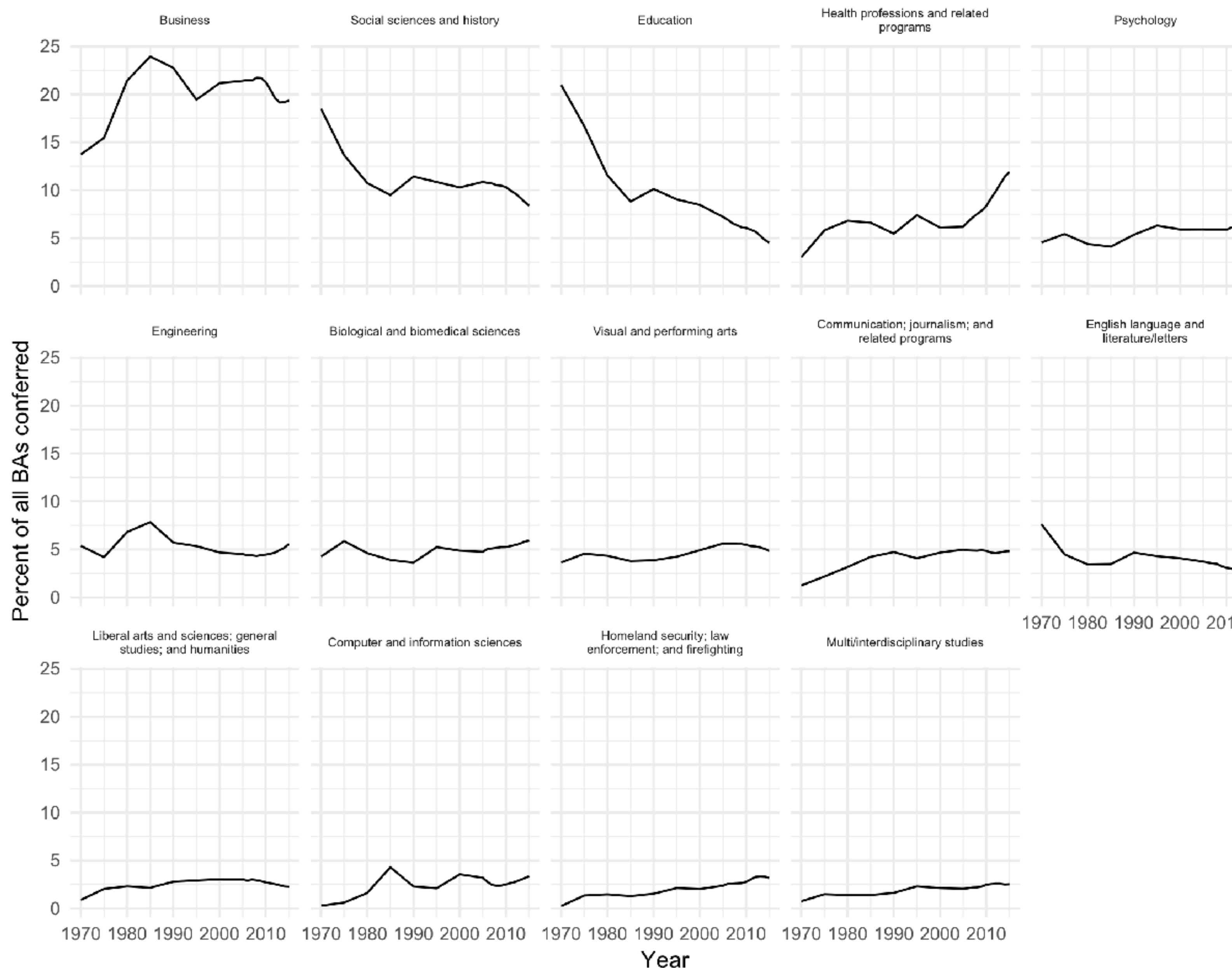
60





US Trends in Bachelor's Degrees Conferred, 1970-2015, for Areas averaging more than 2% of all degrees

Observations are every 5 years from 1970-1995, and annually thereafter



Data from NCES Digest 2017, Table 322.10.

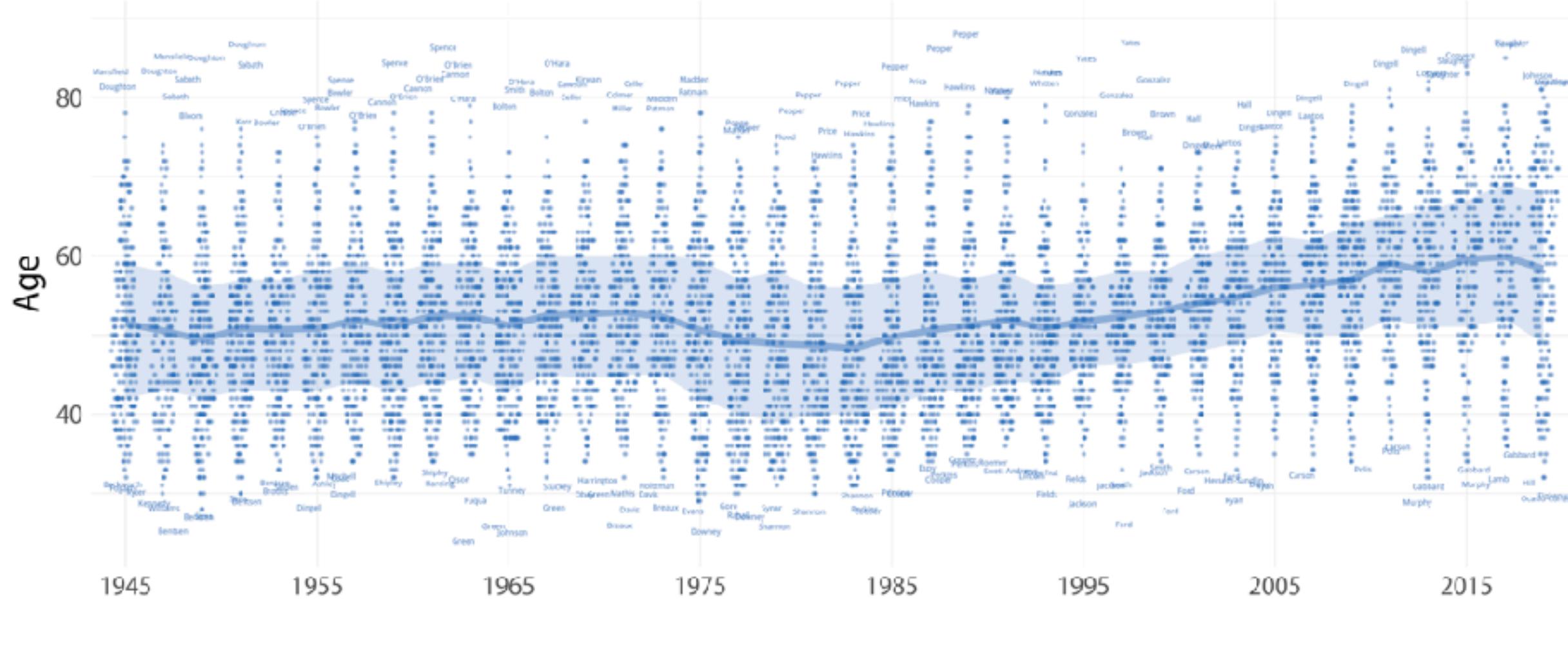
Show Ponies

Age Distribution of Congressional Representatives, 1945-2019

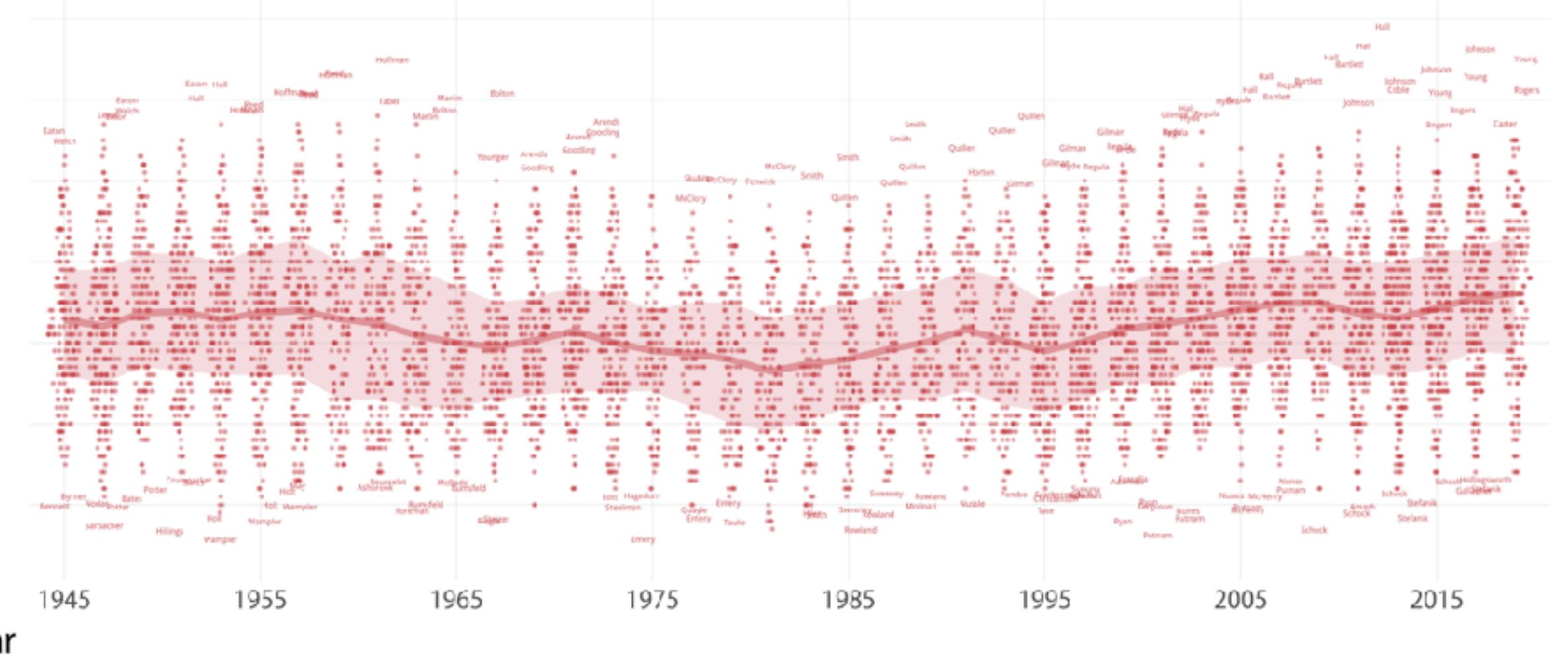
Trend line is mean age; bands are 25th and 75th percentiles of the range.

Youngest and oldest percentiles are named instead of being shown by points.

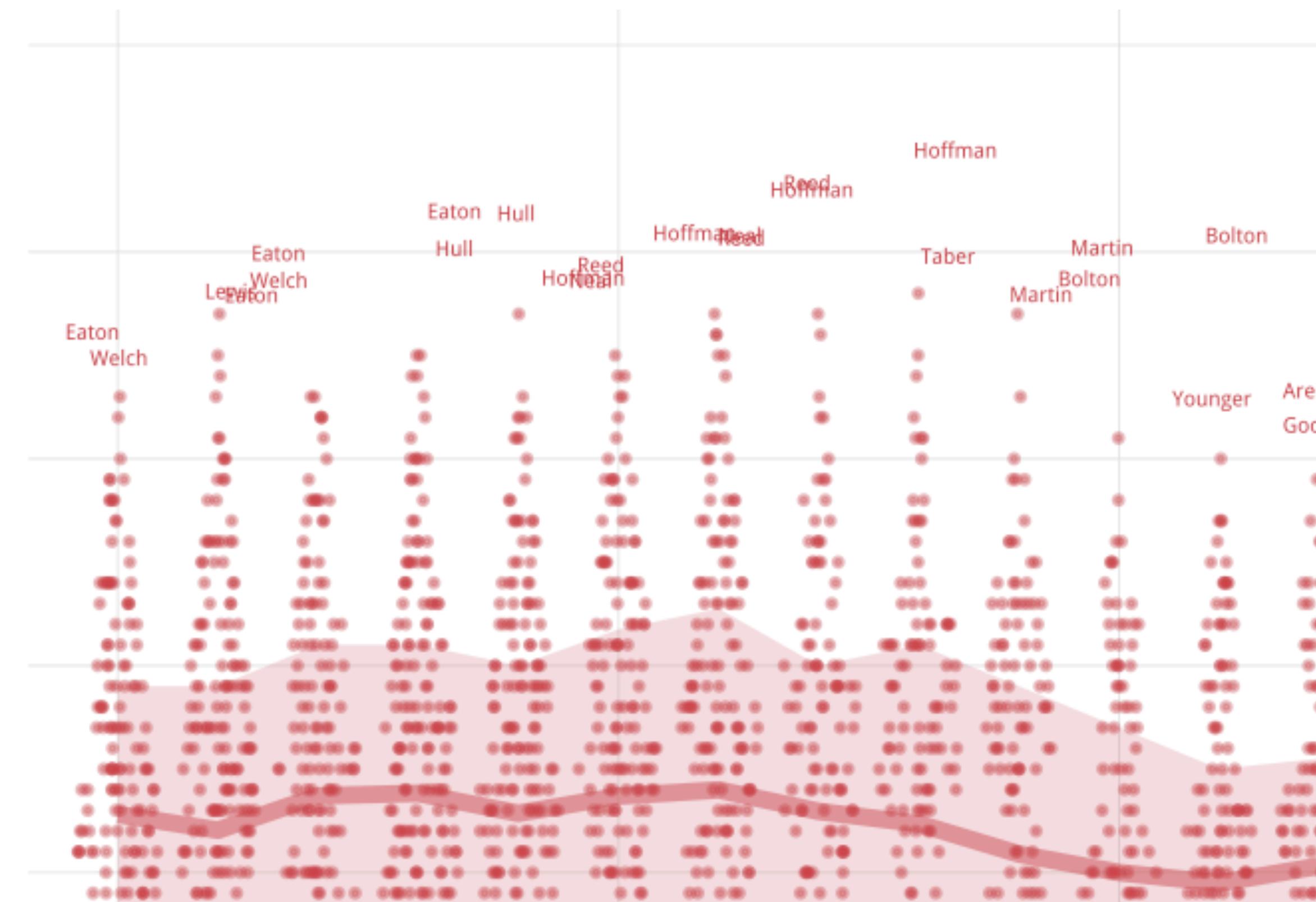
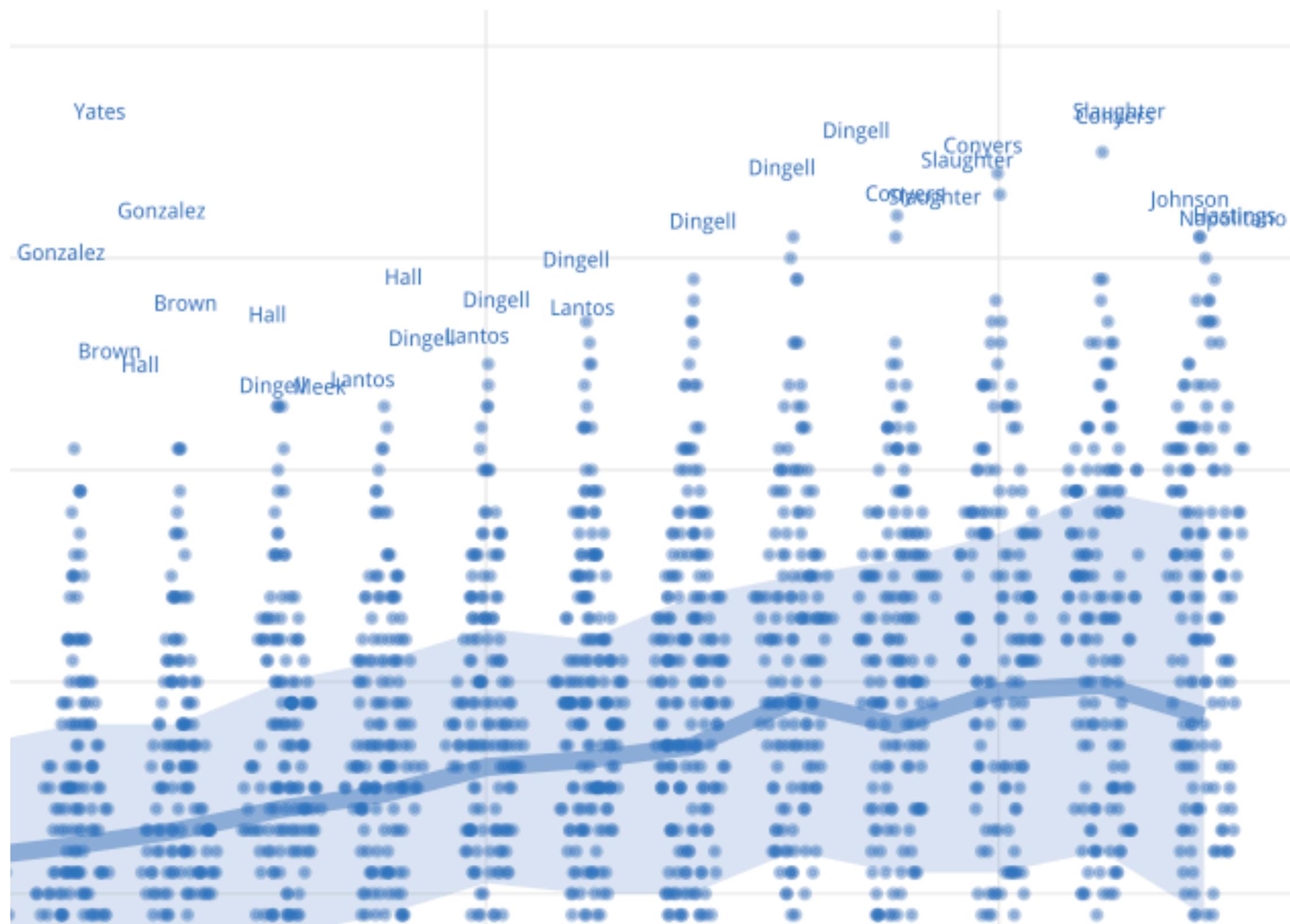
Democrats



Republicans



Show Ponies



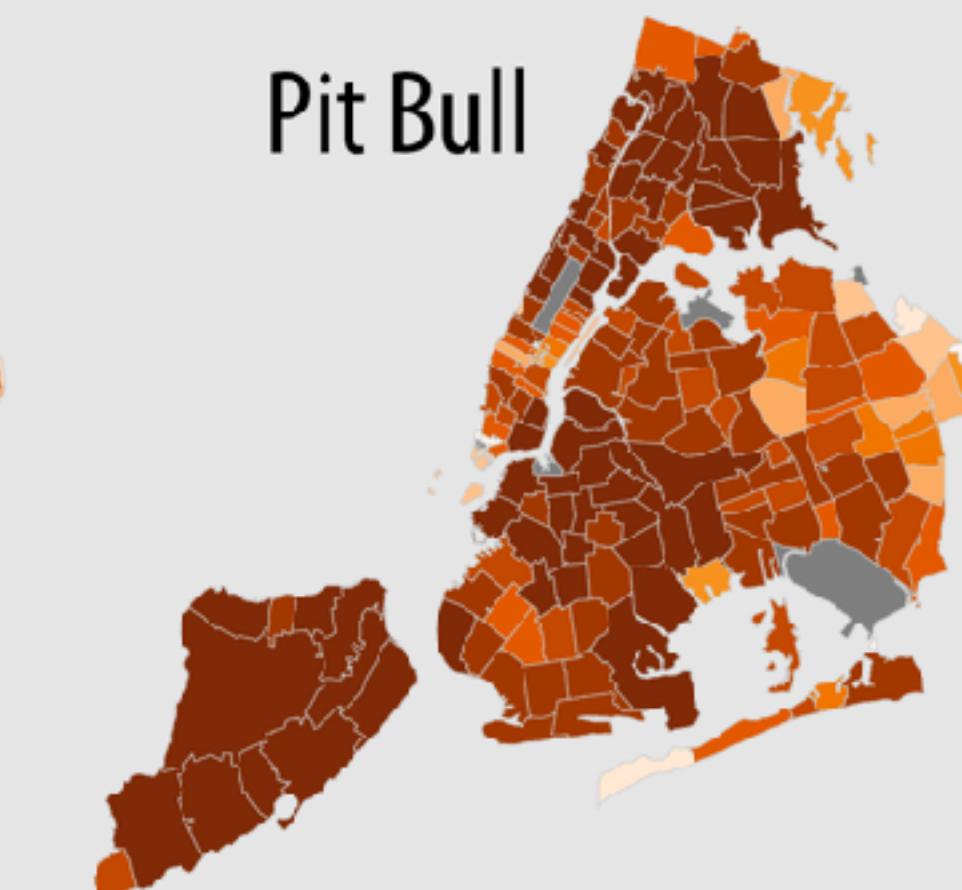
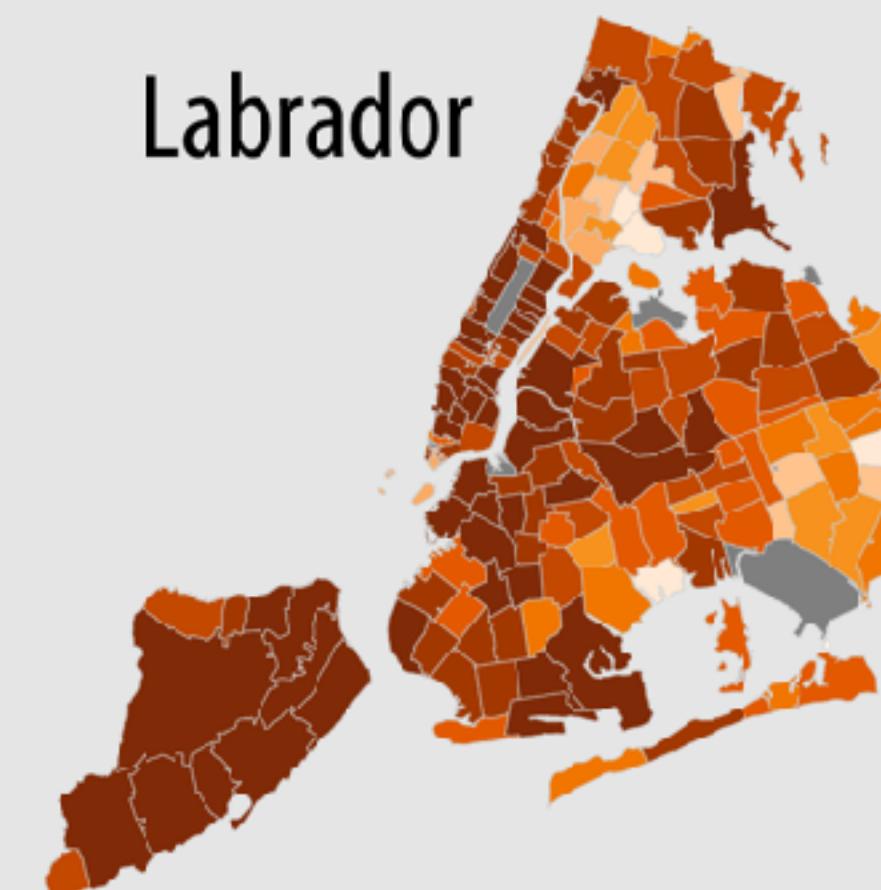
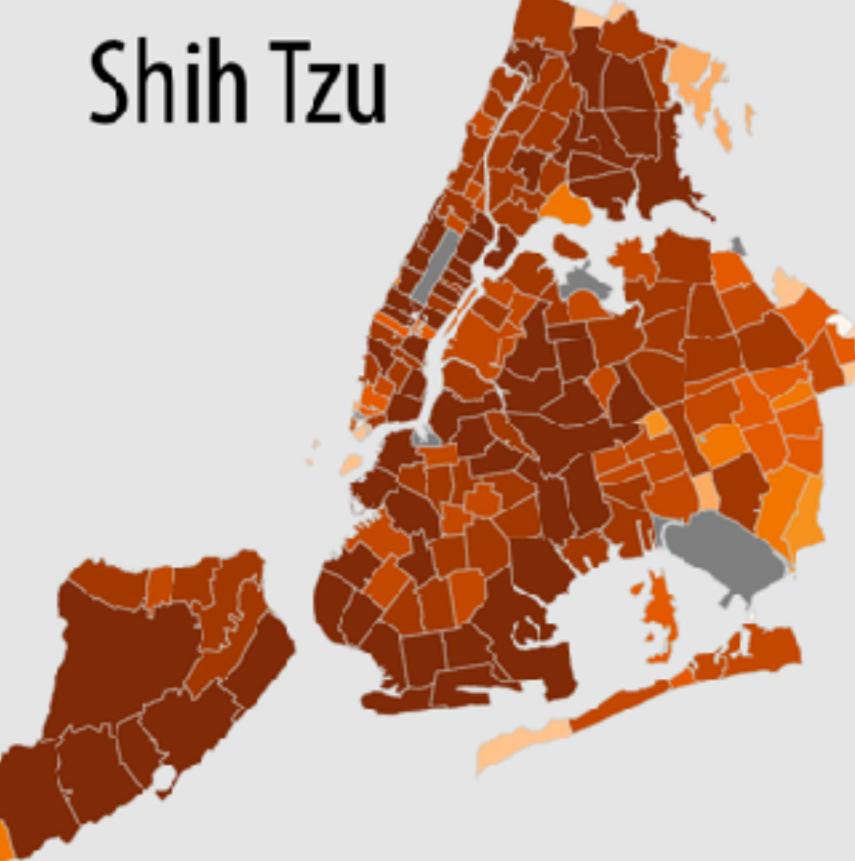
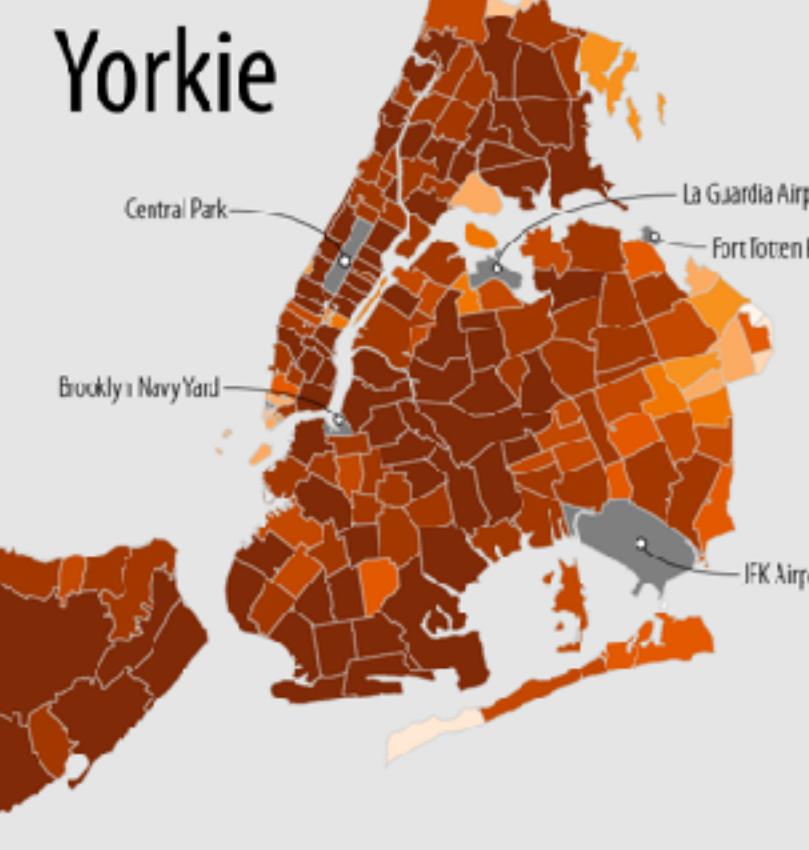
Relative prevalence of breed, from Most to Least



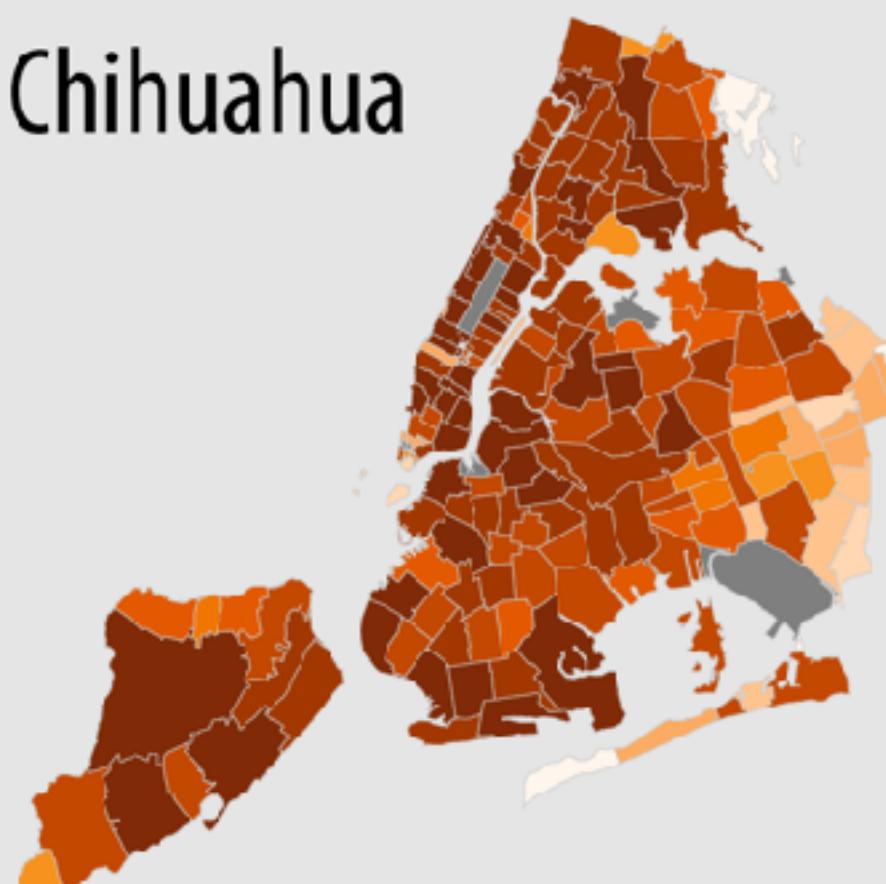
More Common

Less Common

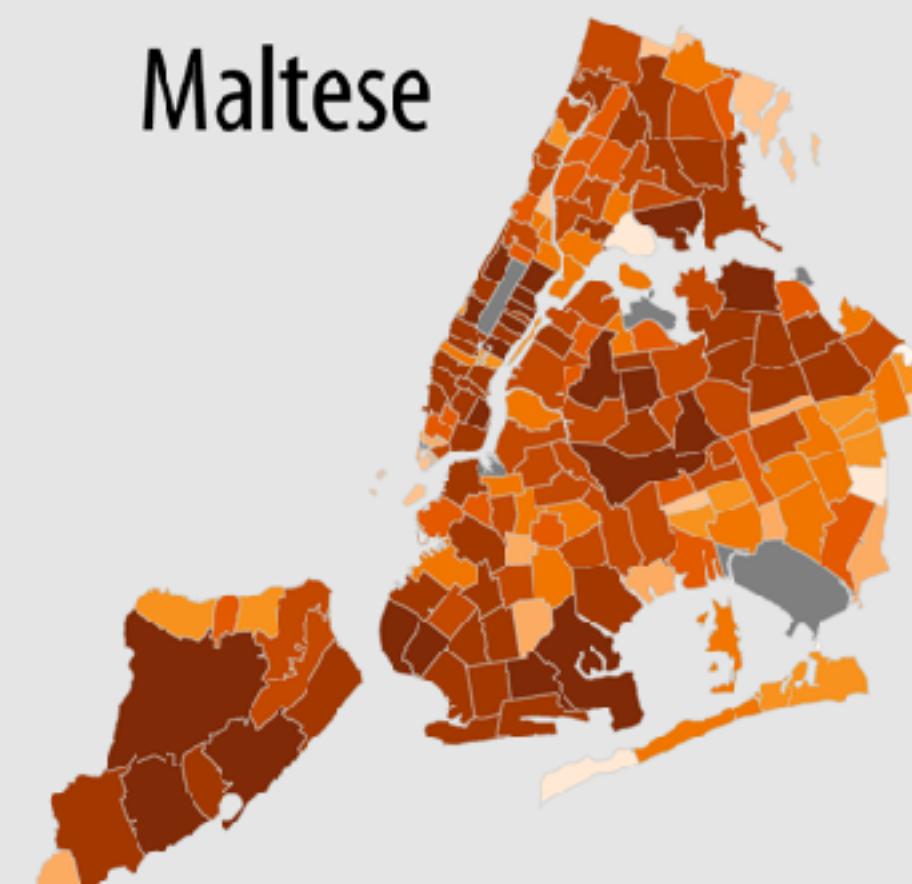
The most common
NYC breed is
“Mixed” or
“Unknown”



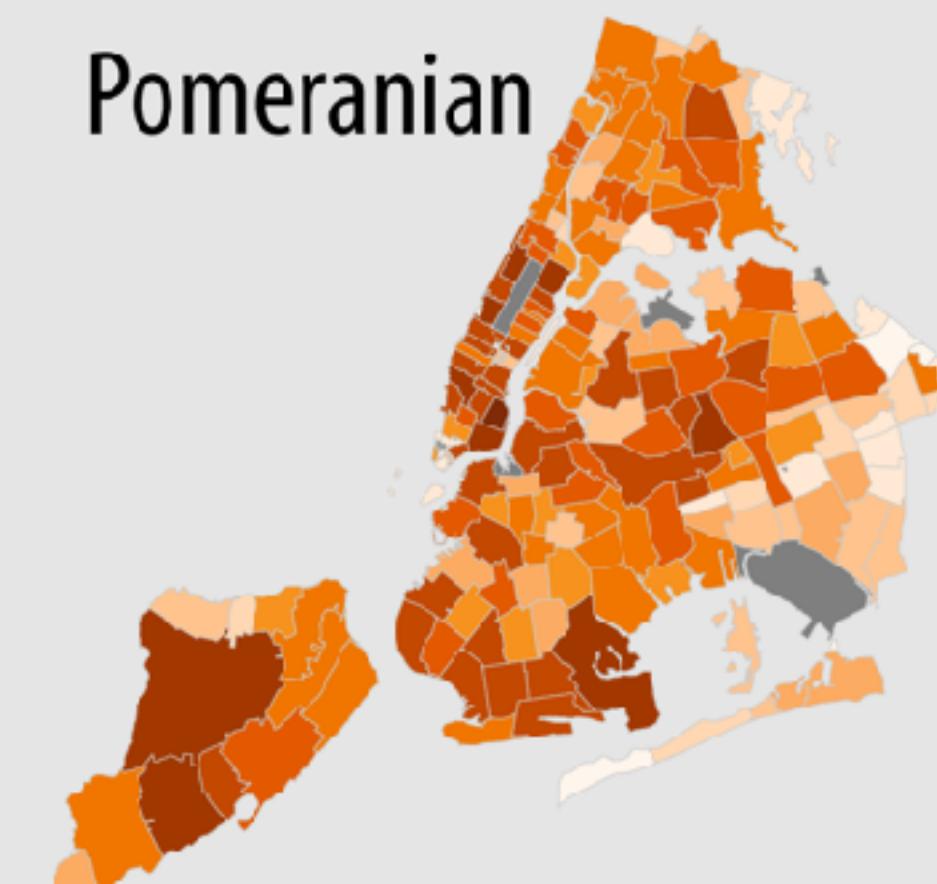
Chihuahua



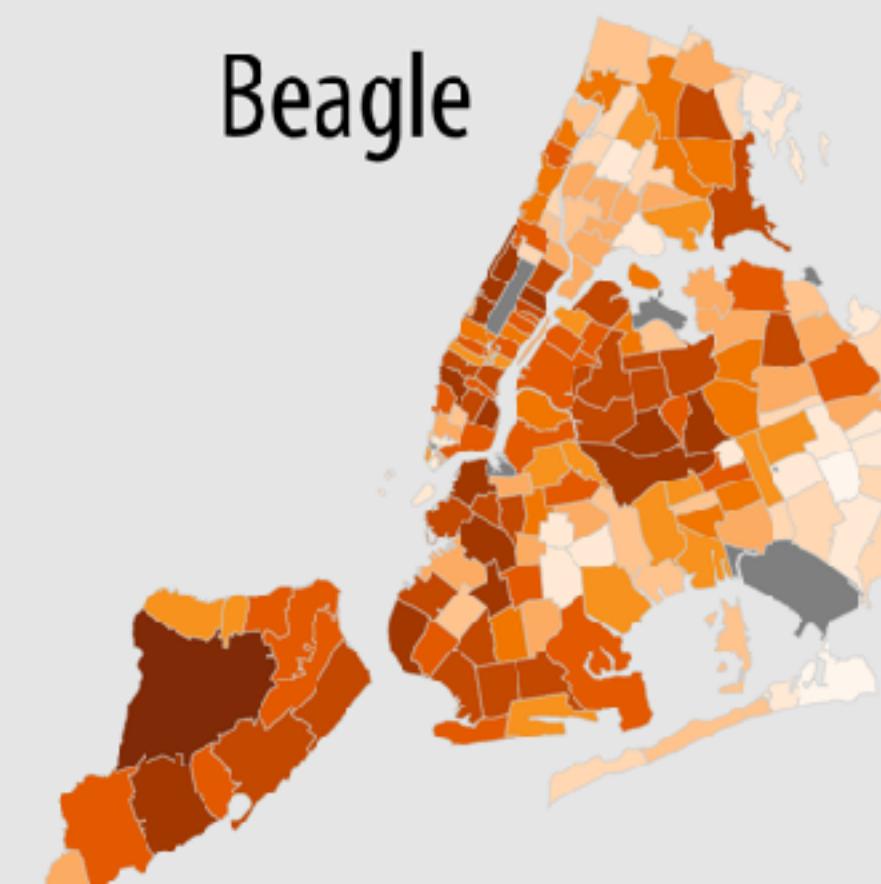
Maltese



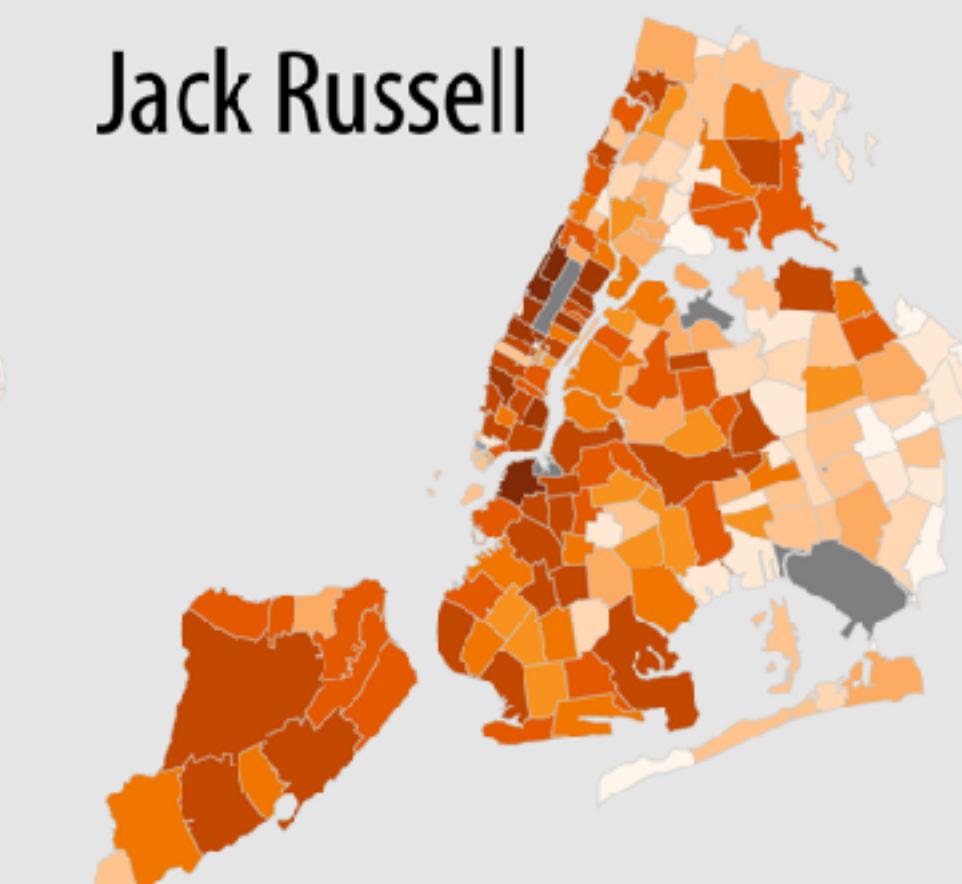
Pomeranian



Beagle

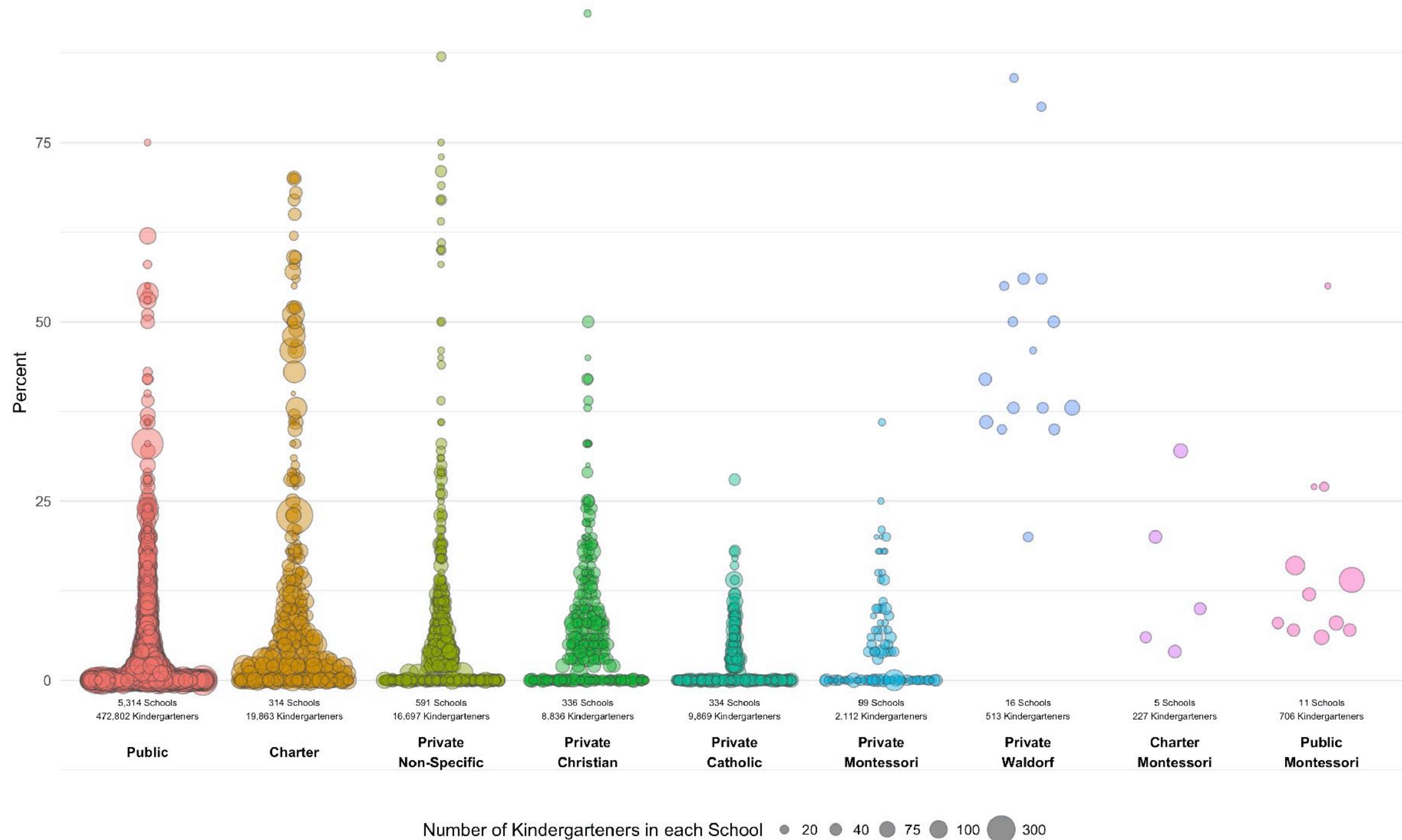


Jack Russell

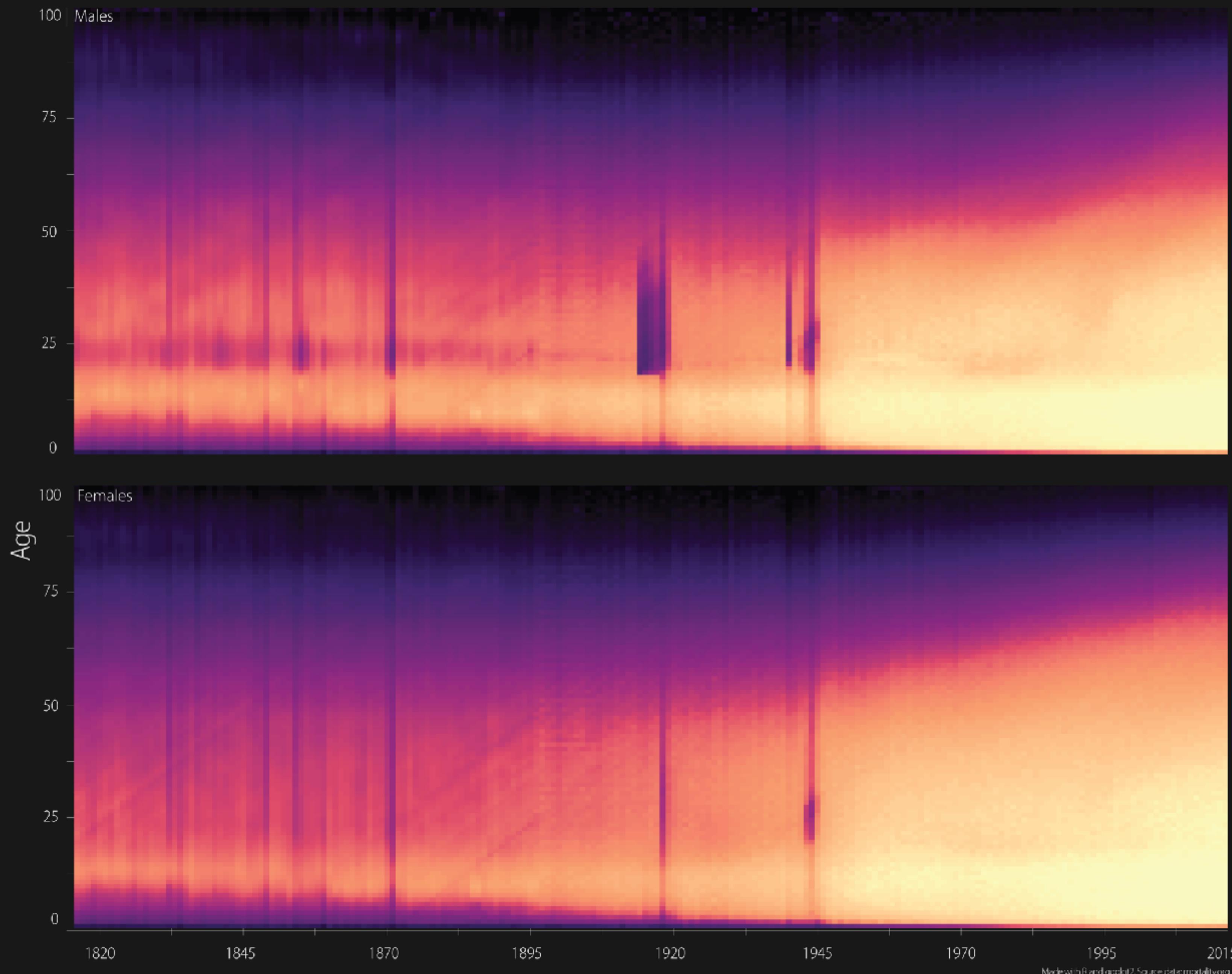


Vaccination Exemption Rates in California Kindergartens

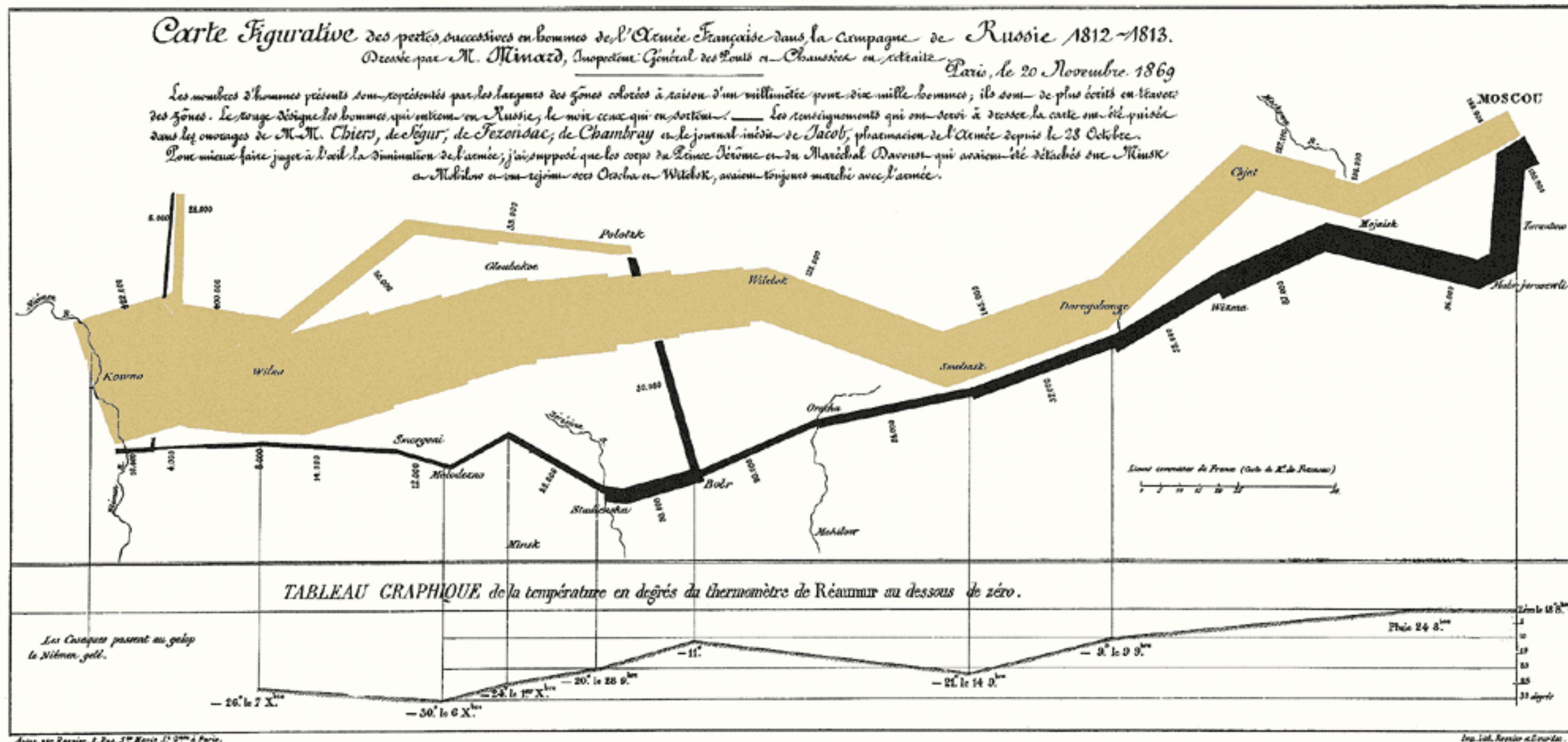
Percent of Kindergarteners with a Personal Belief Exemption, by Type and Size of School.

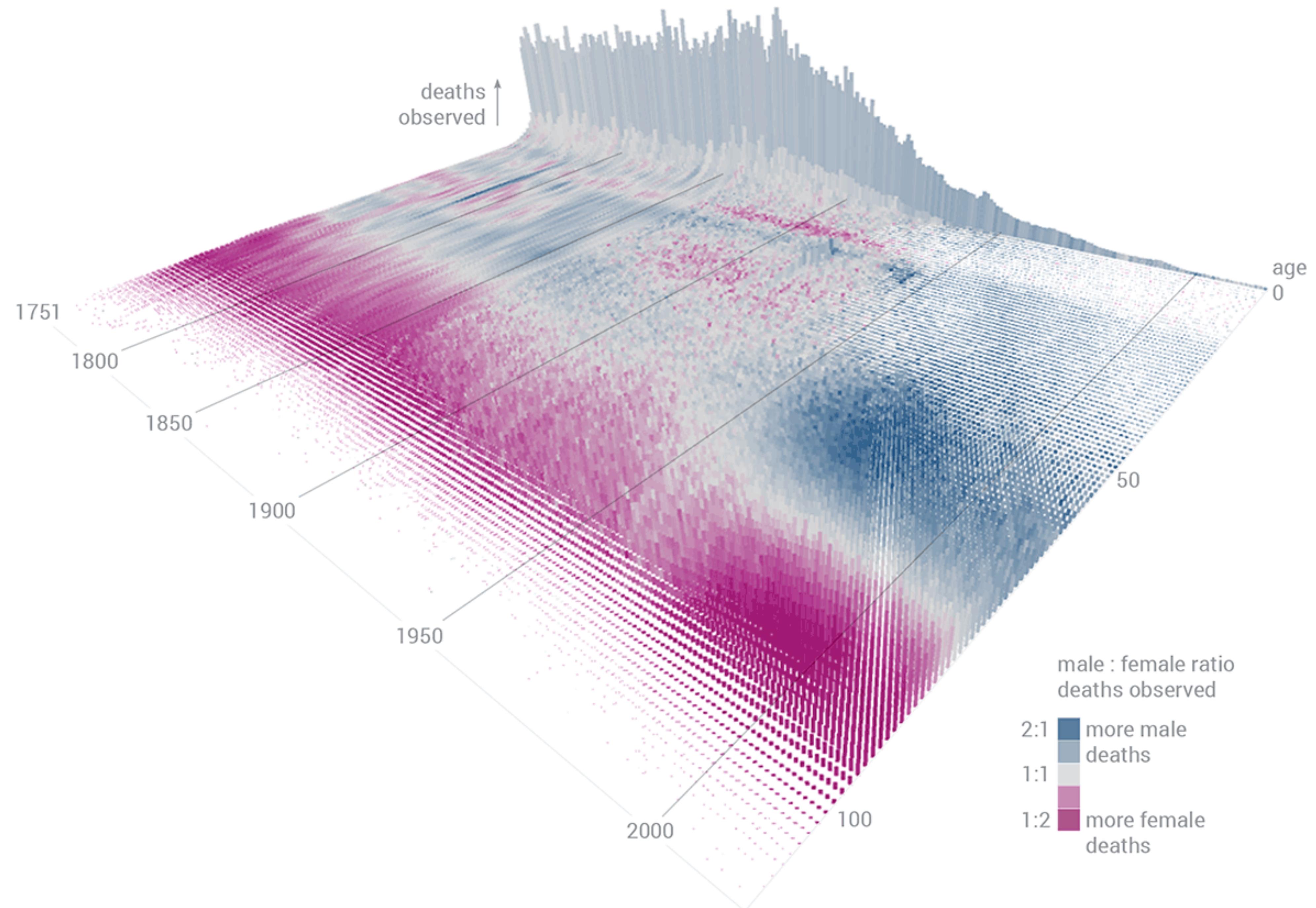


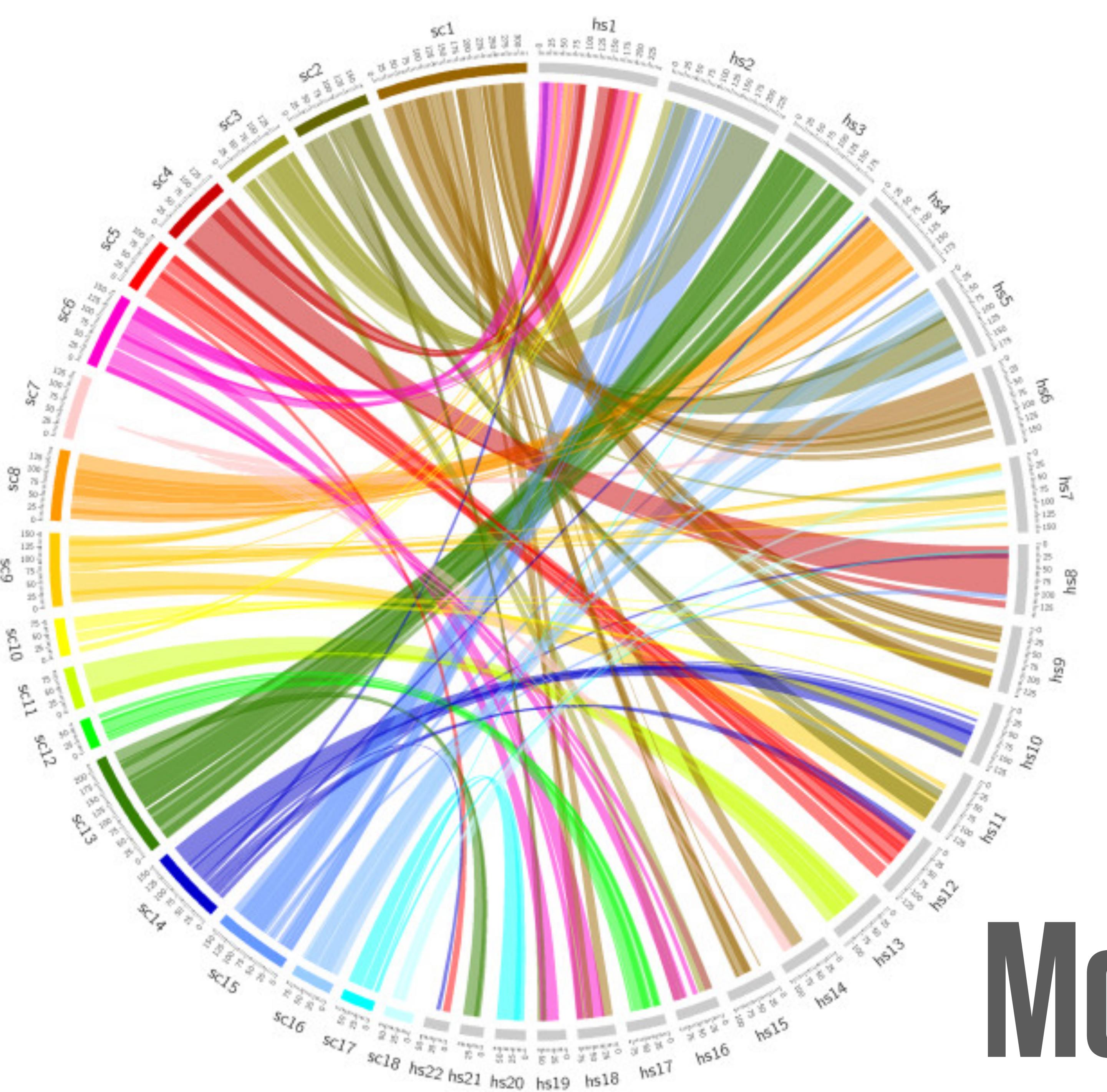
MORTALITY IN FRANCE 1816 – 2016



Unicorns ...







... or
Monsters