

Data Visualization





Introductions

Instructor:

Kieran Healy

[@kjhealy](#)

Assisted by:

Mara Averick

[@dataandme](#)

Arie Spiegel

[@ariespirgel](#)

Dewey Dunnington

[@paleolimbot](#)

Thomas Lin Pedersen

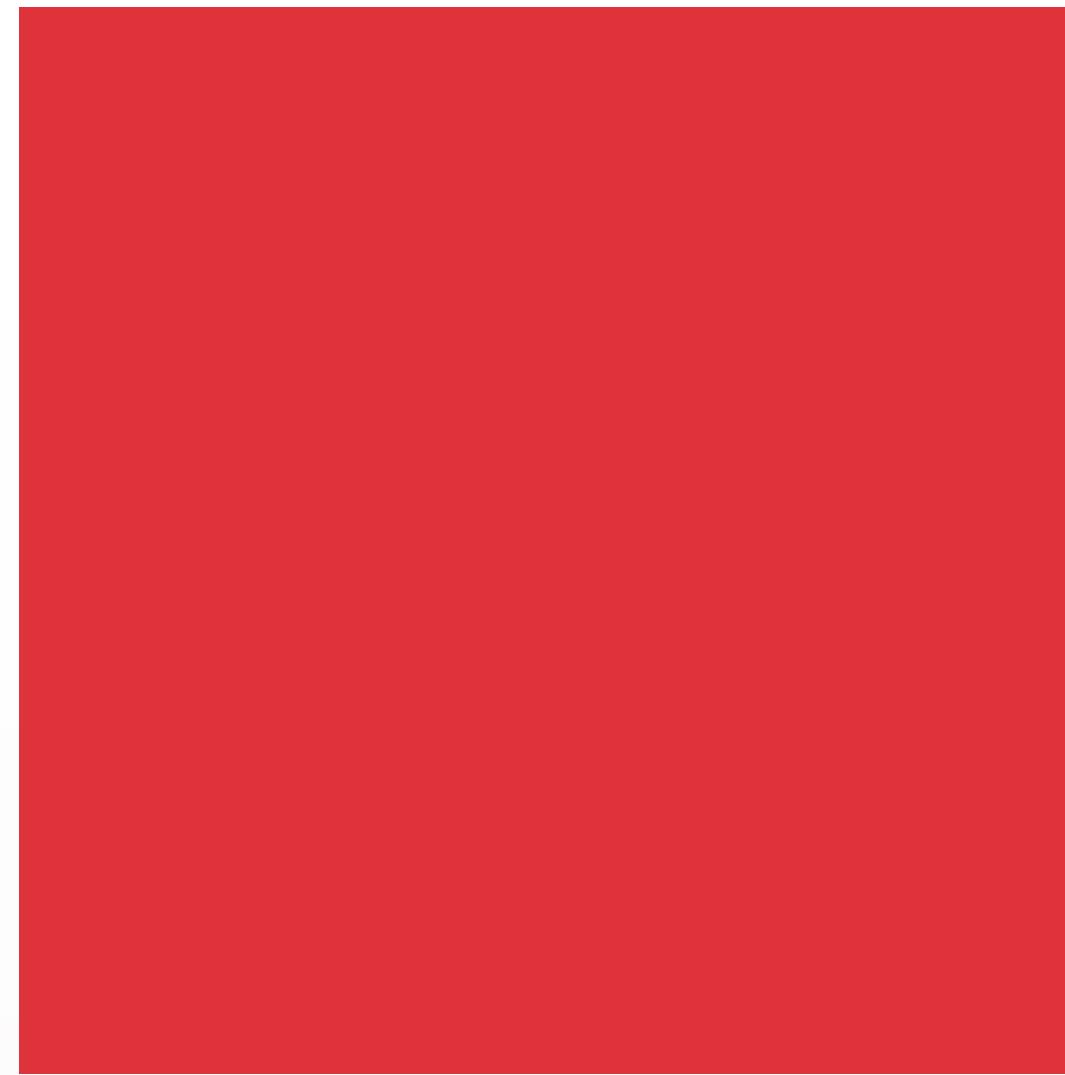
[@thomasp85](#) Coordinator



Workshop Policies

- **Identify** the exits closest to you in case of emergency
- **Please review** the `rstudio::conf` code of conduct that applies to all workshops. Issues can be addressed three ways:
 - In person: contact any `rstudio::conf` staff member or the conference registration desk
 - By email: send a message to conf@rstudio.com
 - By phone: call (844) 448-1212
- **Please do not** photograph people wearing red lanyards
- **A chill-out room** is available for neurologically diverse attendees on the 4th floor of Tower 1

Asking for help



I'm stuck



I'm not stuck, but I need help
on my computer



I need help
understanding something
(which likely means others
do too)

Asking for help



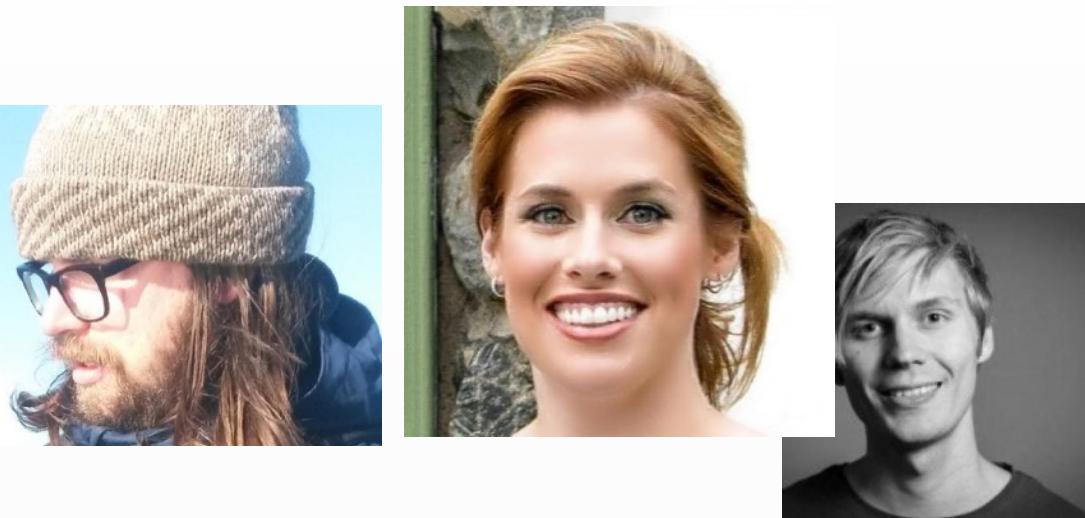
"..."

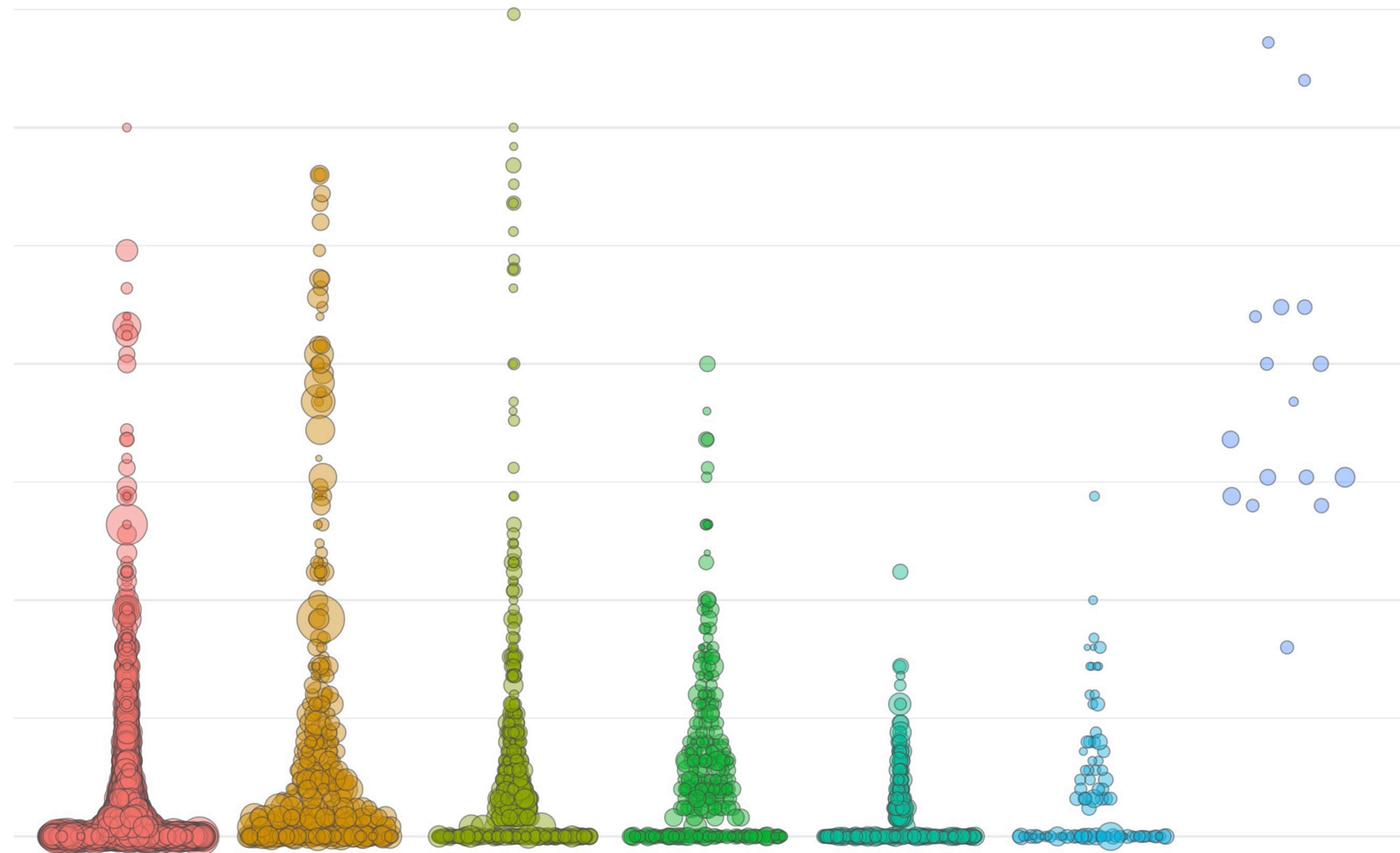


I'm not stuck,

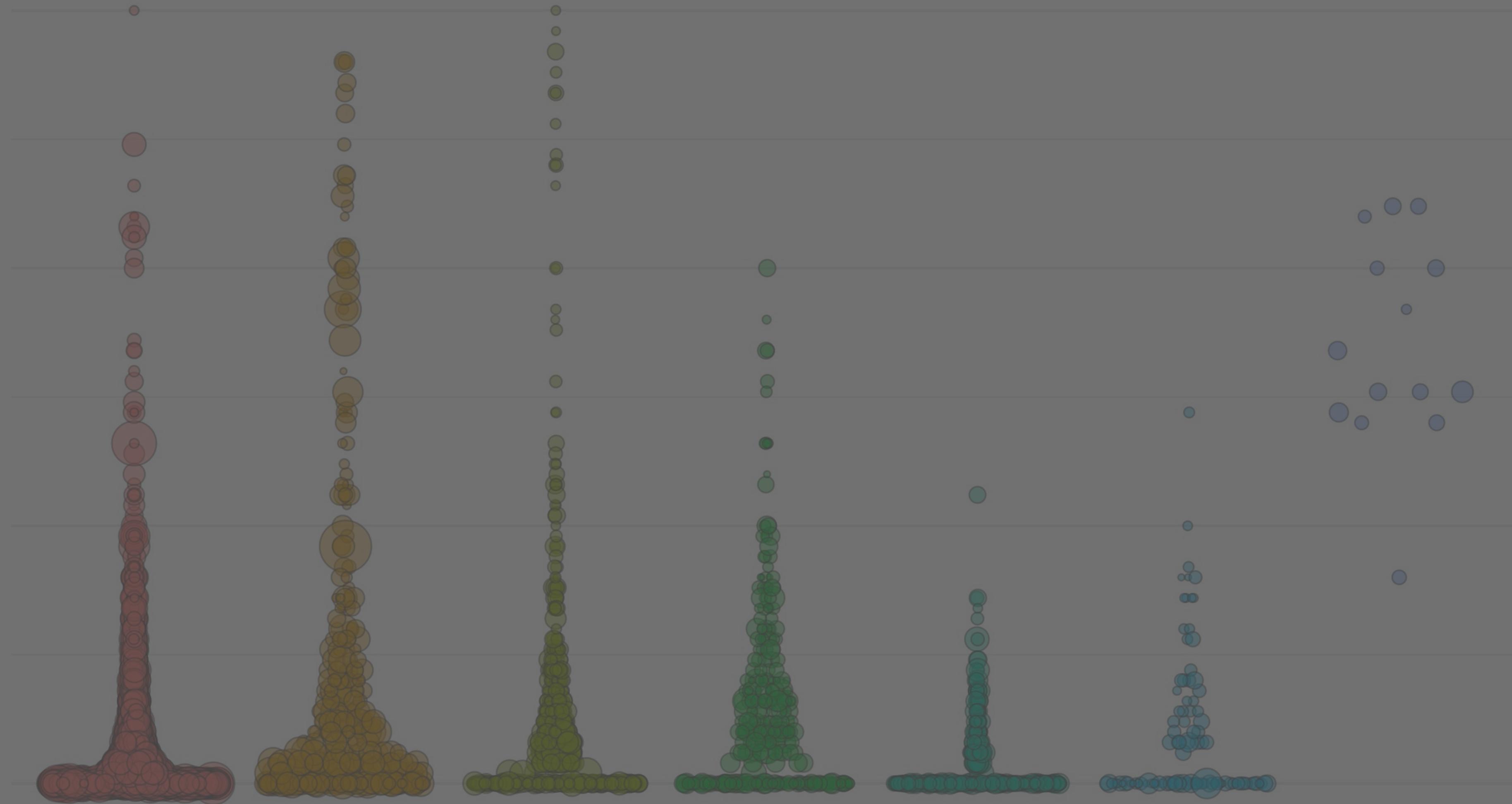


I need help
understandin





DATA VISUALIZATION





RStudio

Get Up & Running

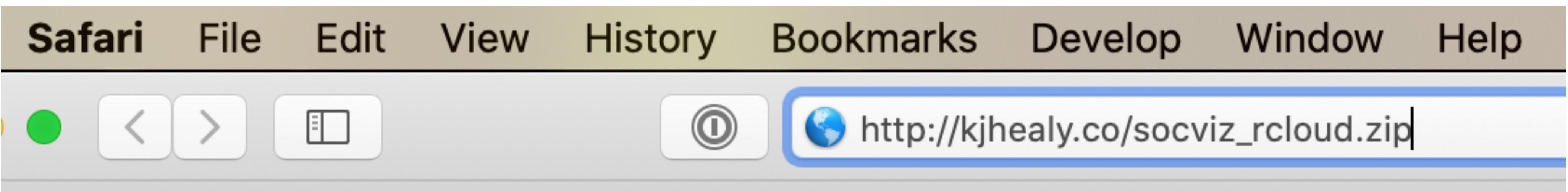
Console

~/Desktop/

Get the Notes from inside RStudio

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

Or just download from your Browser



<https://bit.ly/30VyeKf>

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

51% 24

~/Downloads/socviz_rcloud - master - RStudio

Console Terminal Jobs

~/Downloads/socviz_rcloud/

```
R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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

> |
```

Environment History Files Build Git

New Folder Delete Rename More

Home > Downloads > socviz_rcloud

Name	Size	Modified
..	40 B	Jan 24, 2020, 11:35 AM
.gitignore	4 KB	Jan 18, 2020, 10:12 AM
01_introduction.Rmd	3.3 KB	Jan 18, 2020, 10:12 AM
02_get_started.Rmd	5.8 KB	Jan 18, 2020, 10:12 AM
03_make_a_plot.Rmd	4.8 KB	Jan 23, 2020, 5:23 PM
04_show_the_right_numbers.Rmd	9.9 KB	Jan 23, 2020, 5:24 PM
05_tables_and_labels.Rmd	14.7 KB	Jan 18, 2020, 12:03 PM
06_models.Rmd	12.5 KB	Jan 18, 2020, 10:12 AM
07_maps.Rmd	12.8 KB	Jan 18, 2020, 10:12 AM
08_refine_plots.Rmd	16.9 KB	Jan 18, 2020, 8:27 PM
09_supplementary_material.Rmd		

Plots Connections Packages Help Viewer

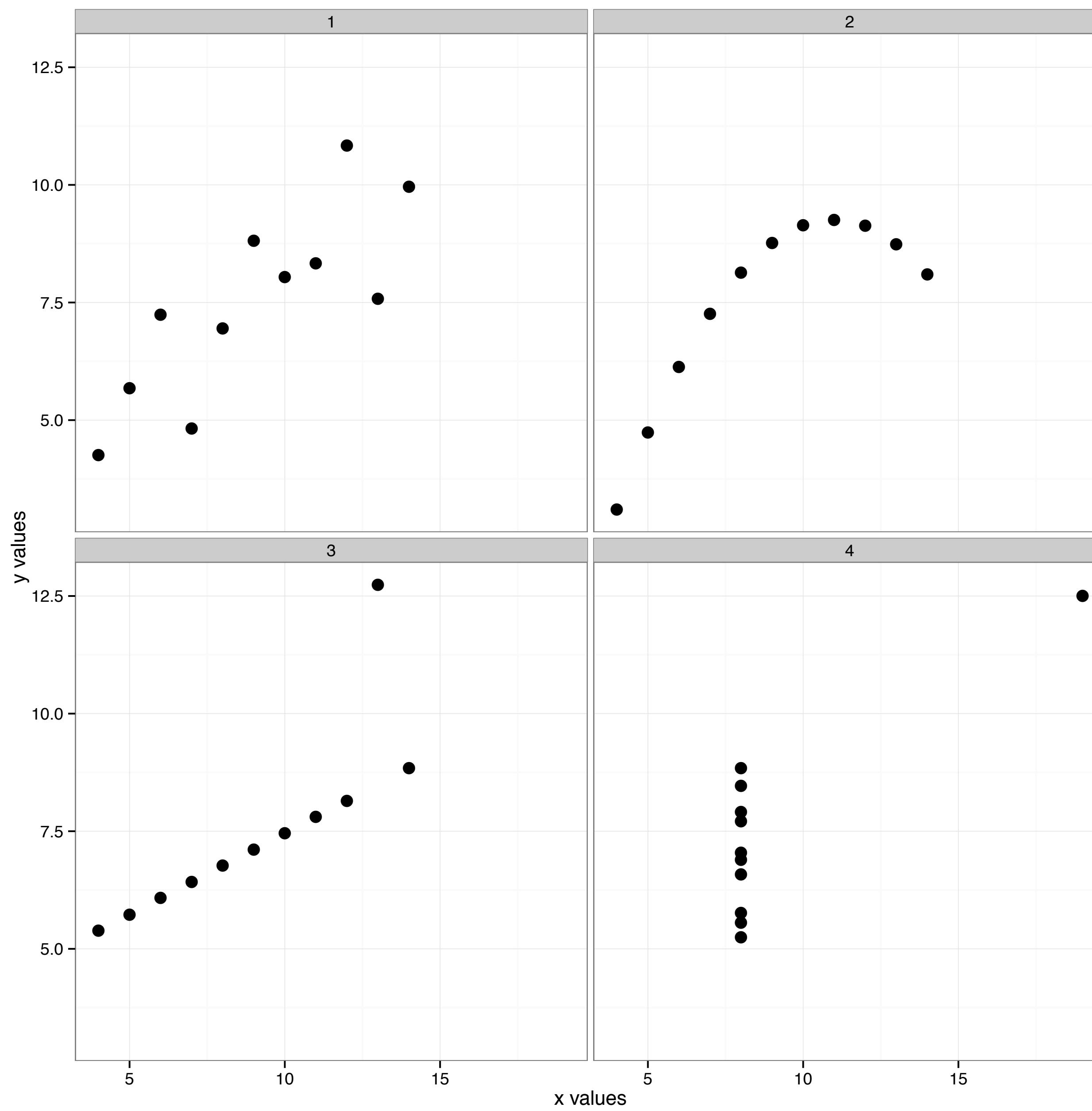
Zoom Export

socviz_rcloud

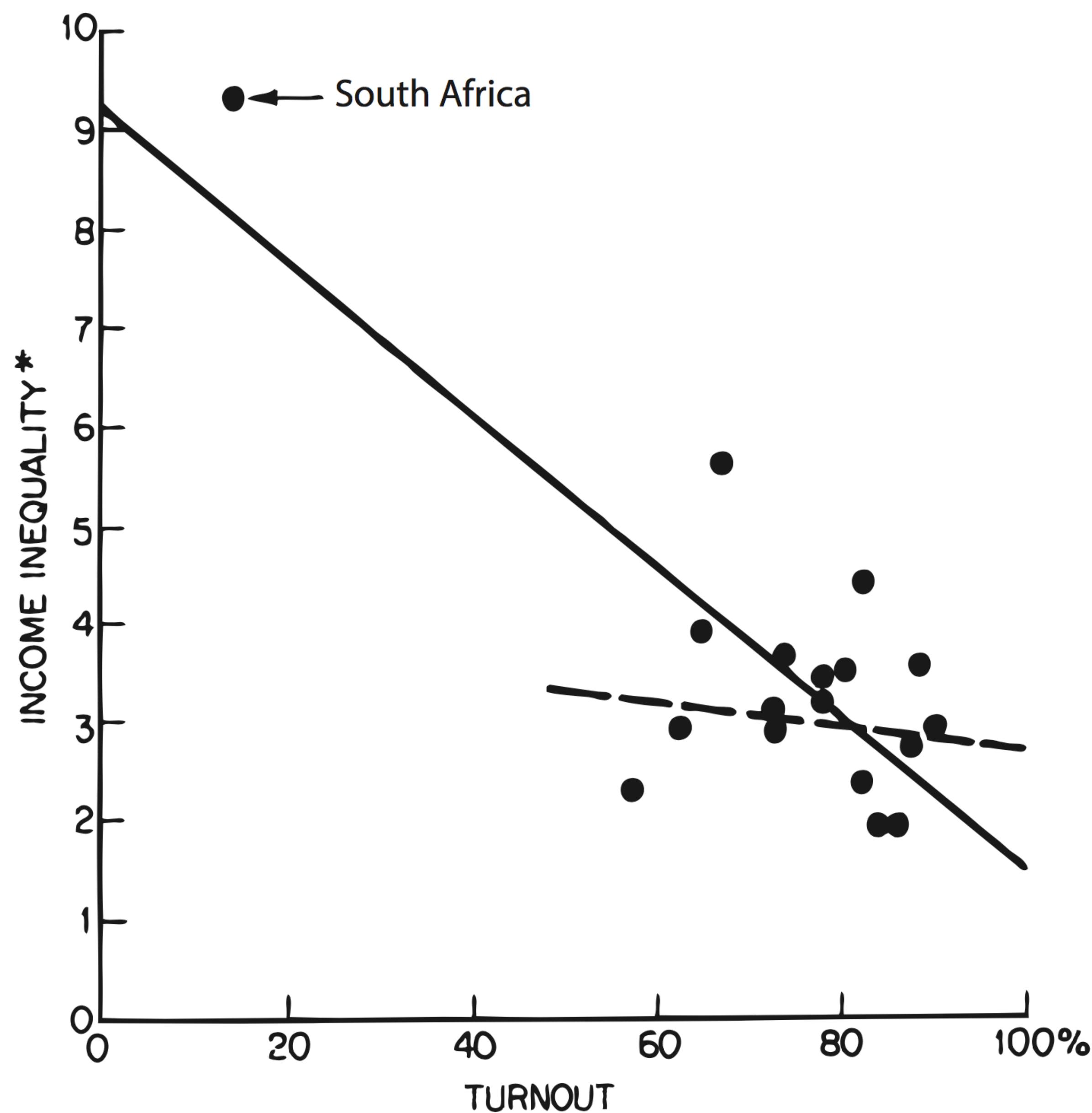
LOOK AT YOUR DATA



SEEING THINGS



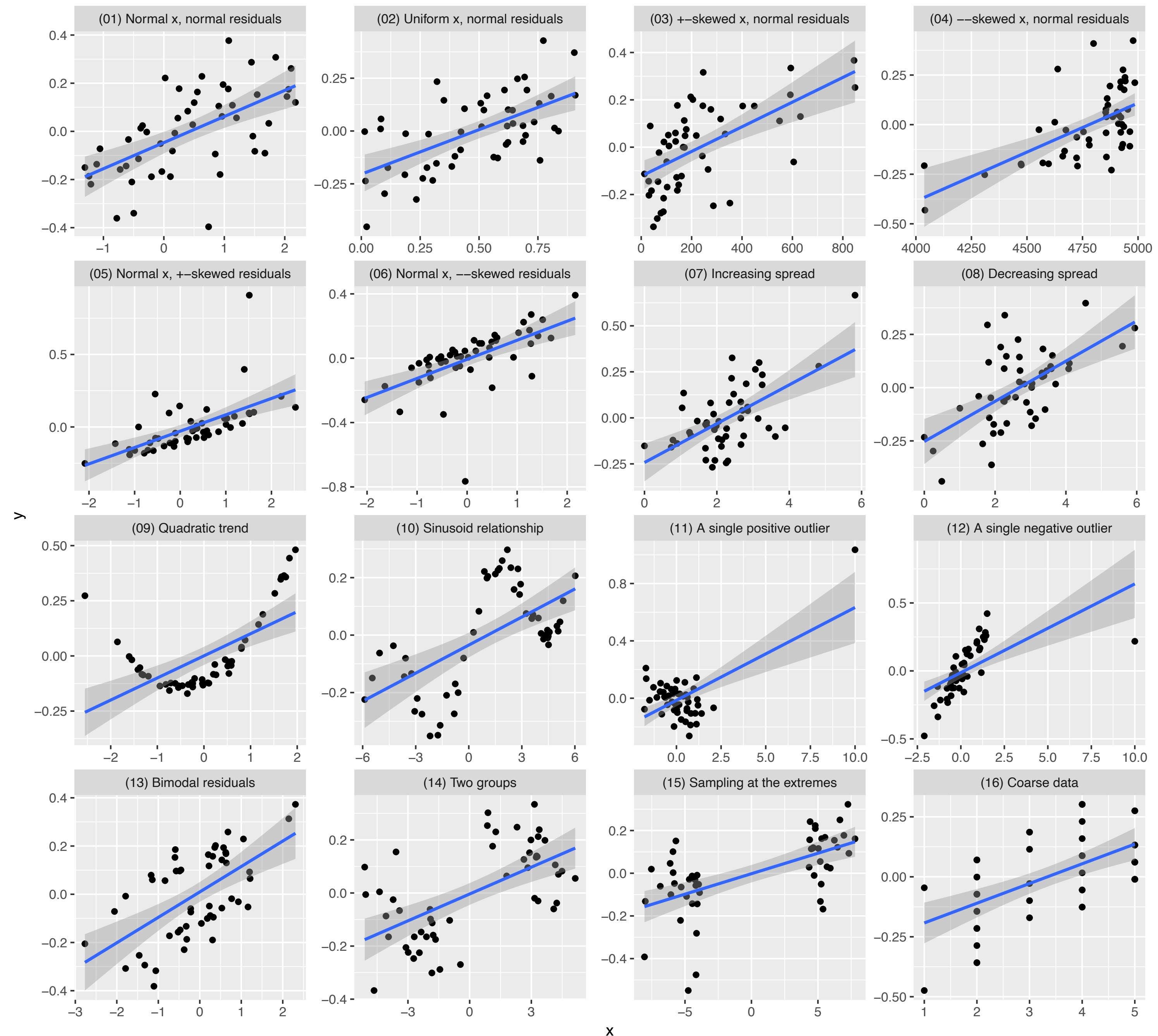
For all panels, N=11; Mean=7.5; Regression: $Y=3 + 0.5(X)$; SE of slope estimate: 0.118, t=4.24; Sum of Squares (X-X): 100; r=0.82.

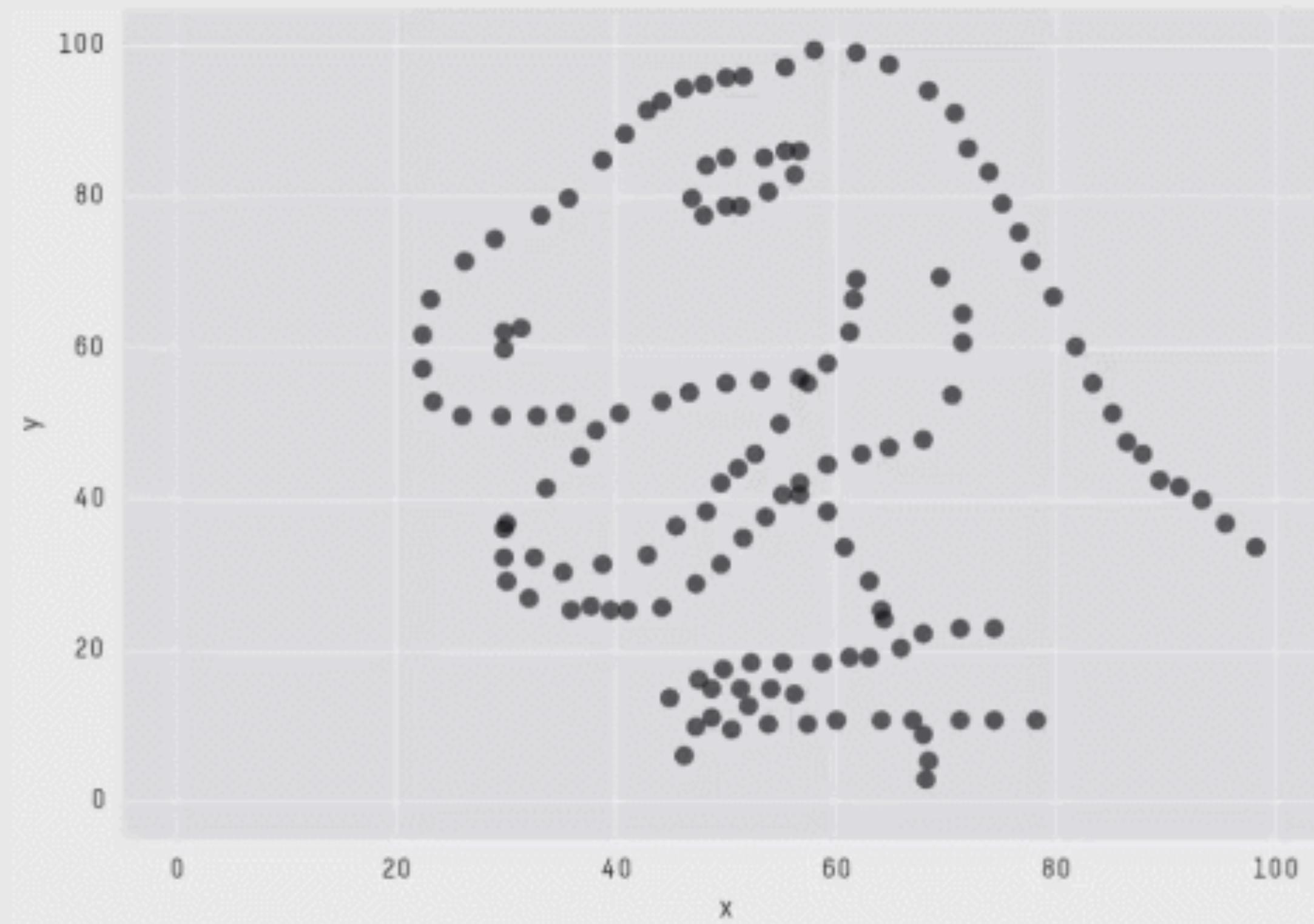


Key.

— Bivariate slope including South Africa ($N = 18$)

- - - Bivariate slope excluding South Africa ($N = 17$)

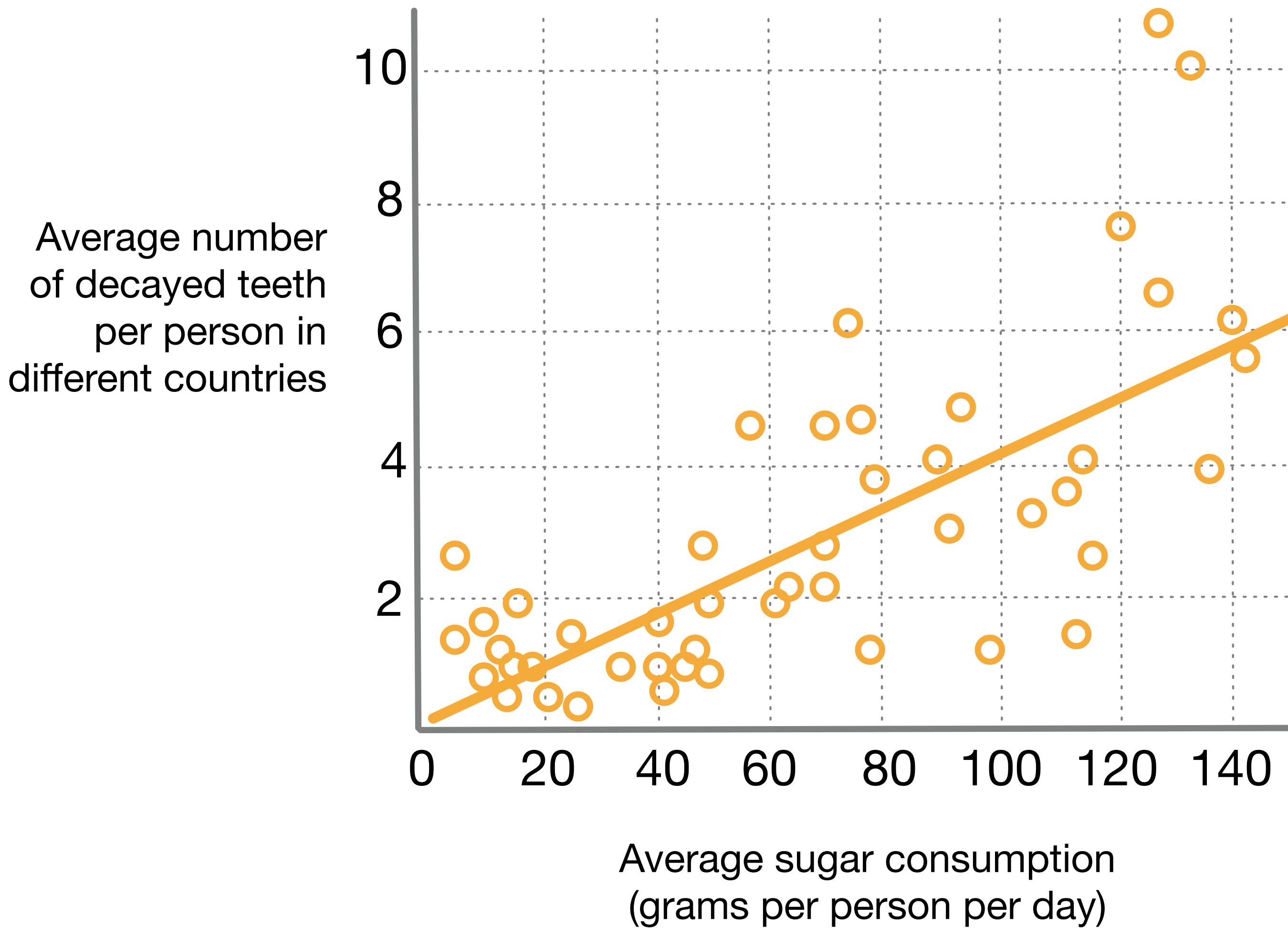




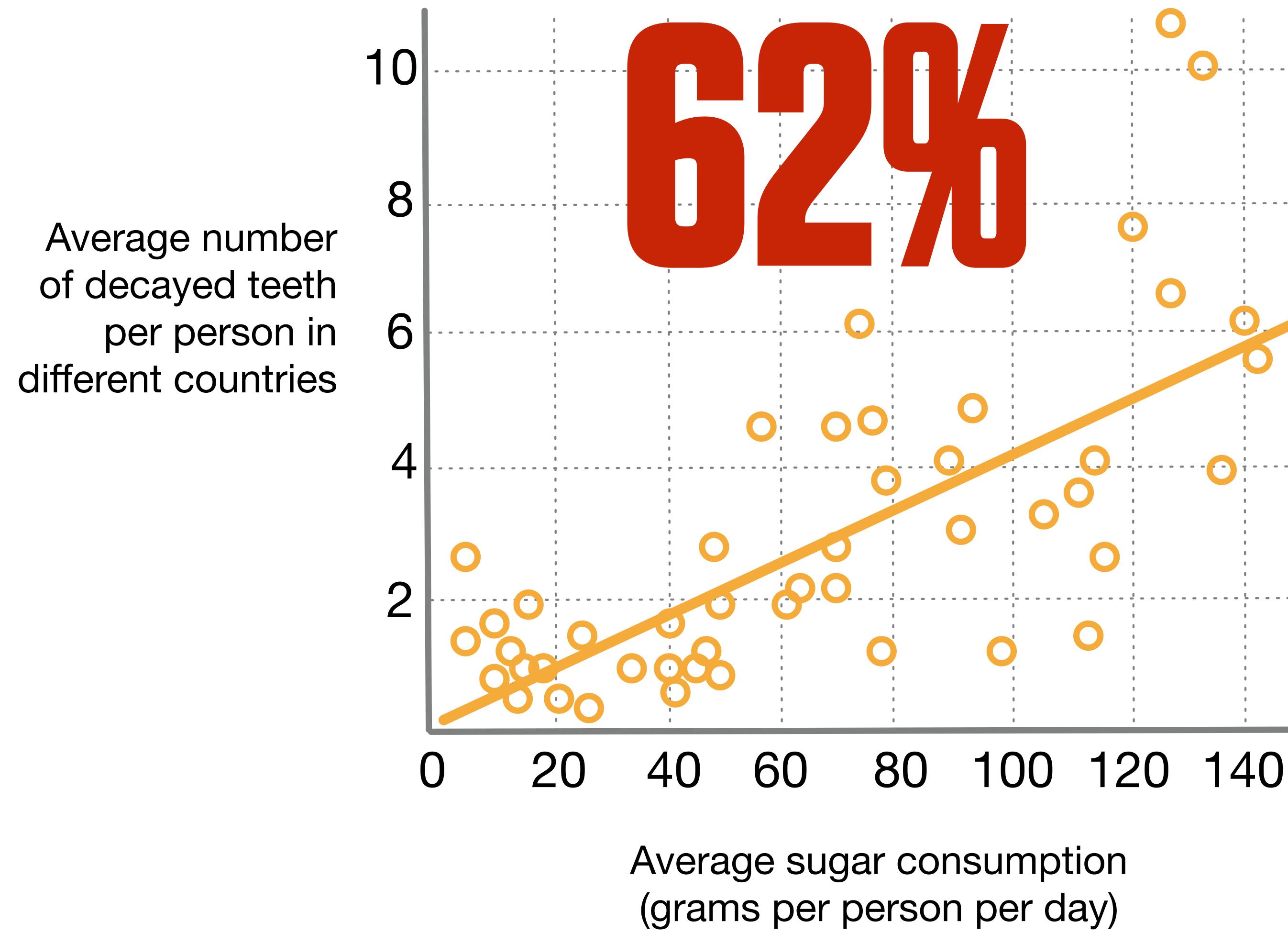
X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Cairo; Matejka & Fitzmaurice

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



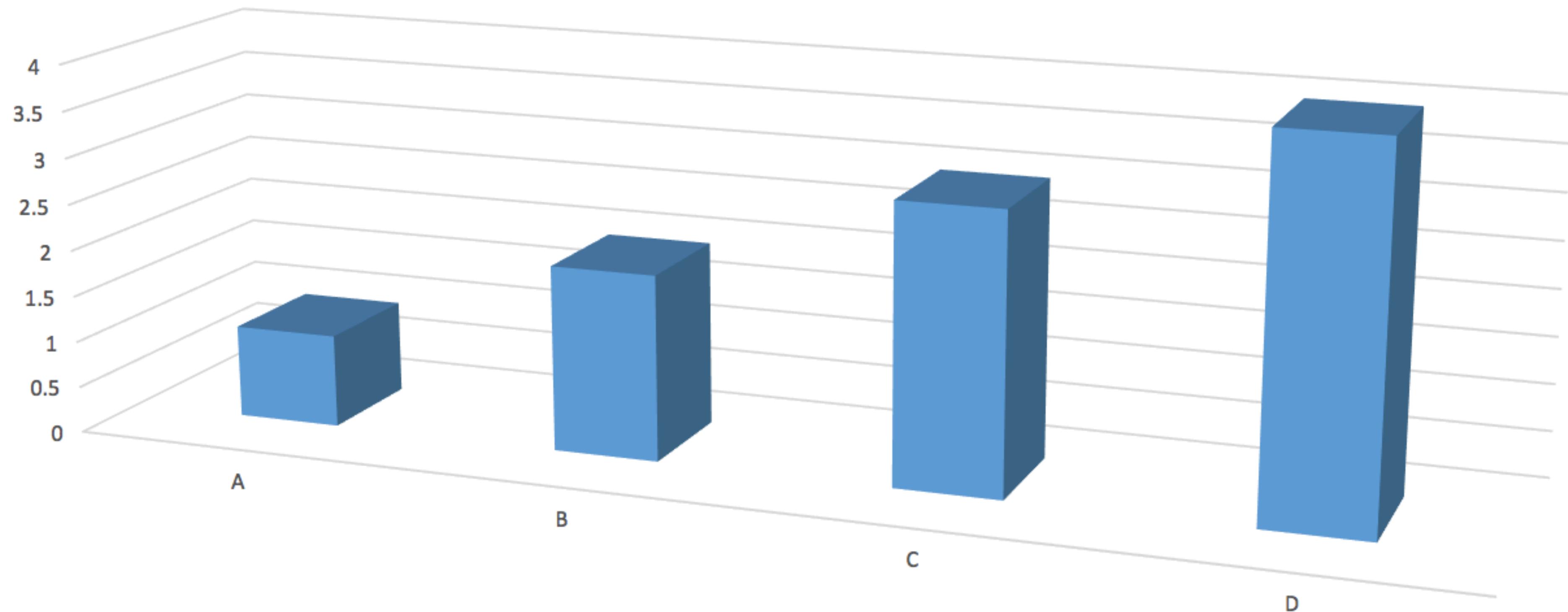
Which of the following statements best describes the data in the graph below?

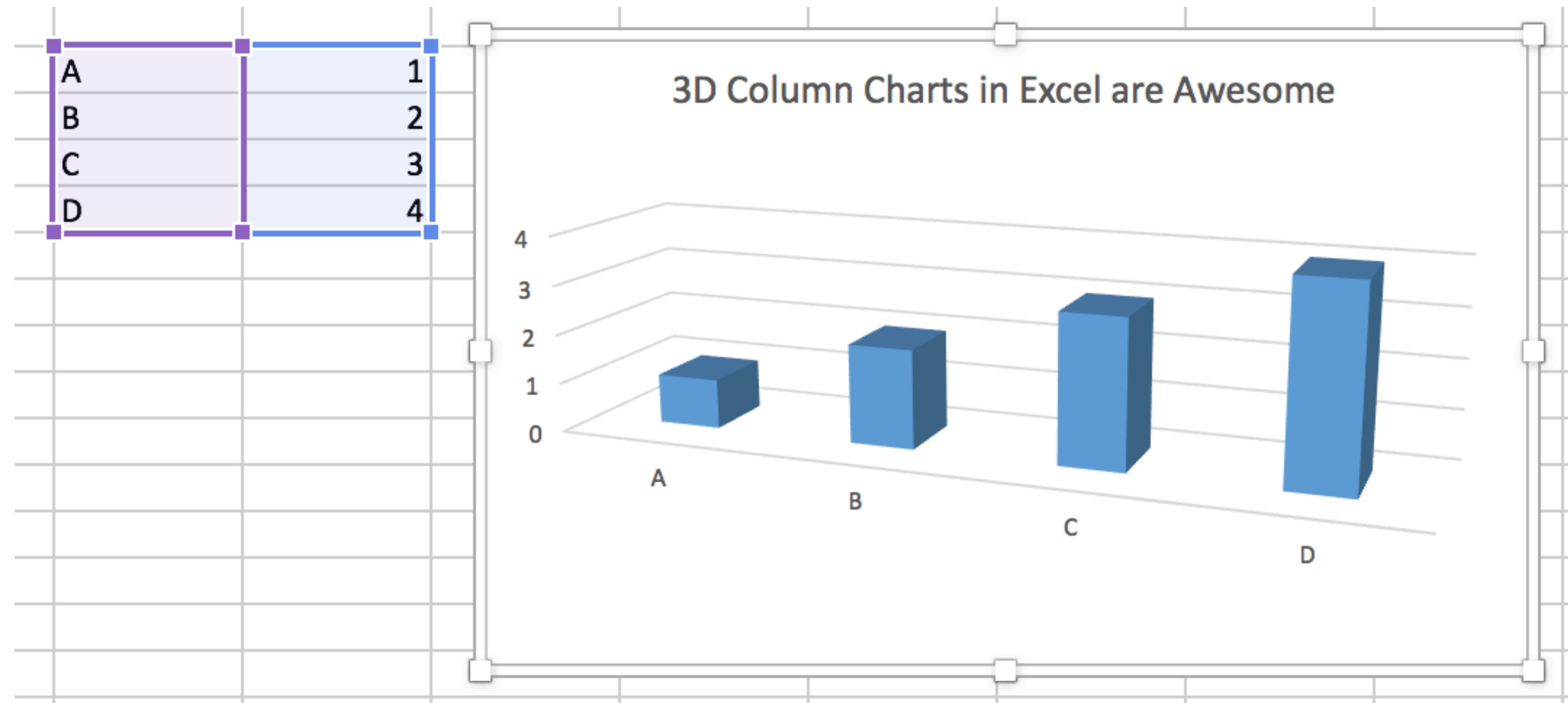


- A. In recent years, the rate of cavities has increased in many countries
- B. In some countries, people brush their teeth more frequently than in other countries
- C. The more sugar people eat, the more likely they are to get cavities**
- D. In recent years, the consumption of sugar has increased in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

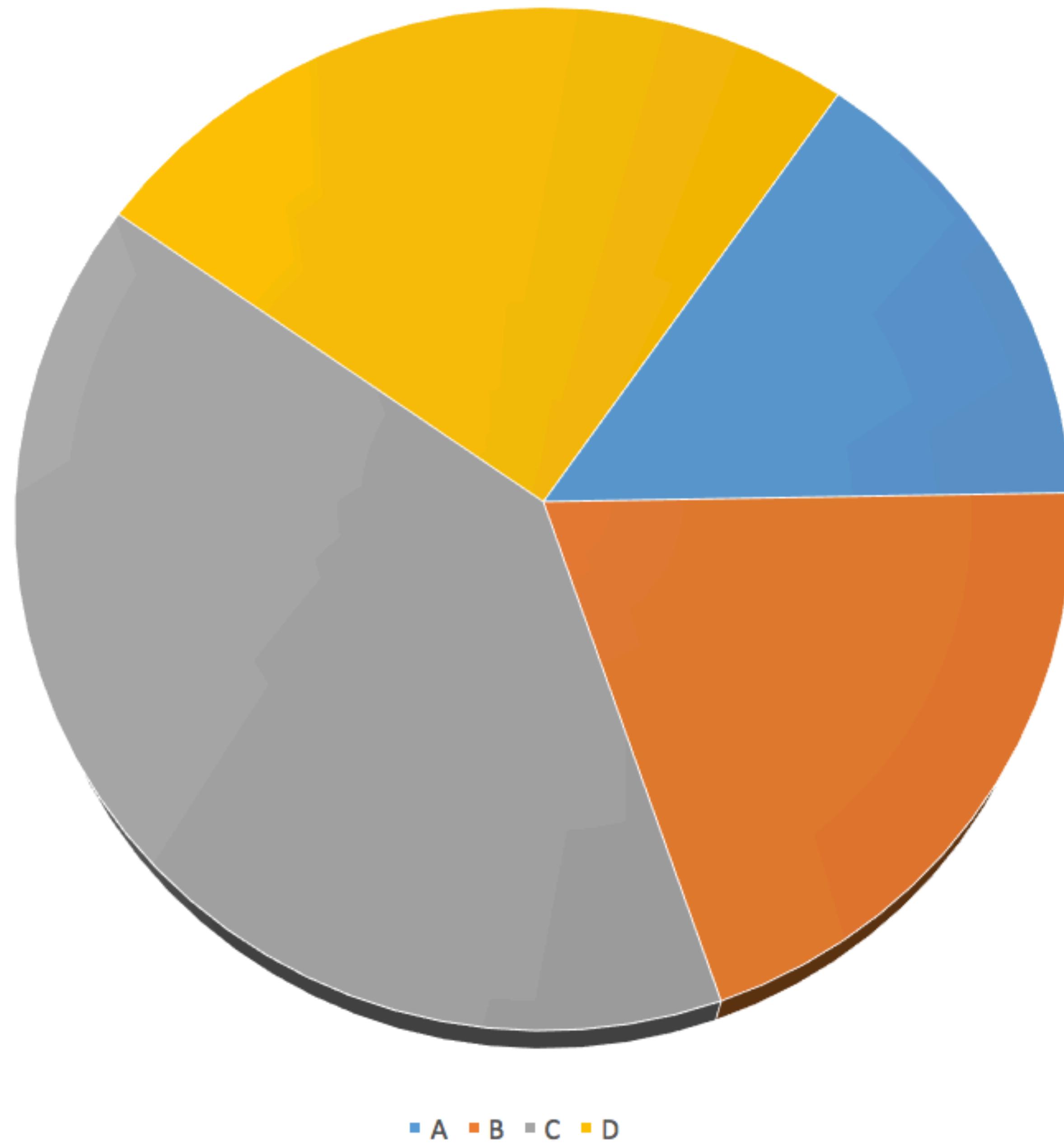
NOT
SEEING THINGS

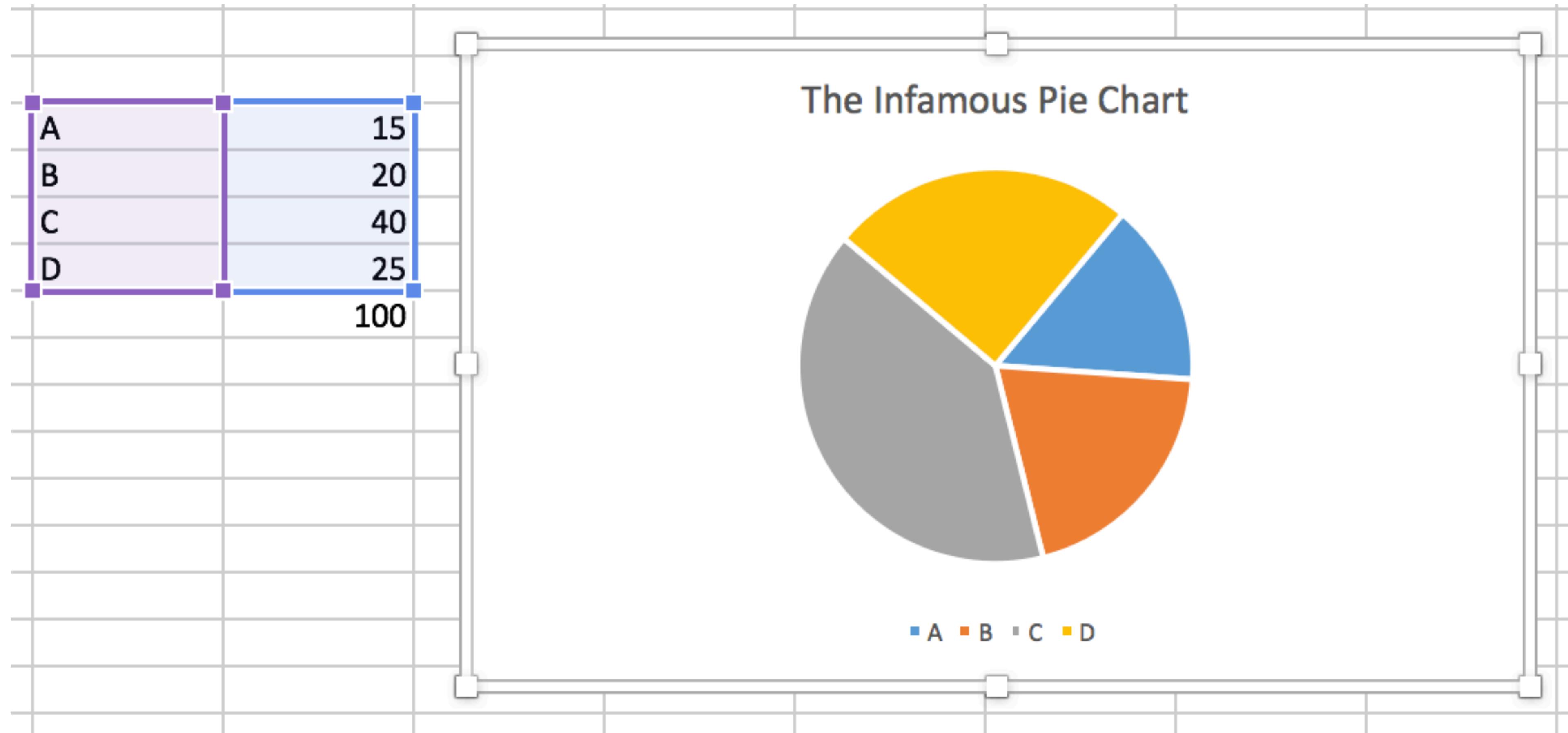
3D Column Charts in Excel are Awesome





The Infamous Pie Chart





BAD TASTE

BAD DATA

BAD PERCEPTION

TASTE

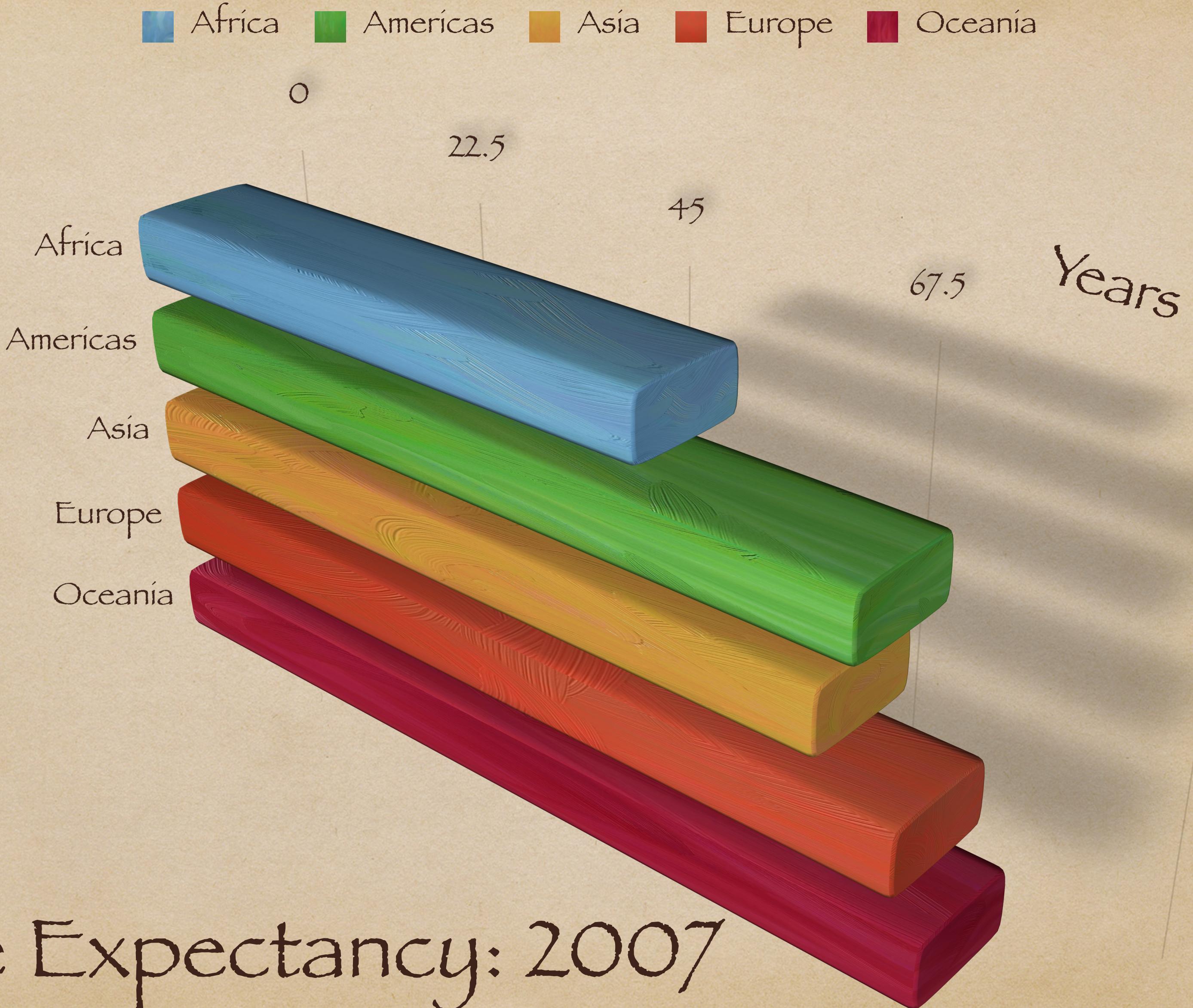
SIMPLIFY, SIMPLIFY?

The Visual Display
of Quantitative Information

EDWARD R. TUFTE

Life Expectancy: 2007

CONTINENT



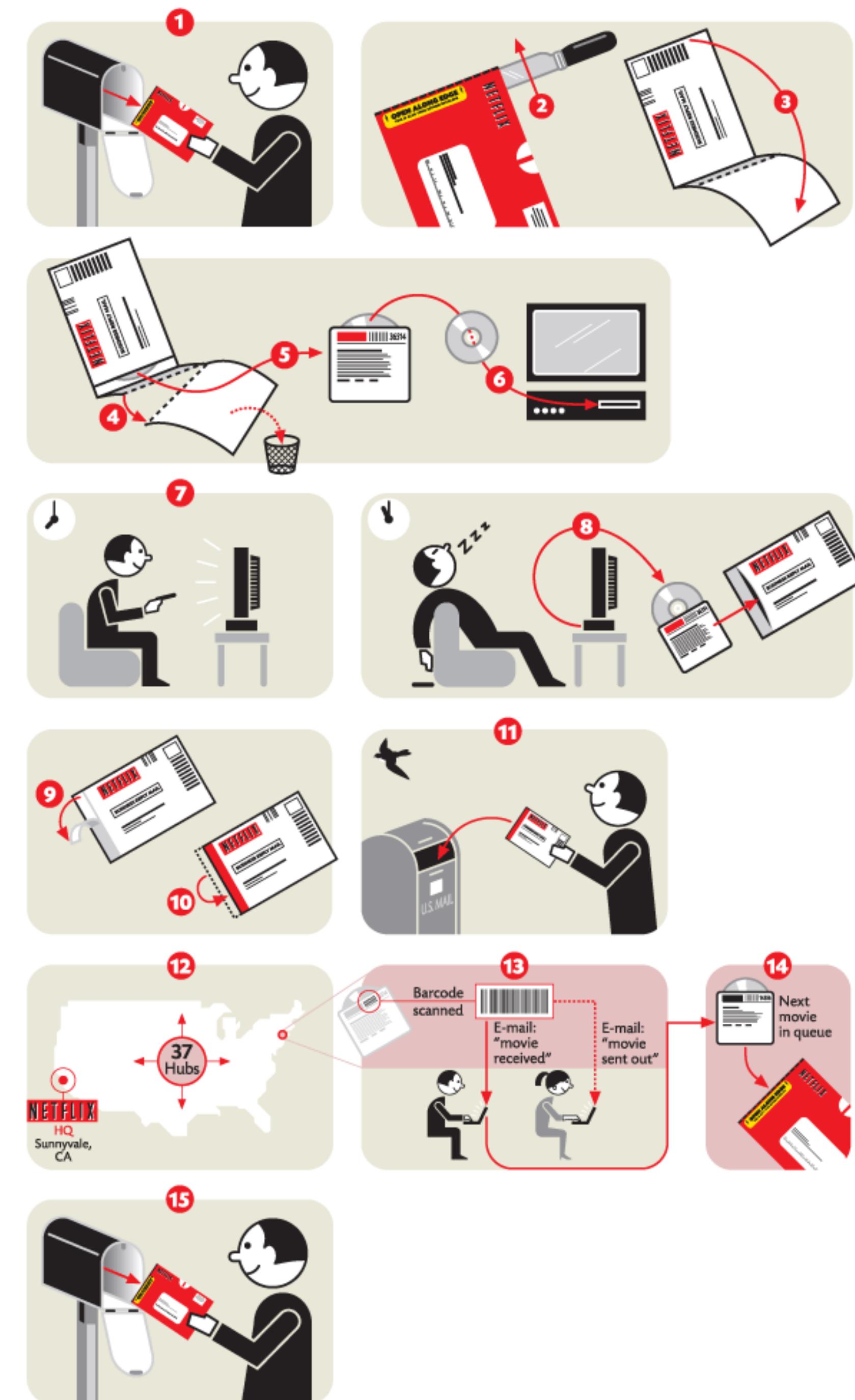
Years

MONSTROUS COSTS

Total House and Senate campaign expenditures,
in millions



HOW NETFLIX WORKS

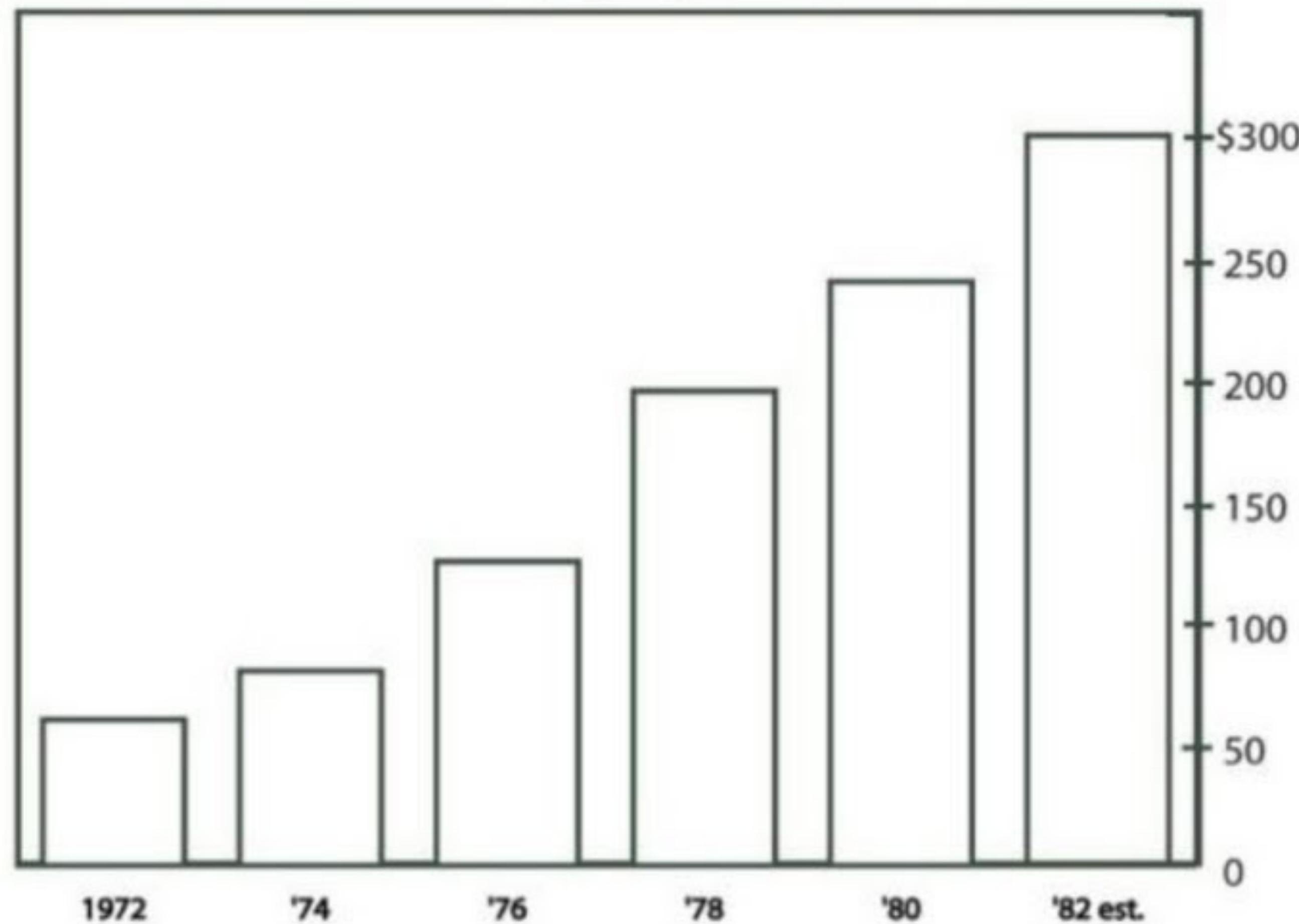


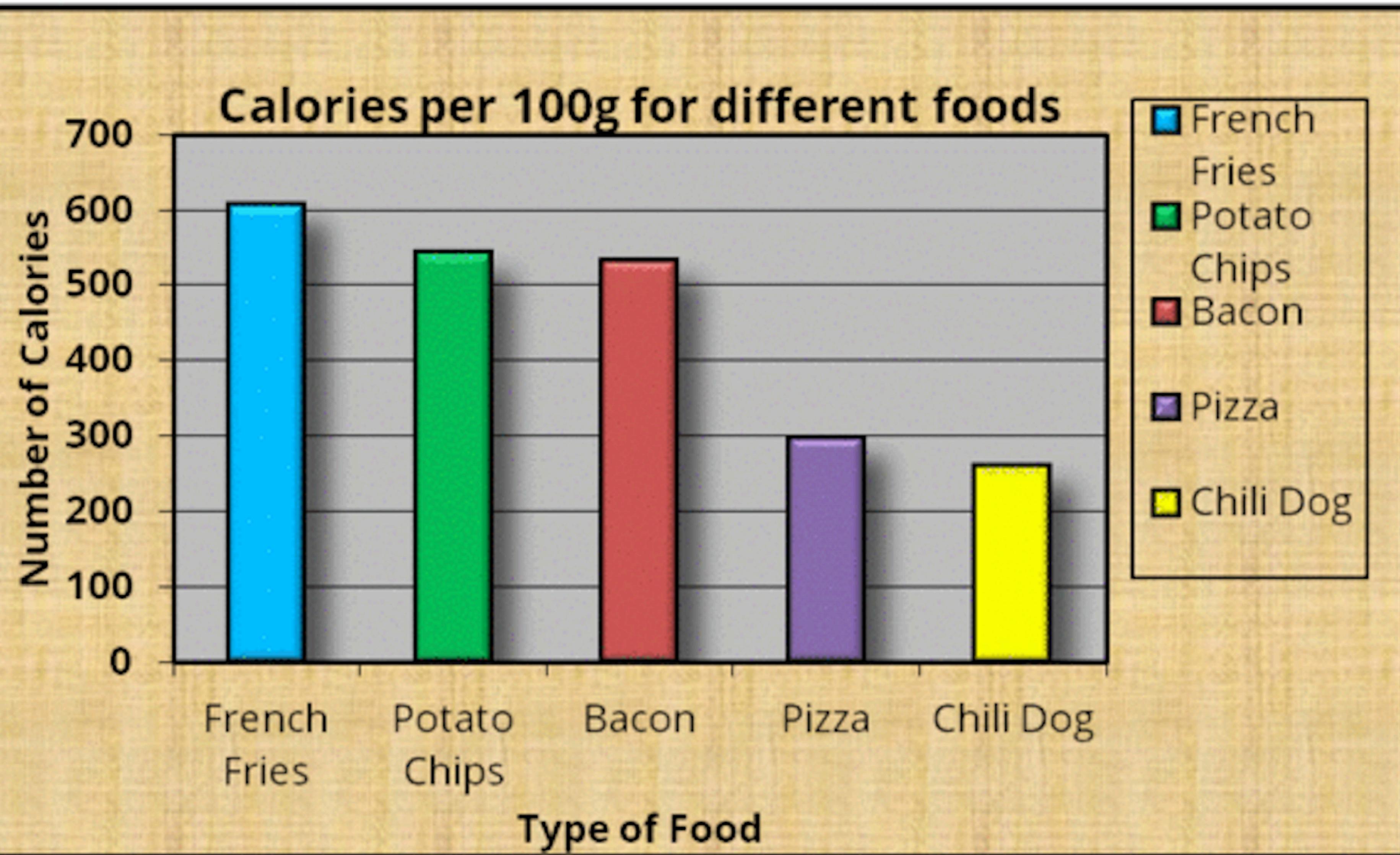
MONSTROUS COSTS

Total House and Senate campaign expenditures,
in millions

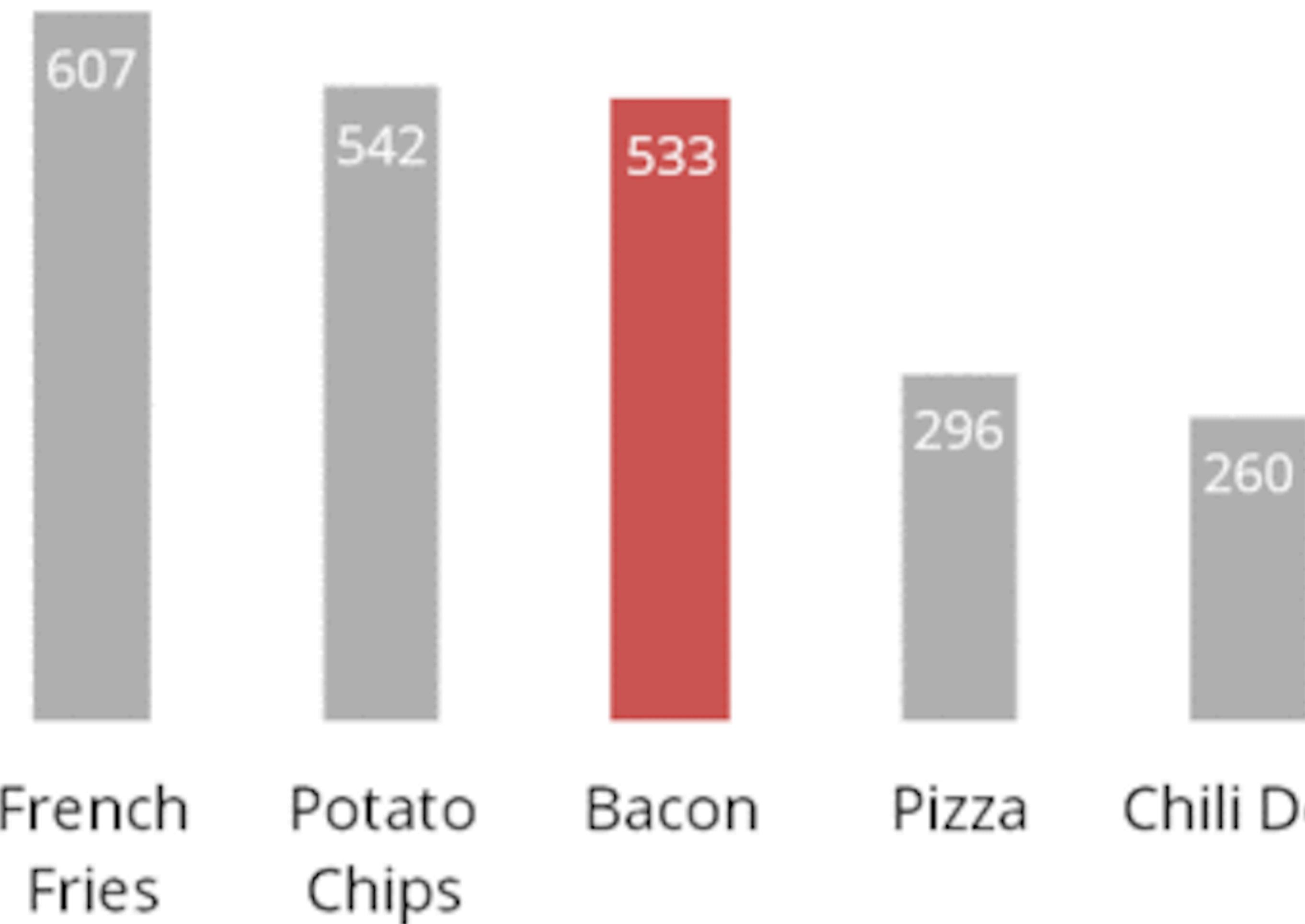


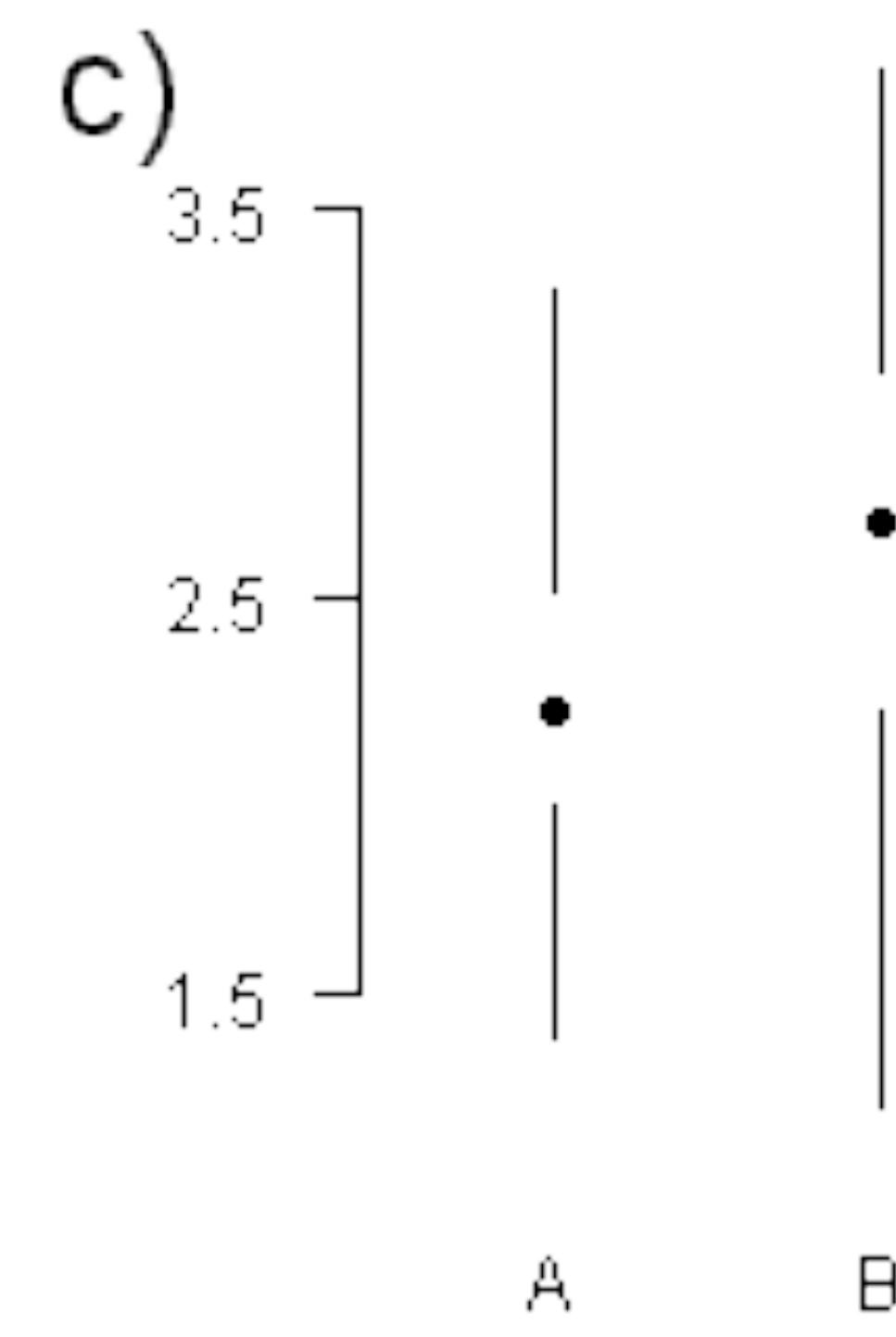
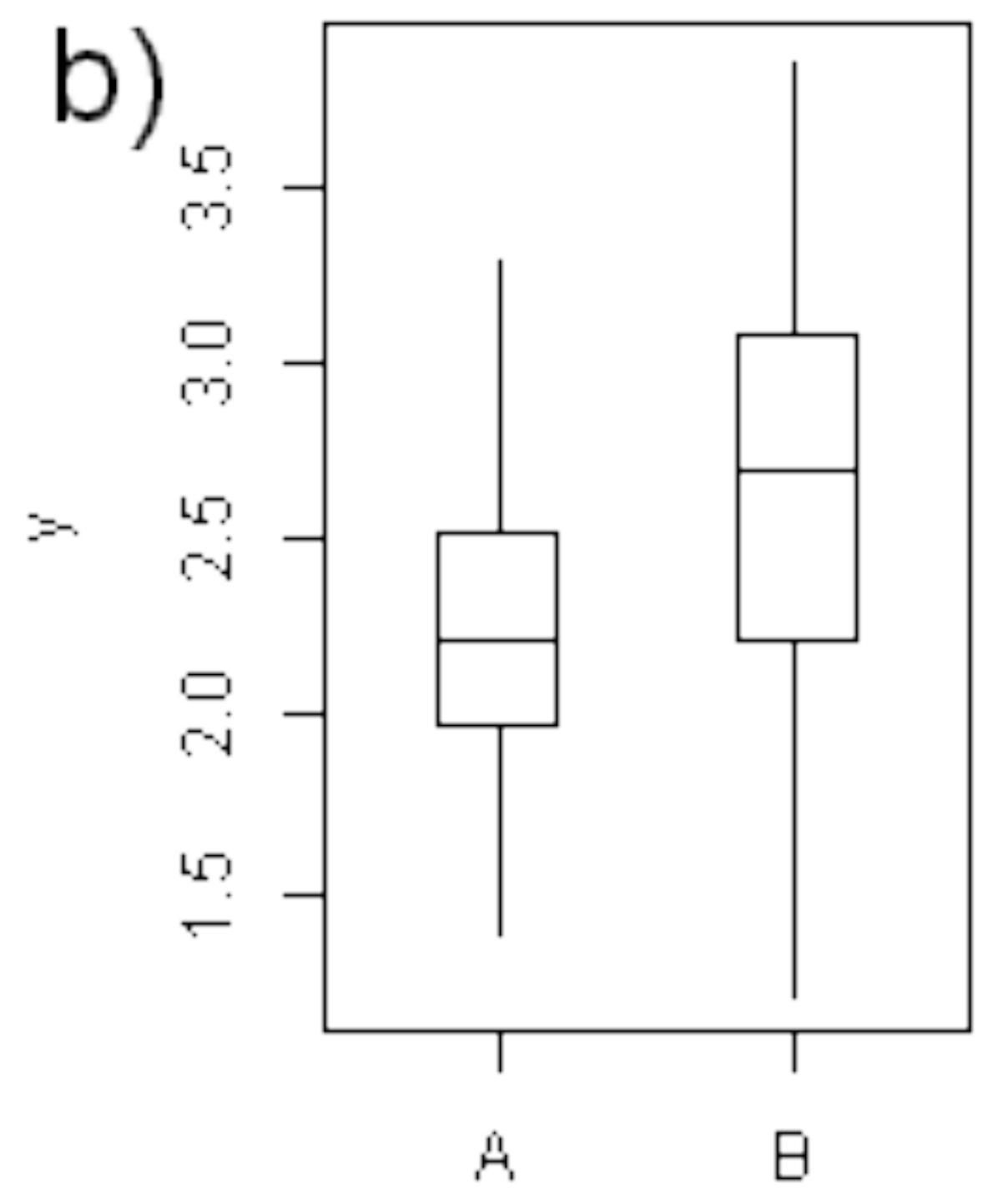
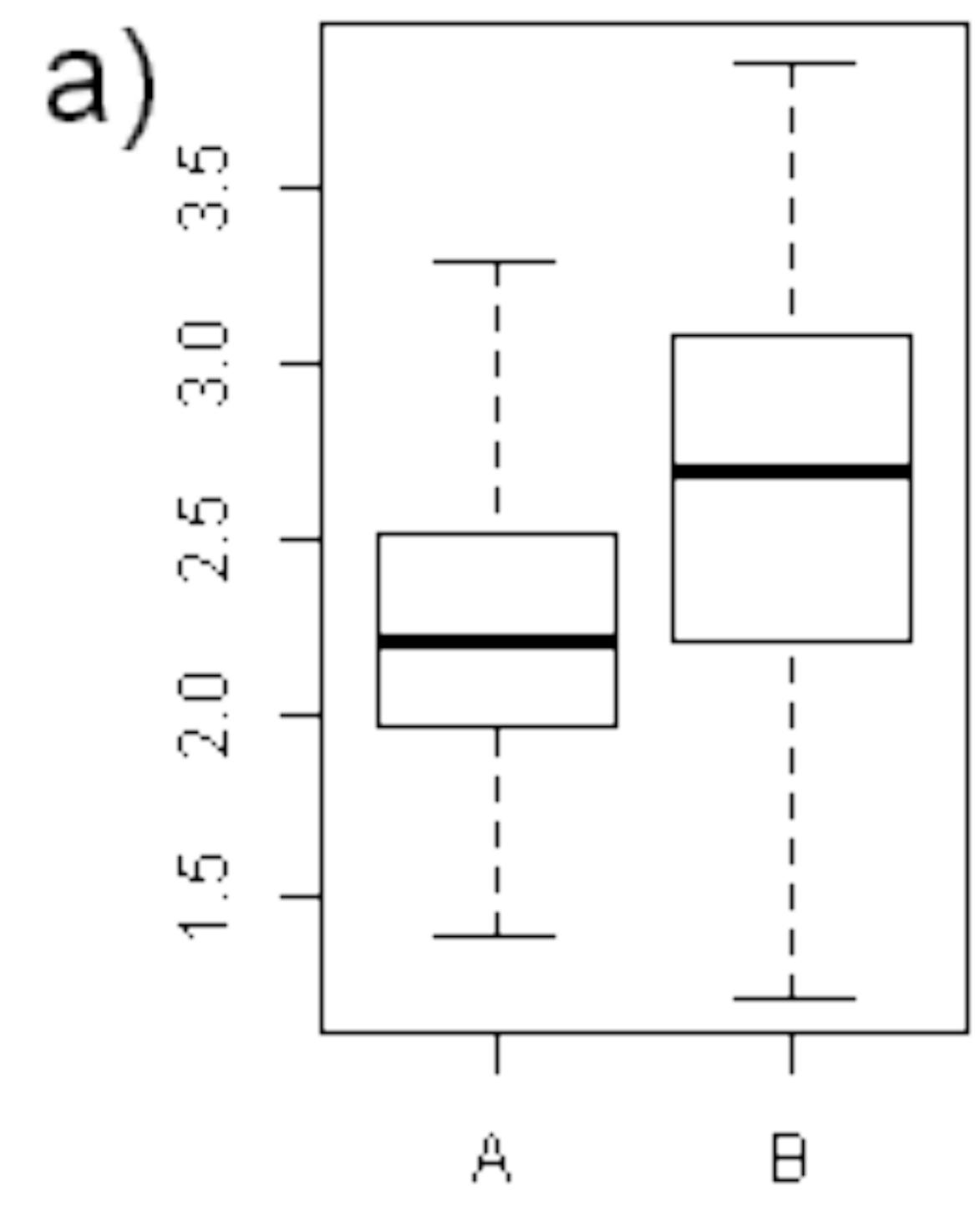
MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions

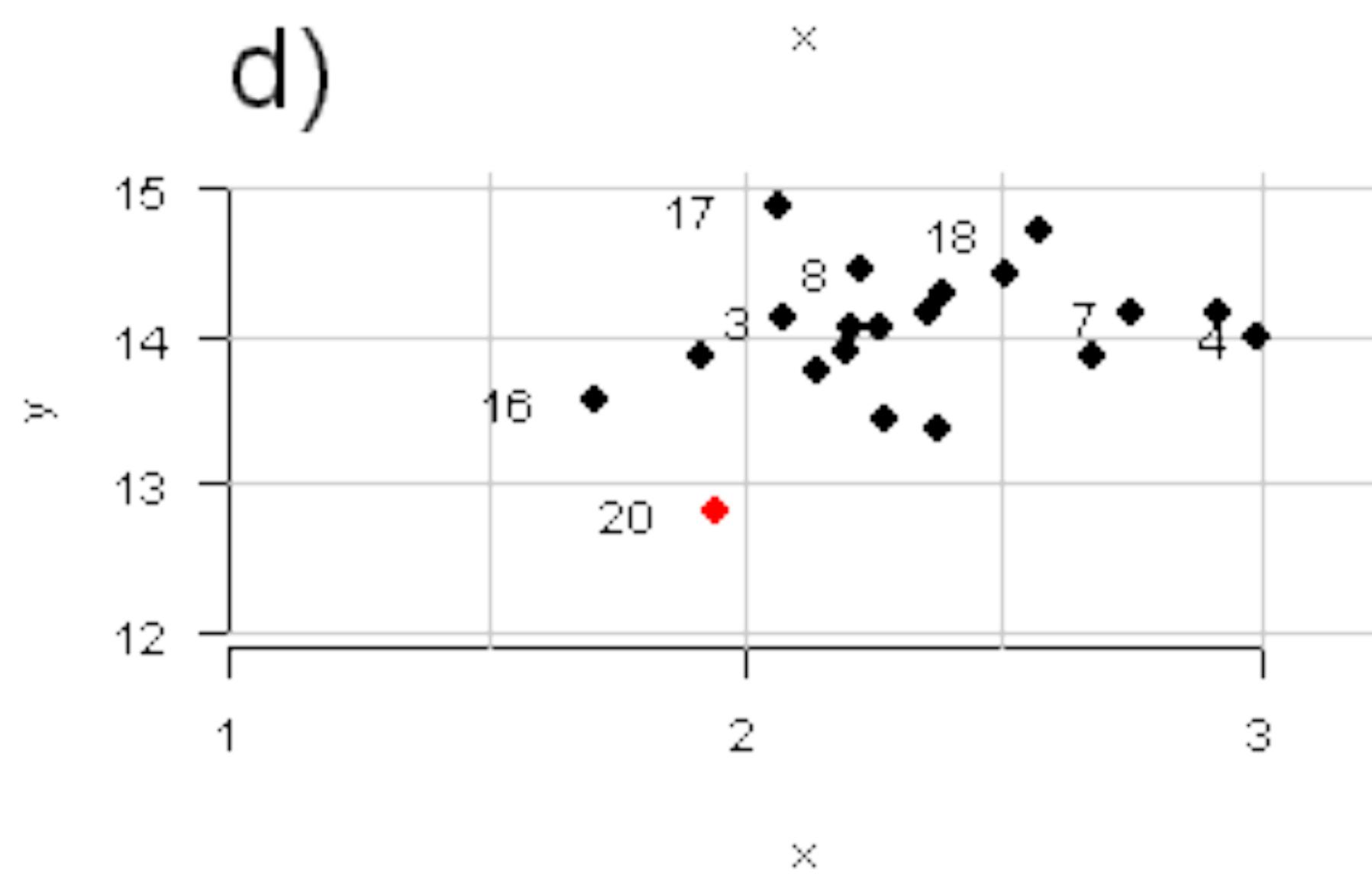
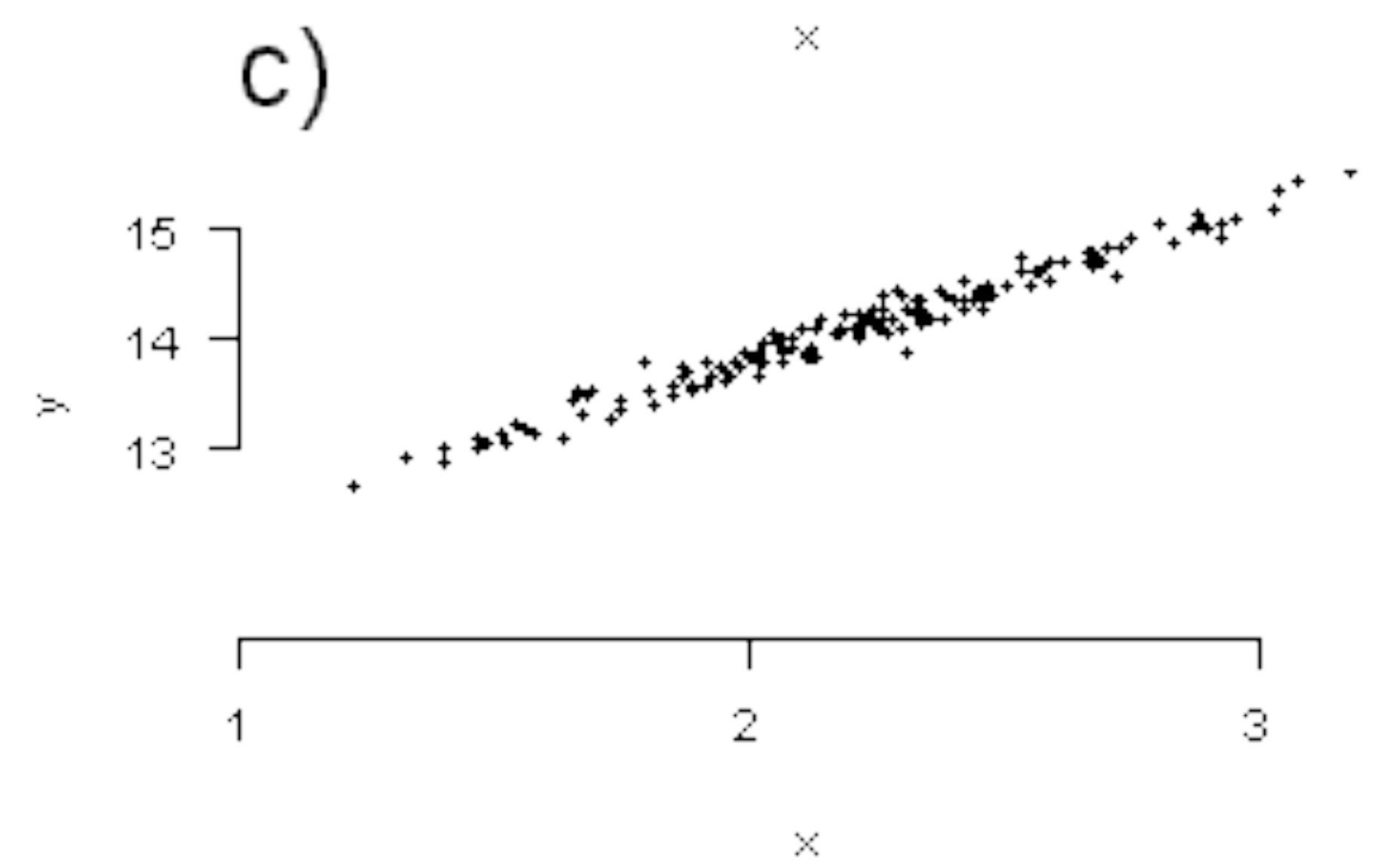
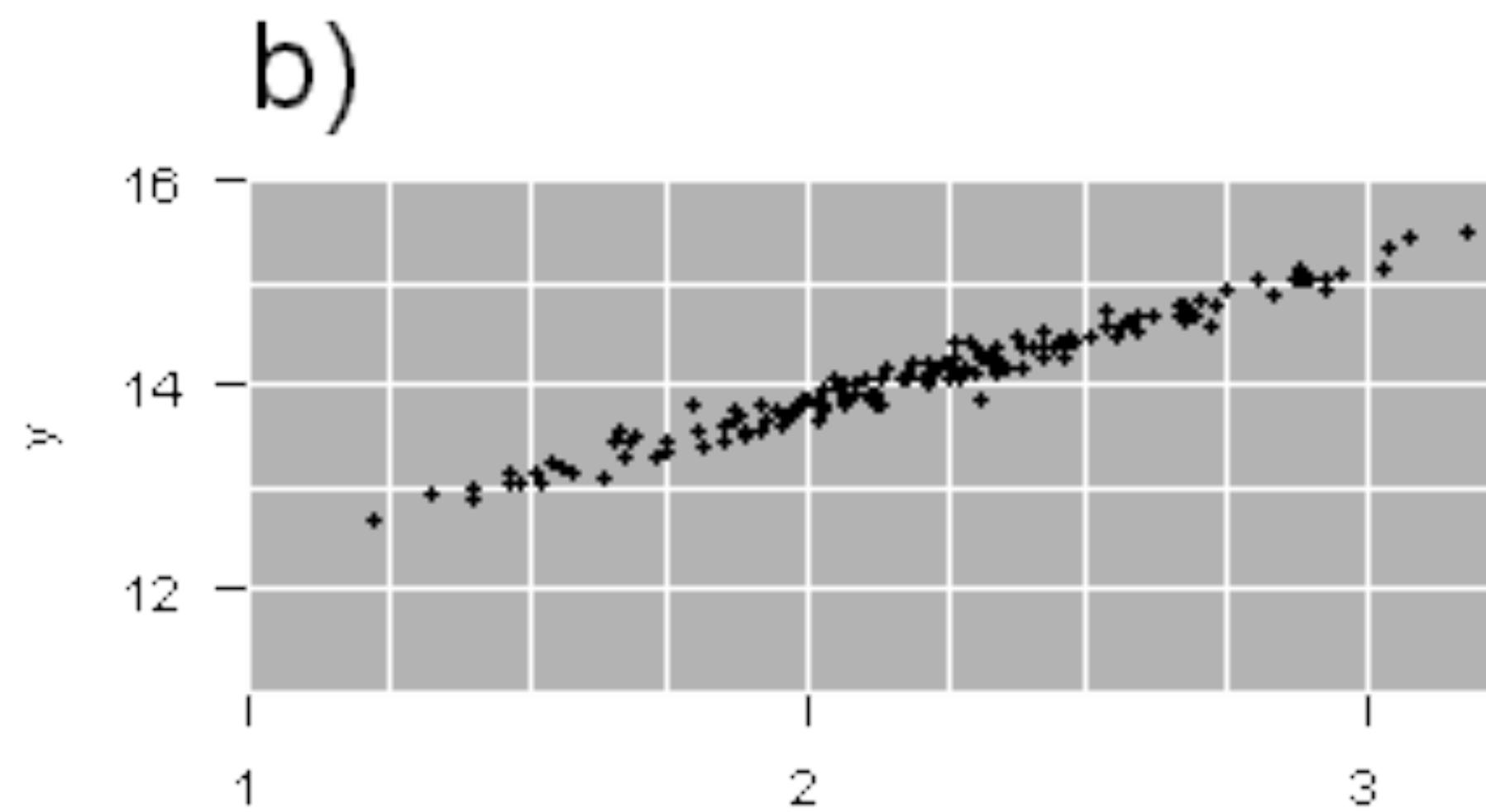
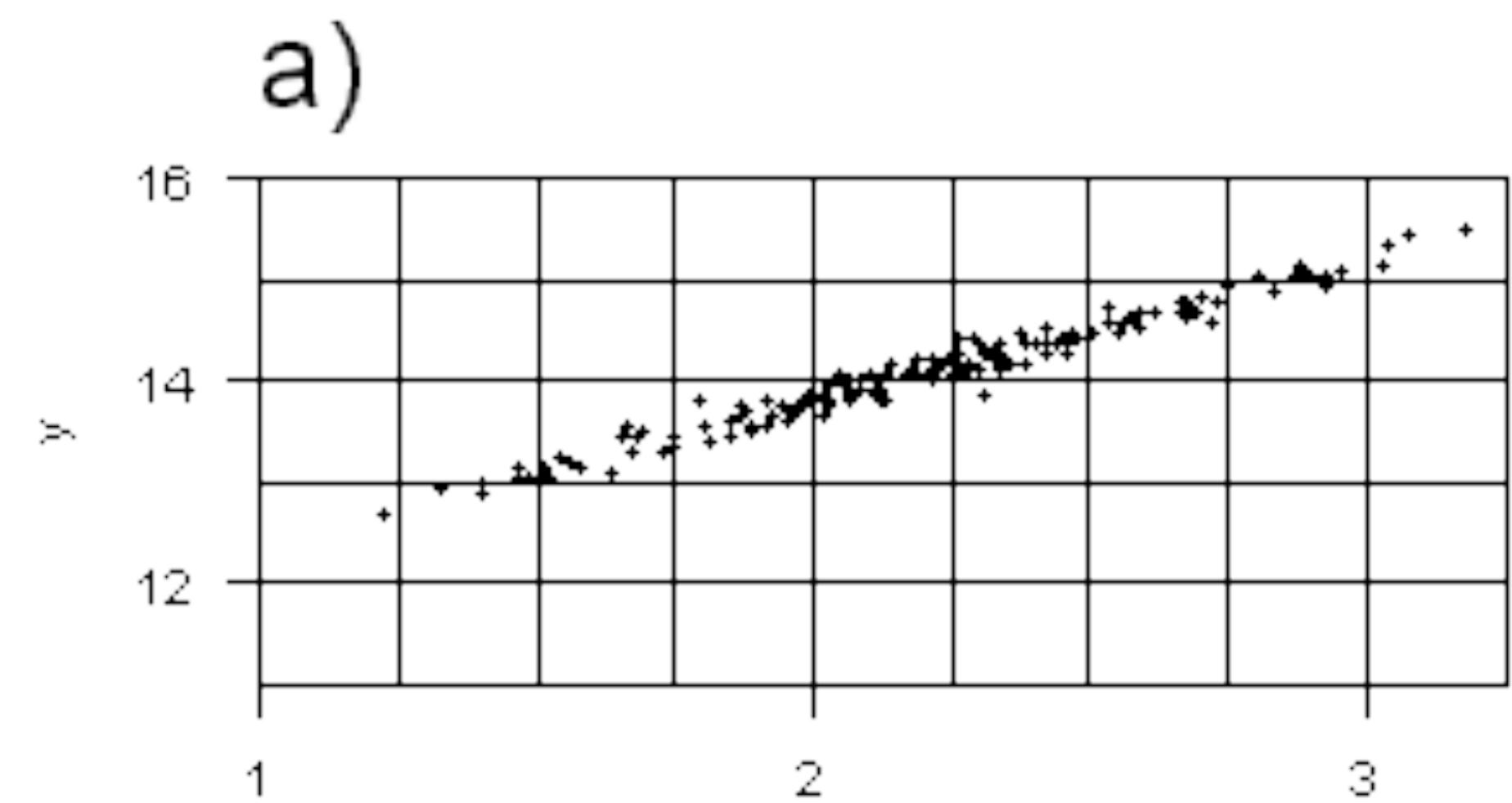




Calories per 100g

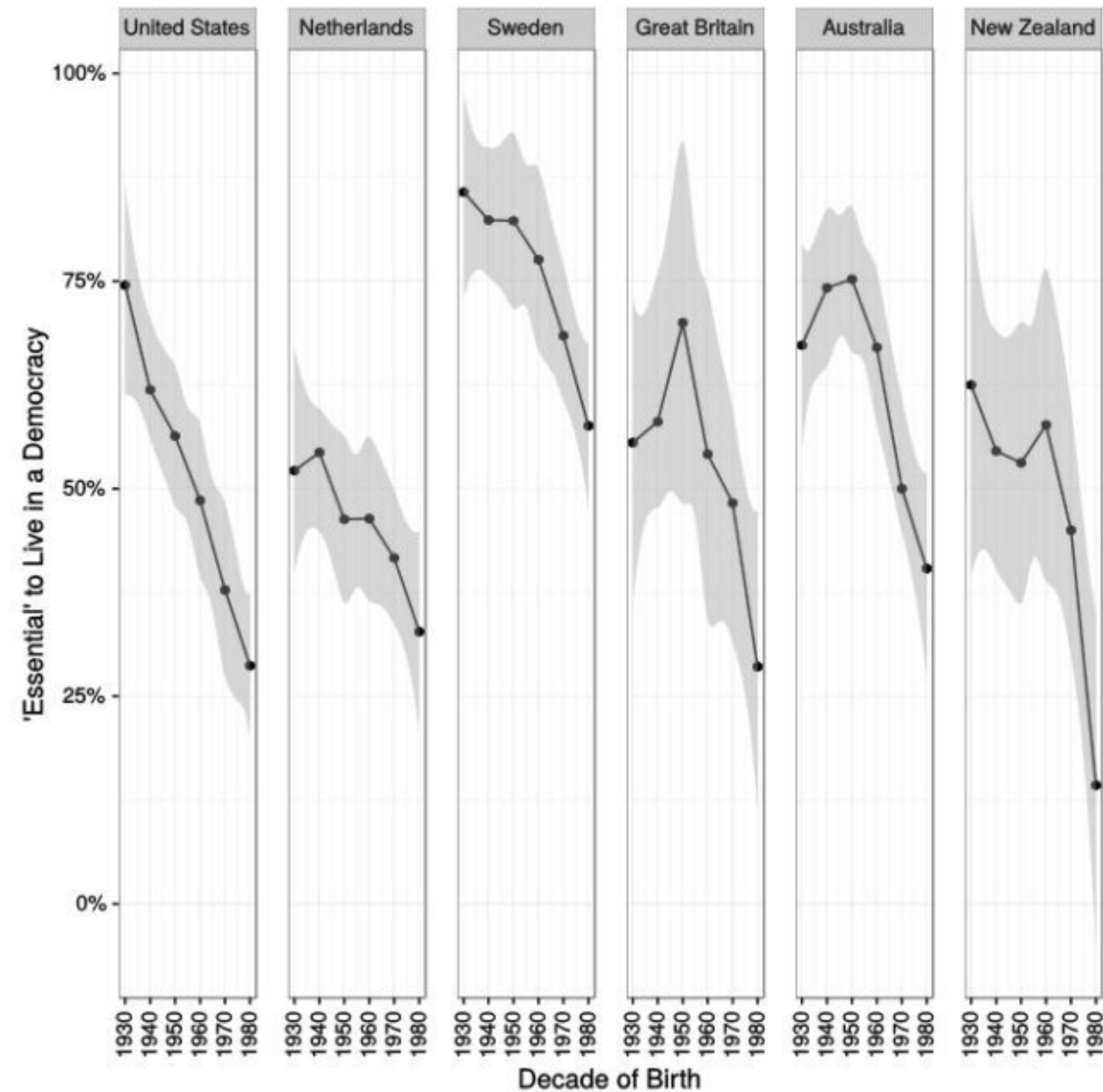




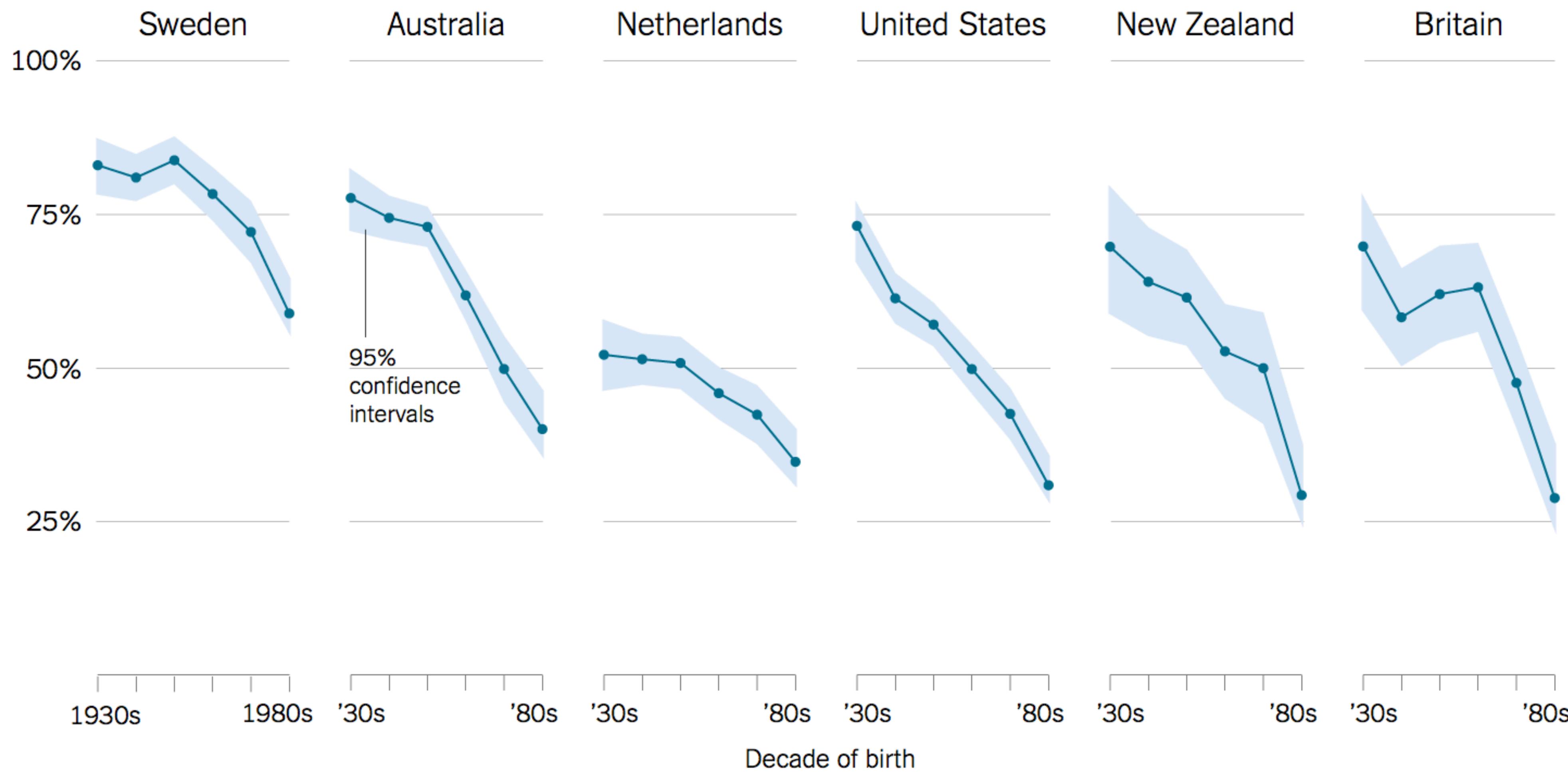


DATA

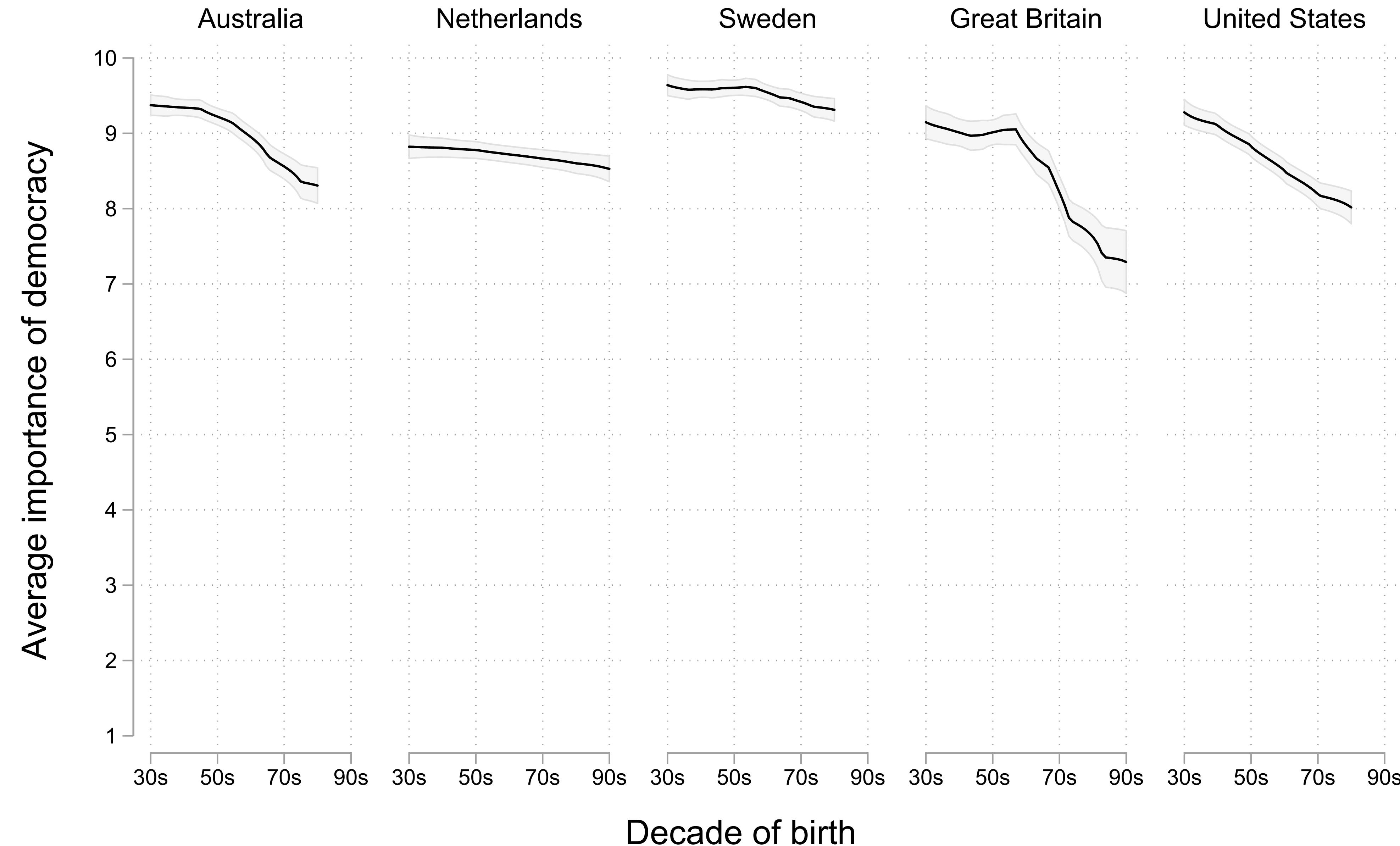
**JUNK-FREE
JUNK CHARTS**



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

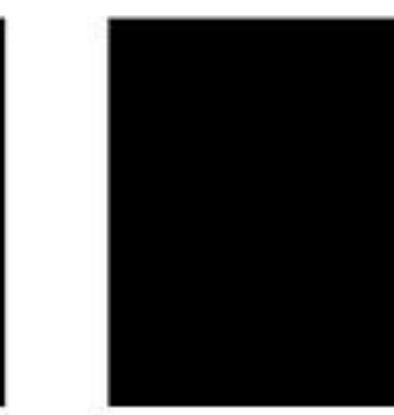
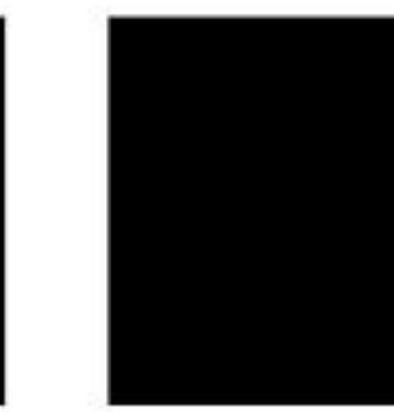
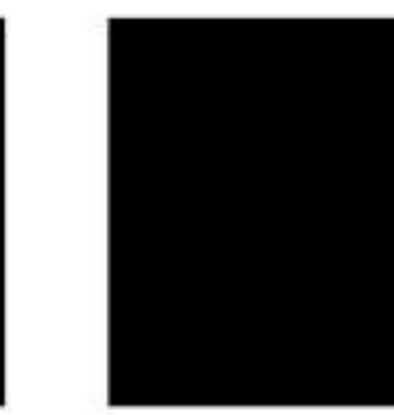


Graph by Erik Voeten, based on WVS 5

PERCEPTION

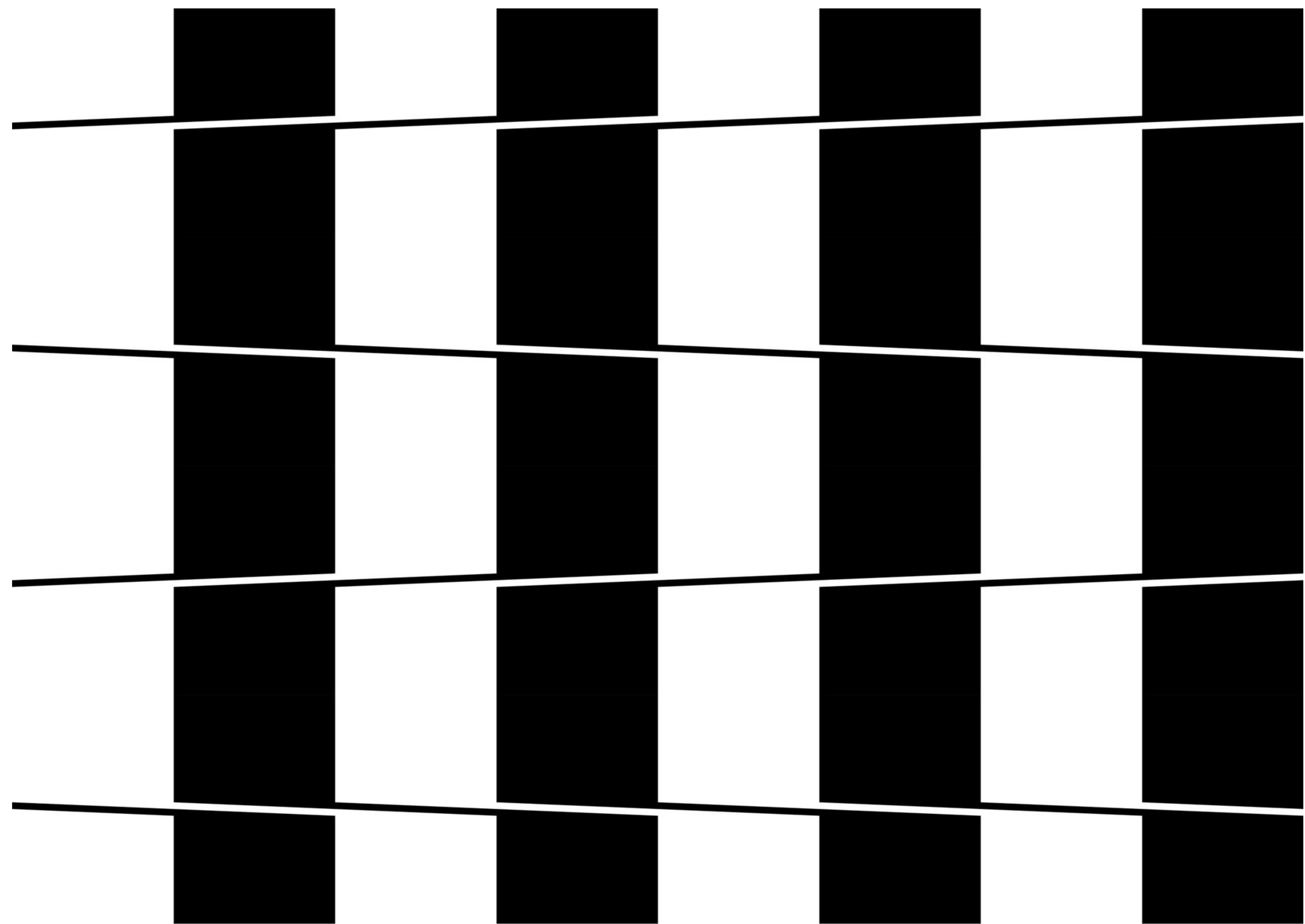
SIGHT AND INSIGHT

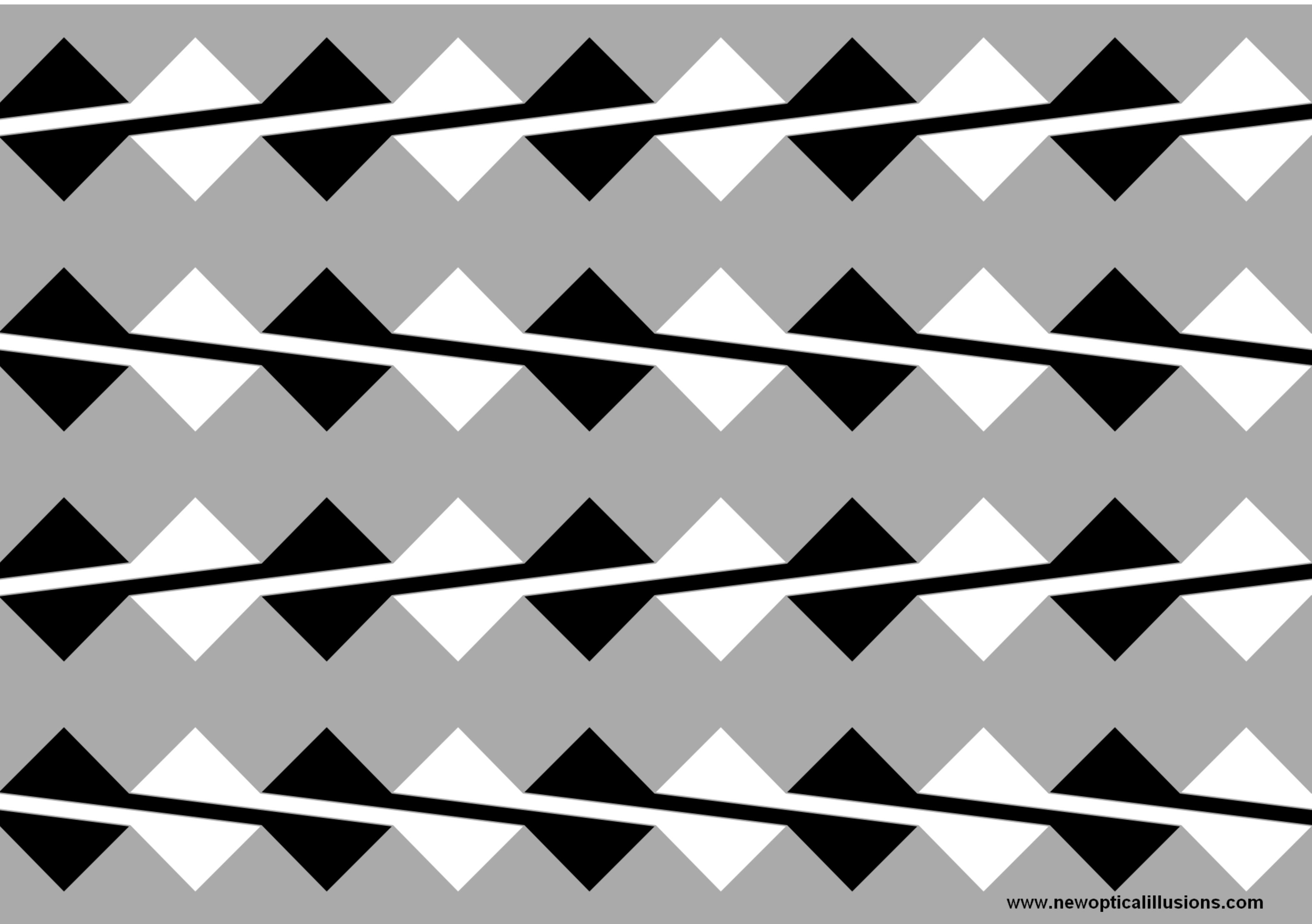
**Edges &
Contrasts**

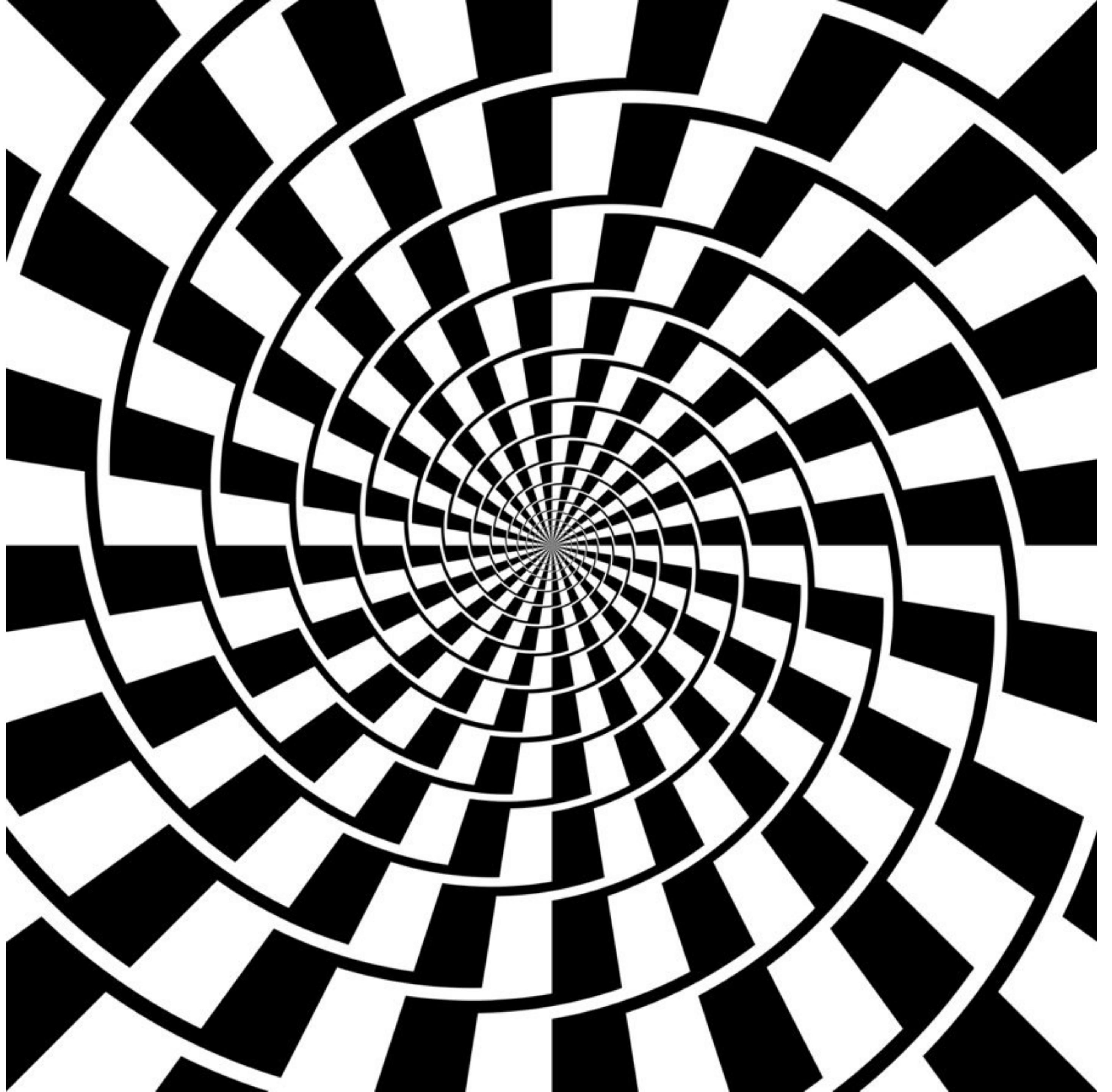


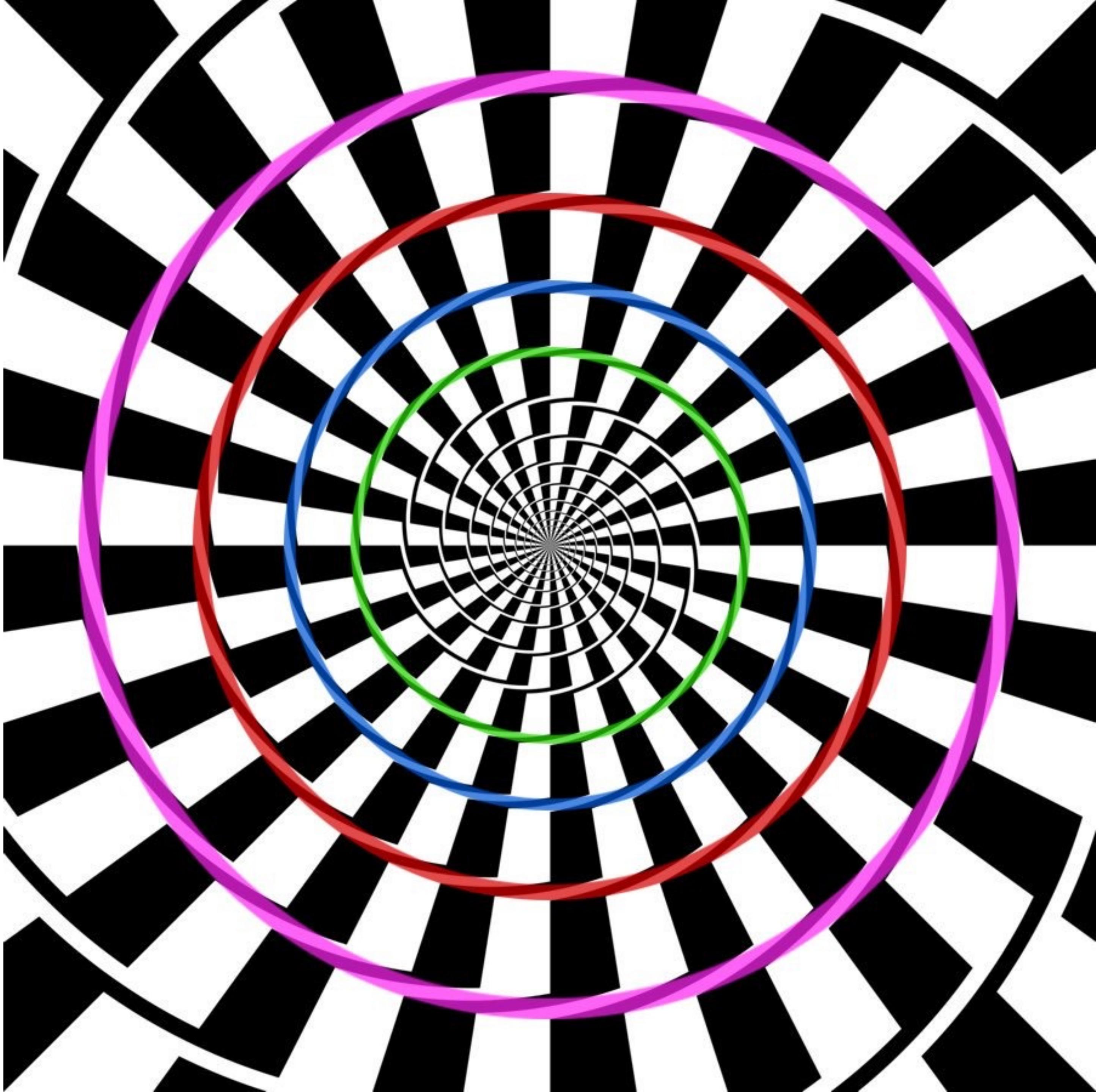
fin

© 2013

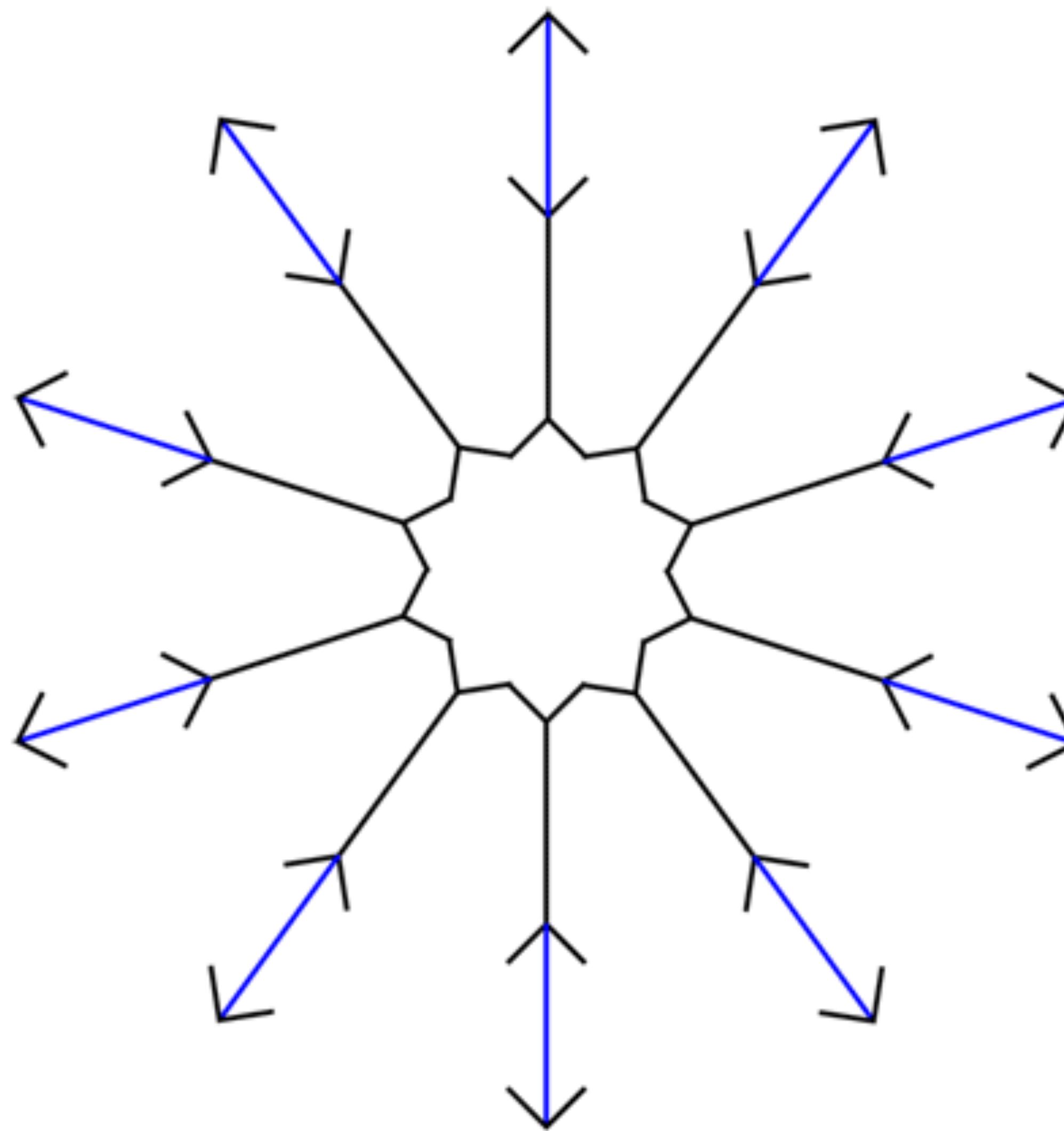


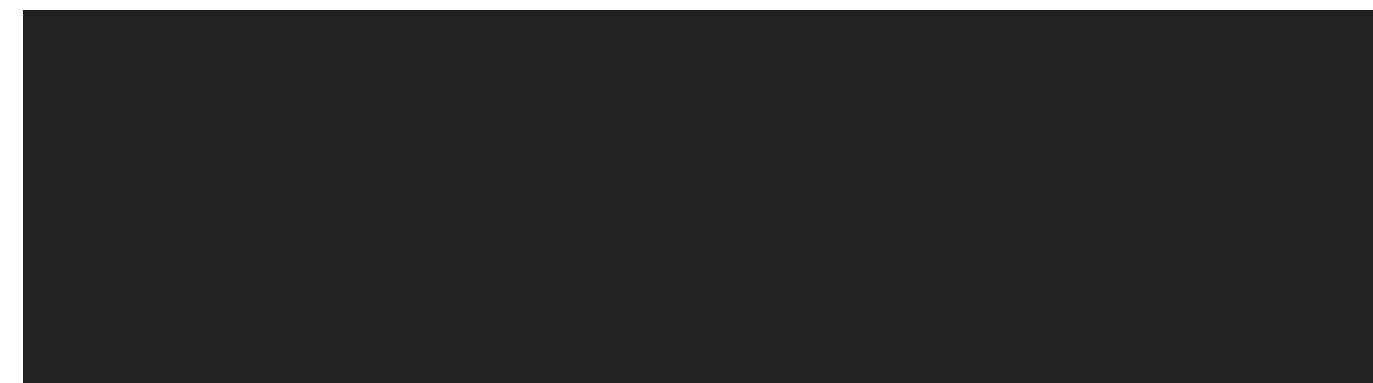




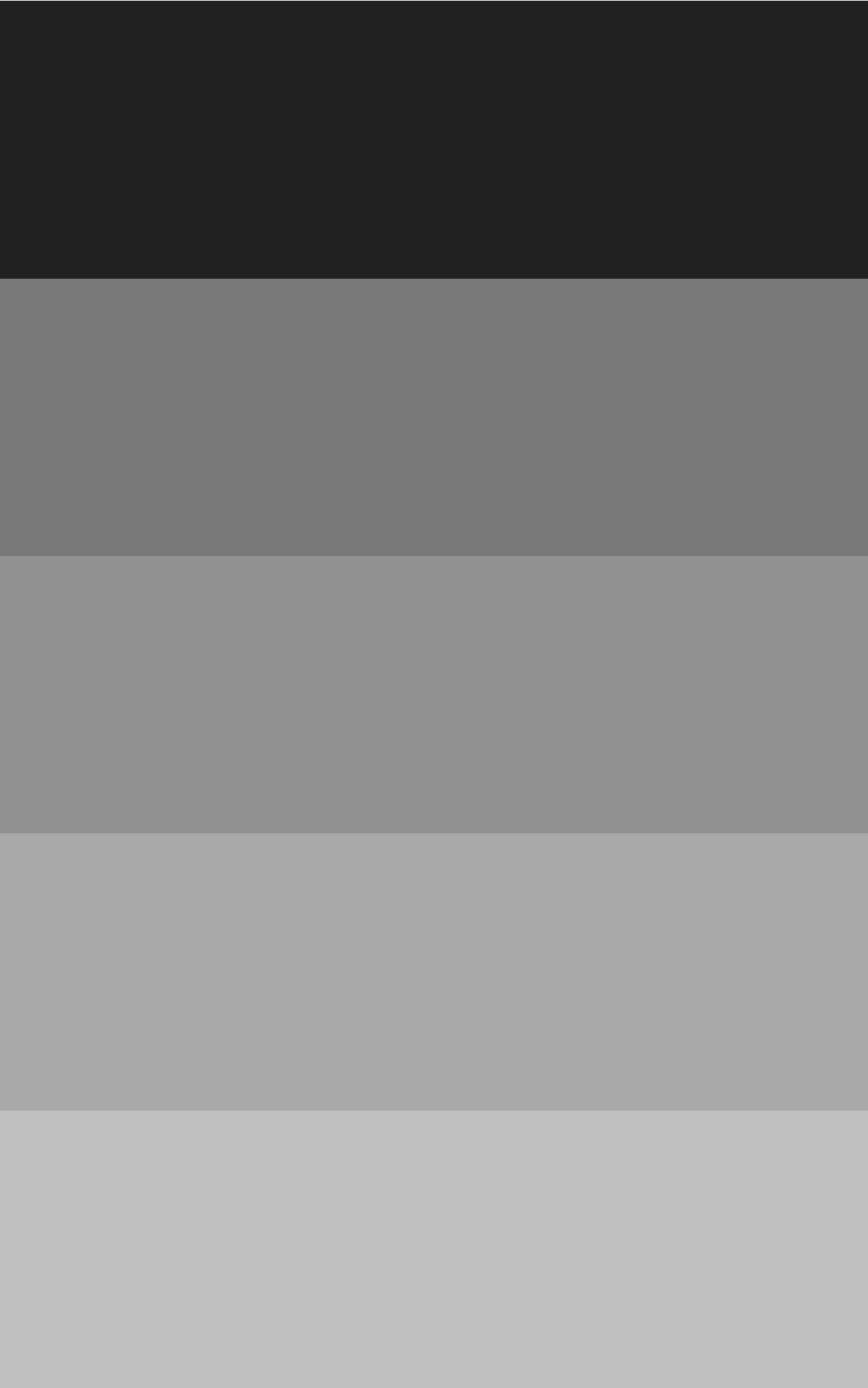


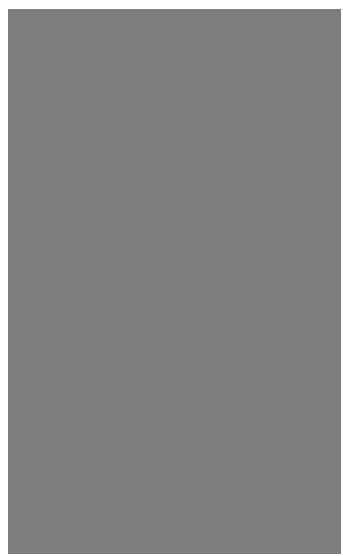
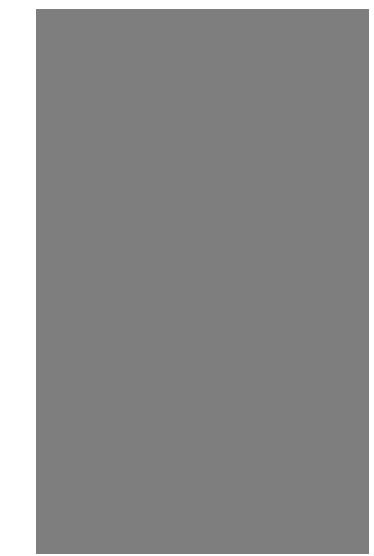
Sarcone's Dynamic Müller-Lyer Illusion

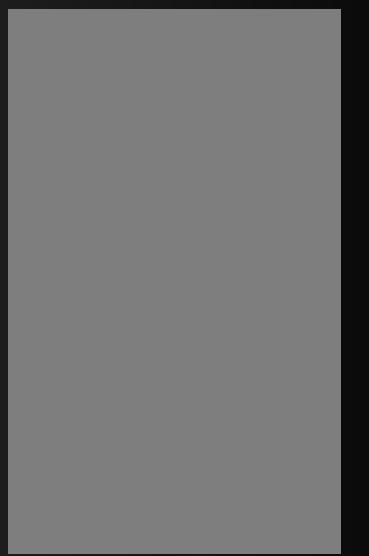
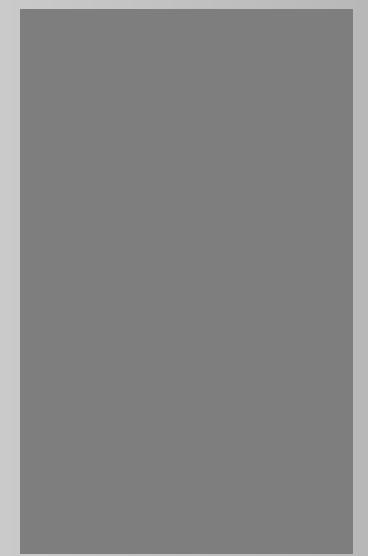
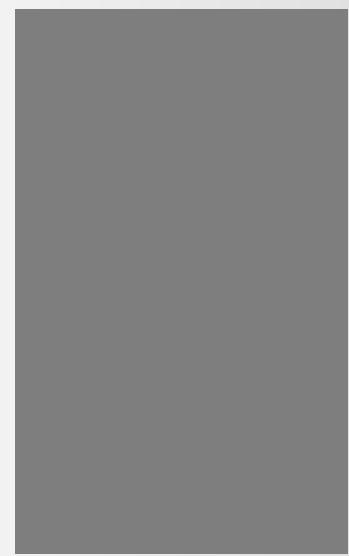


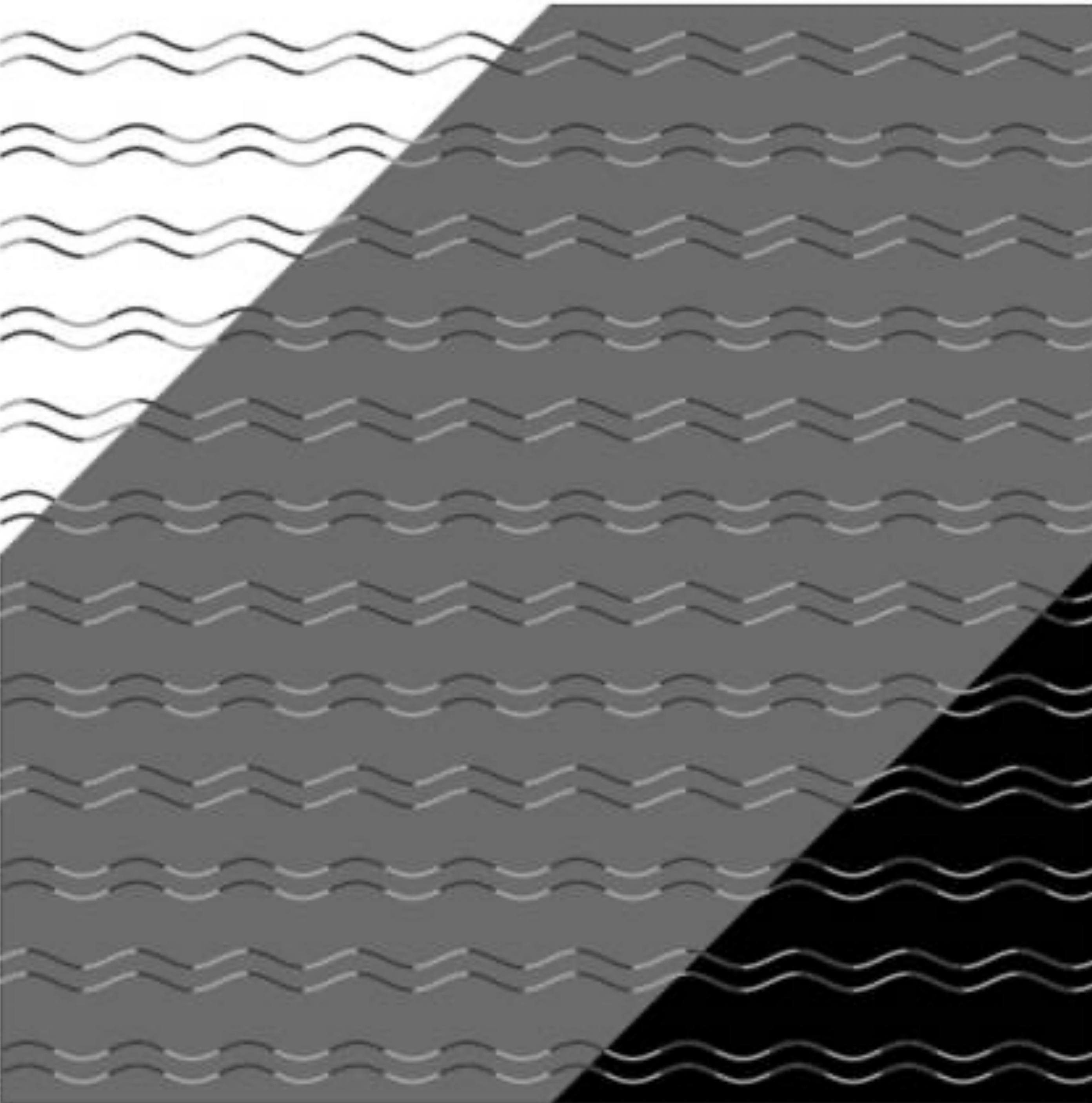


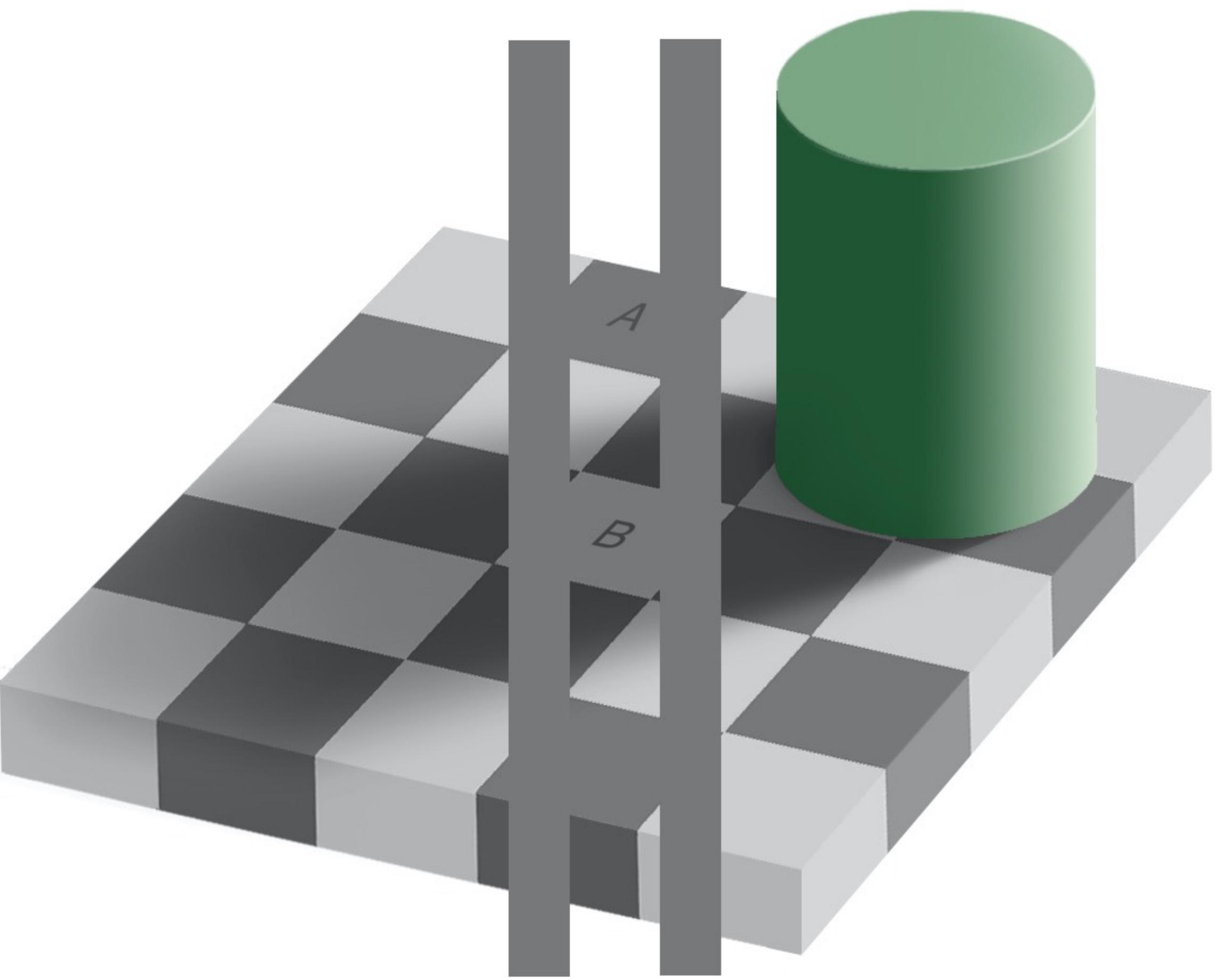
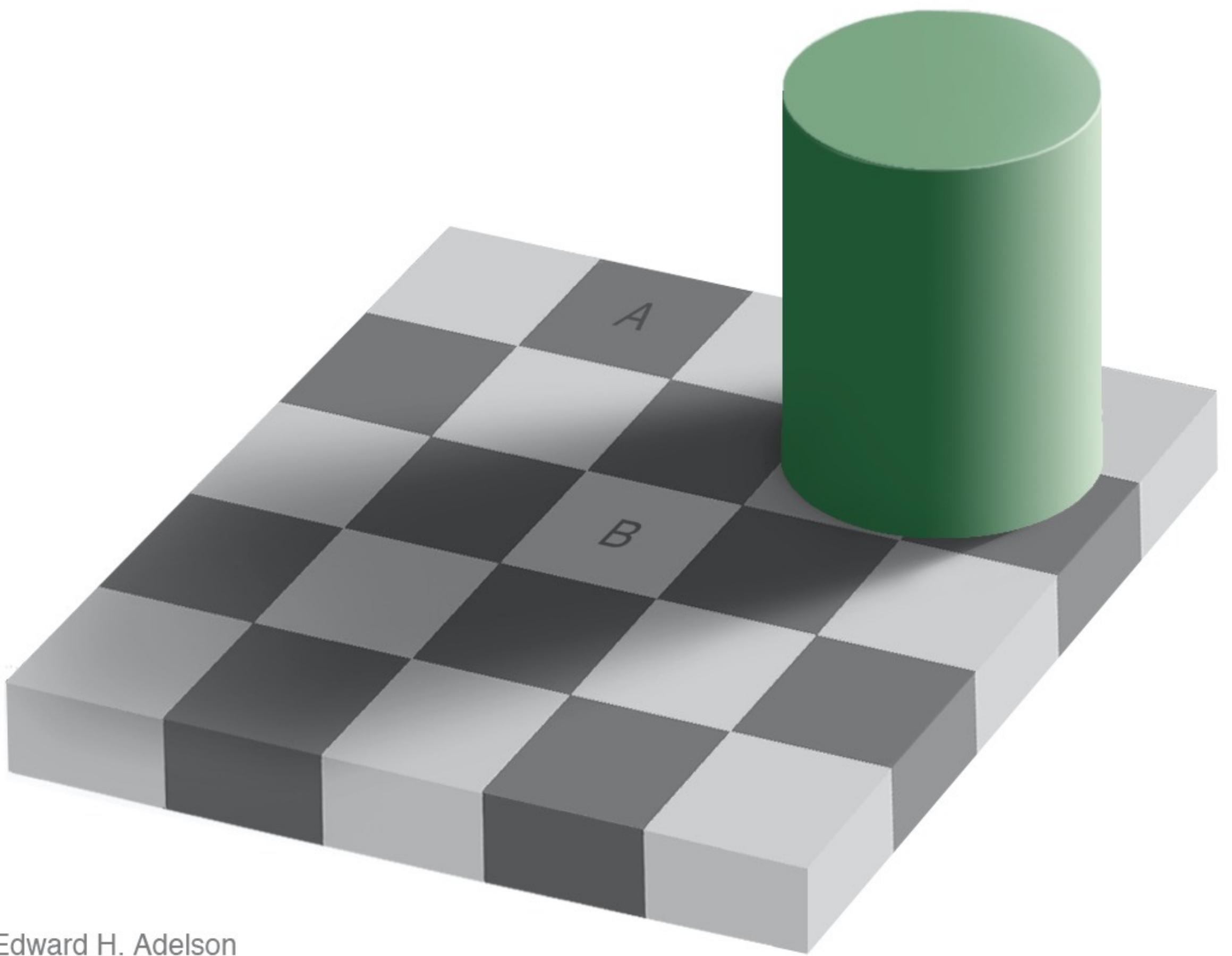




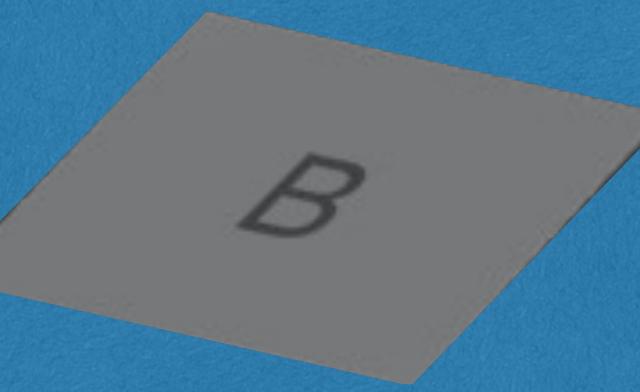
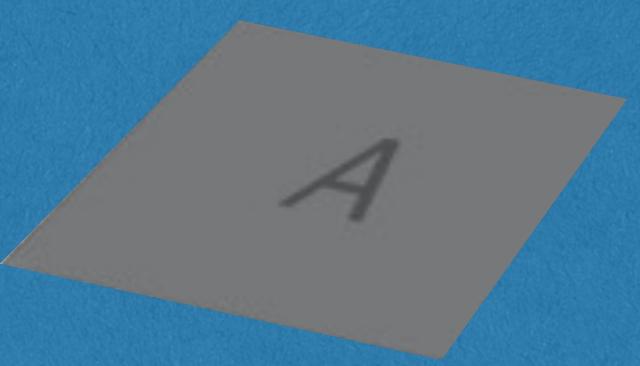


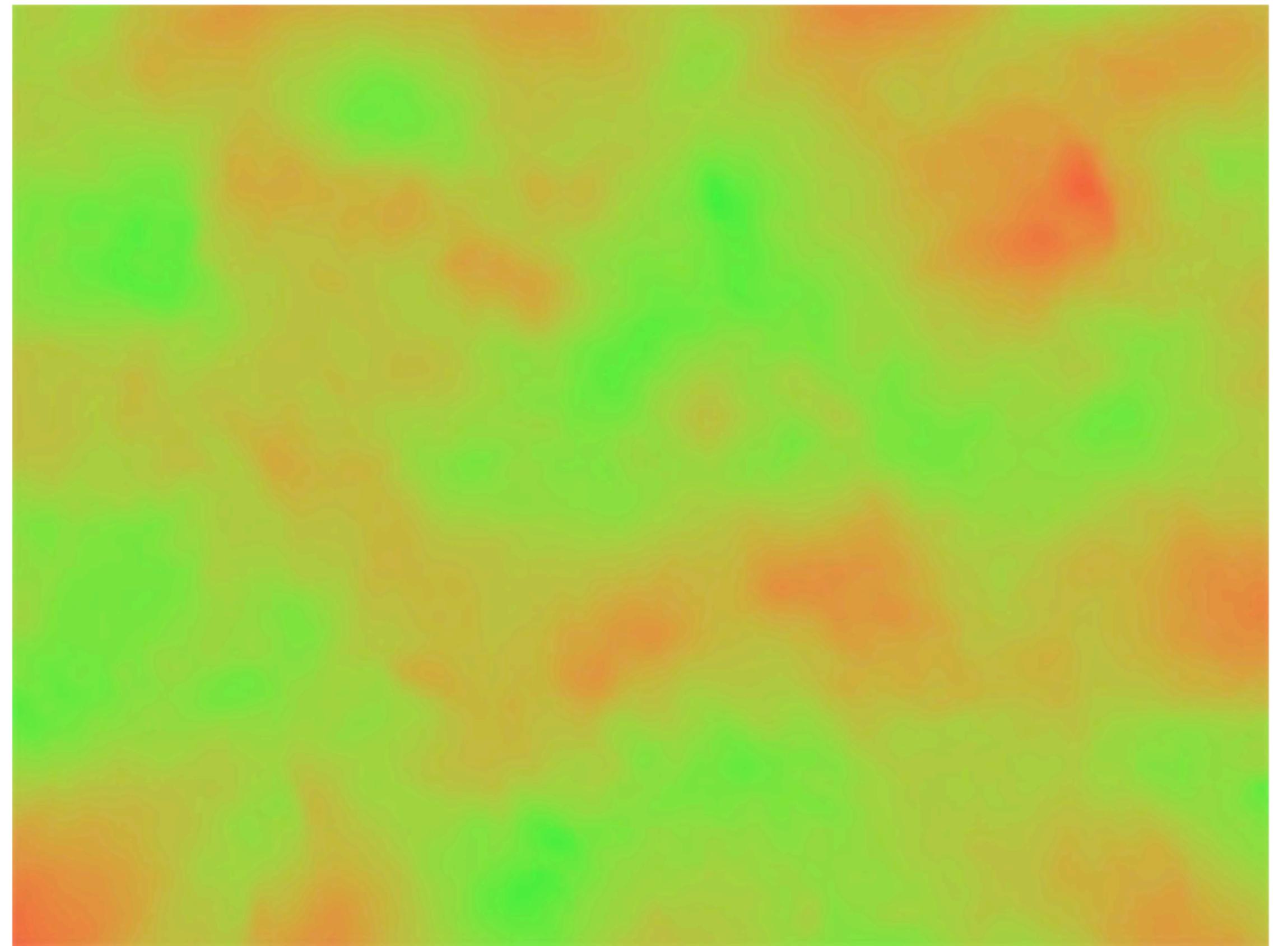
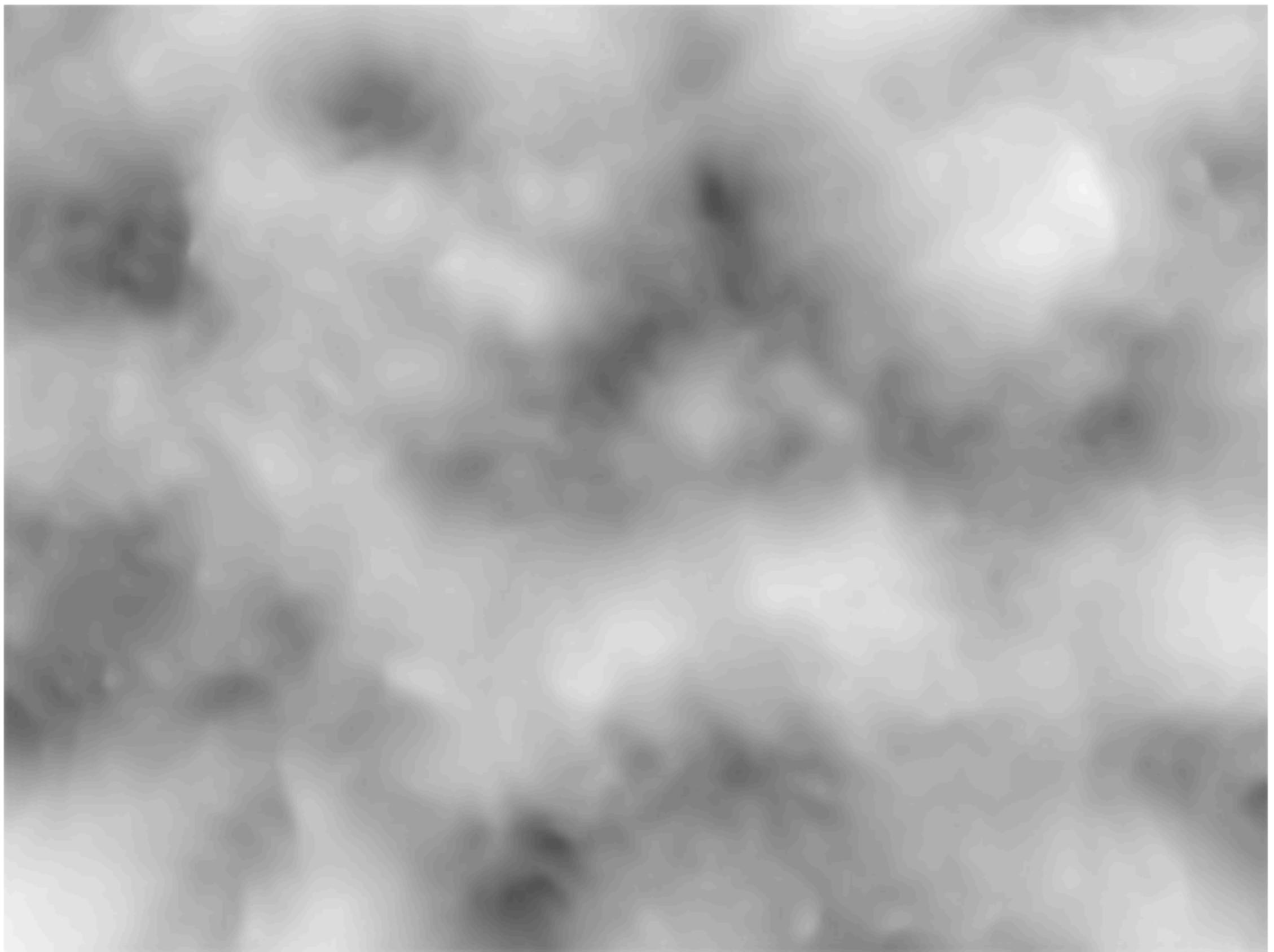


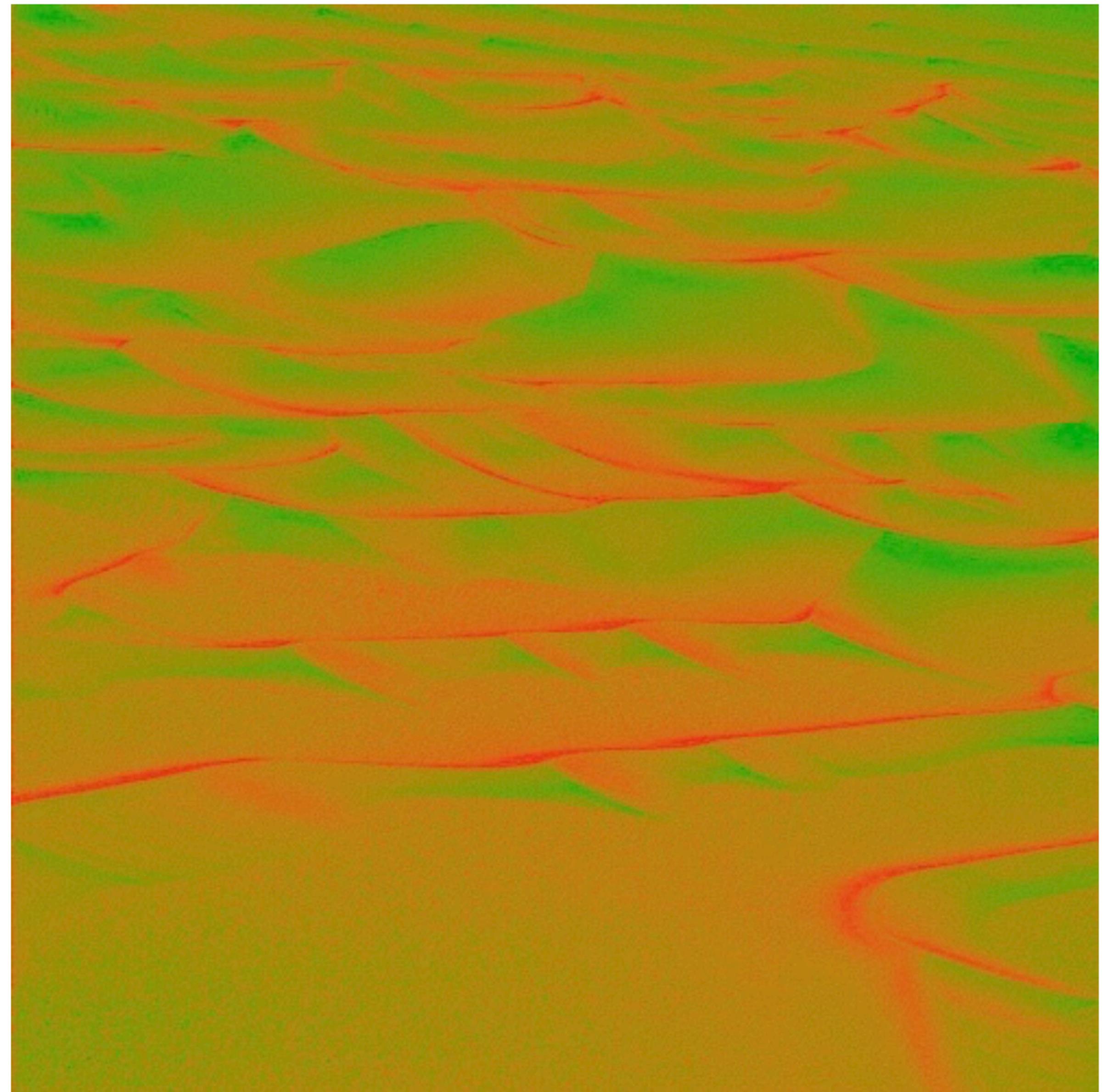


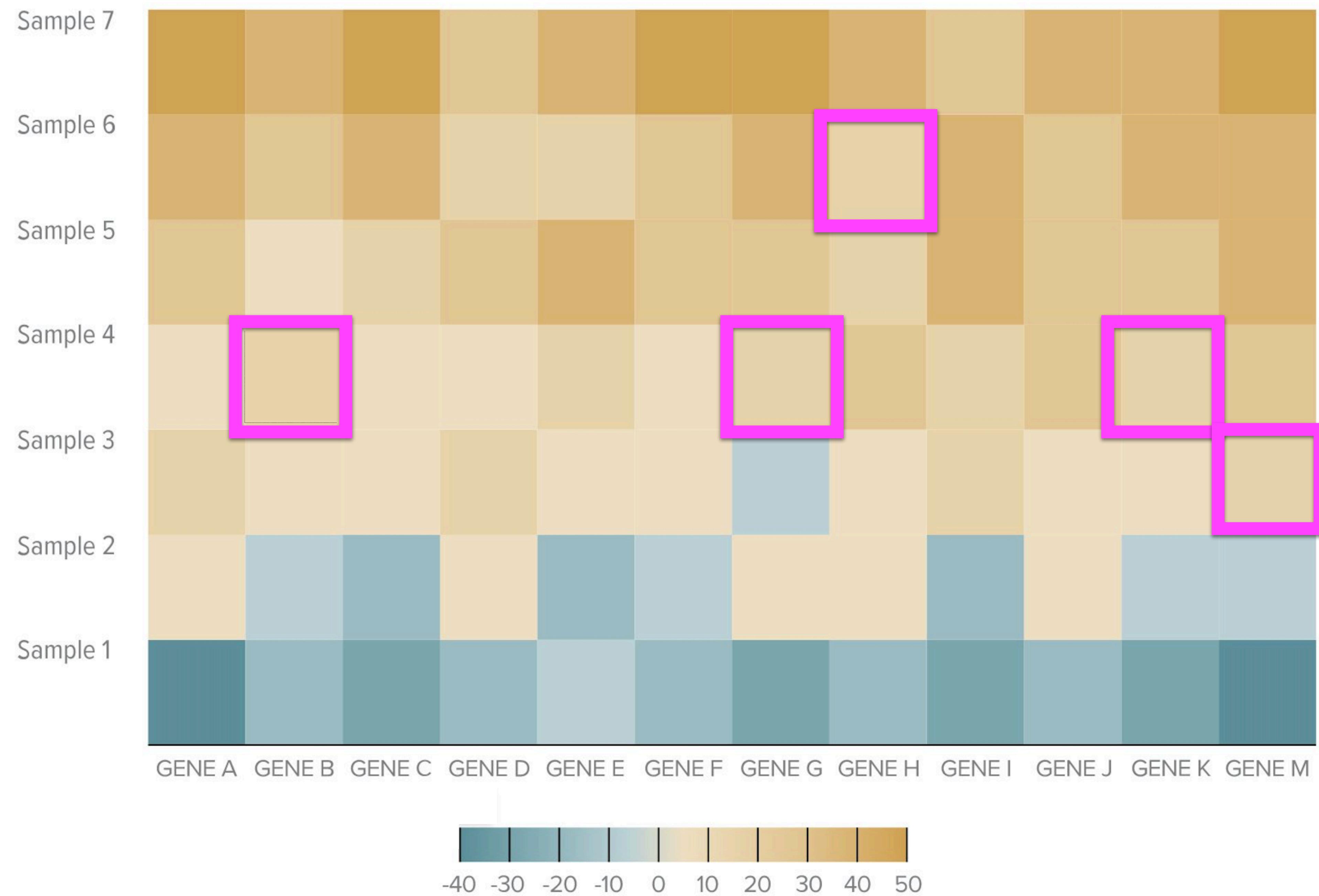


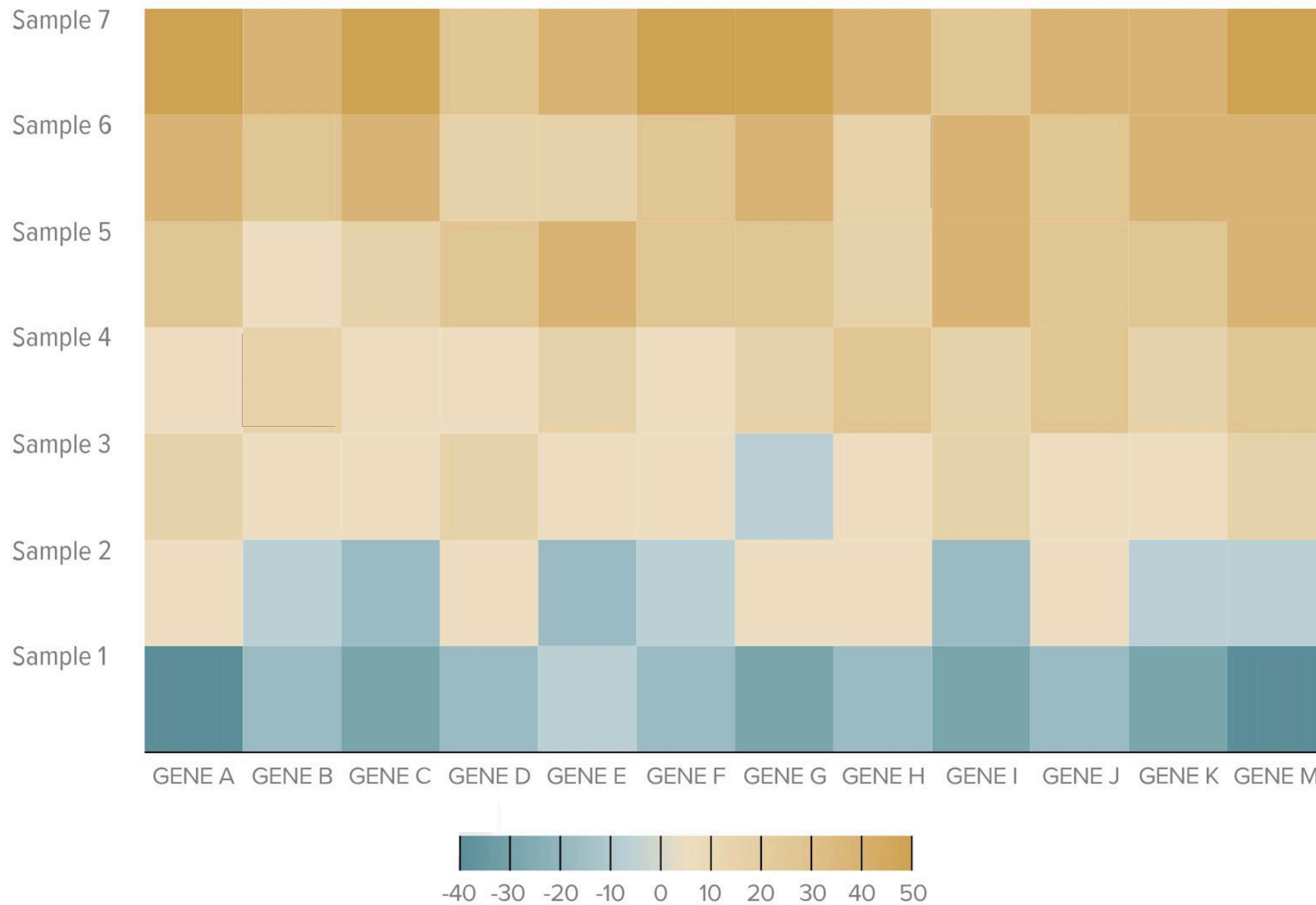
Edward H. Adelson







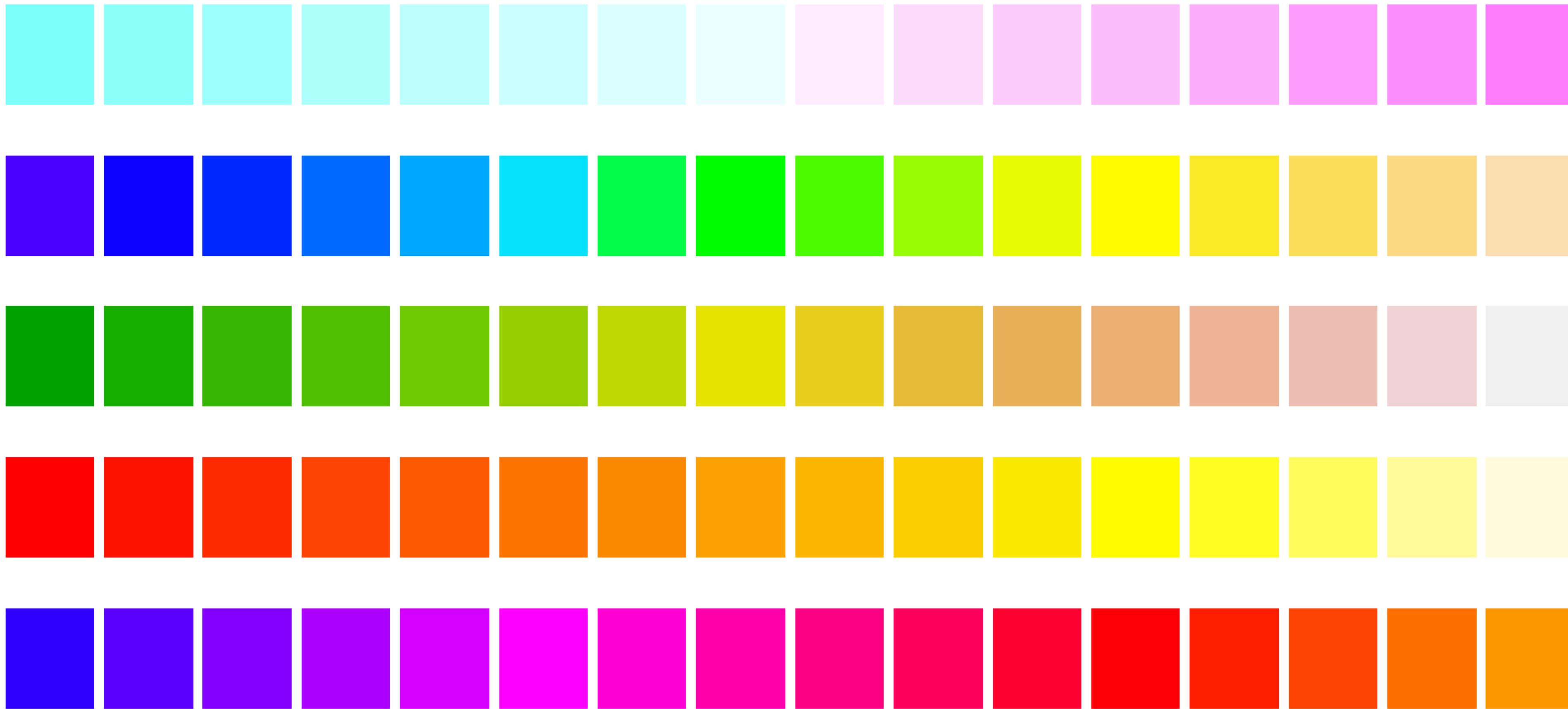




Color

+





viridis



magma



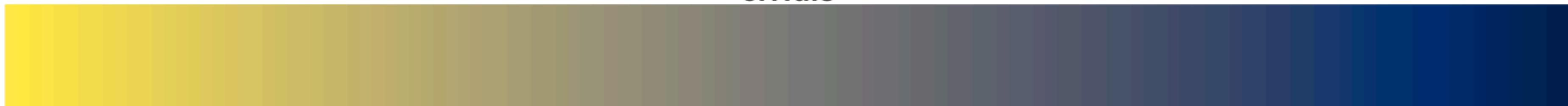
plasma

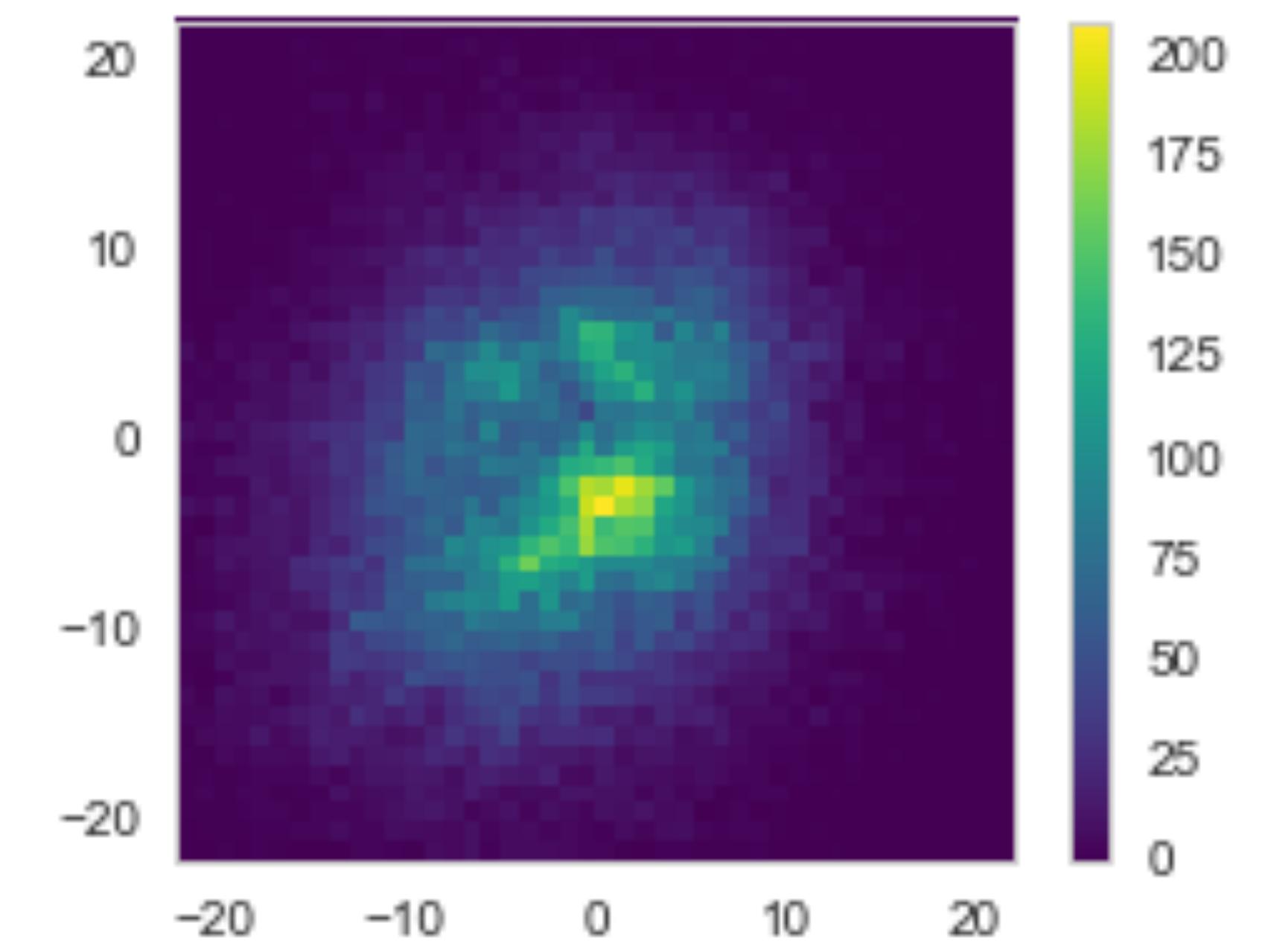
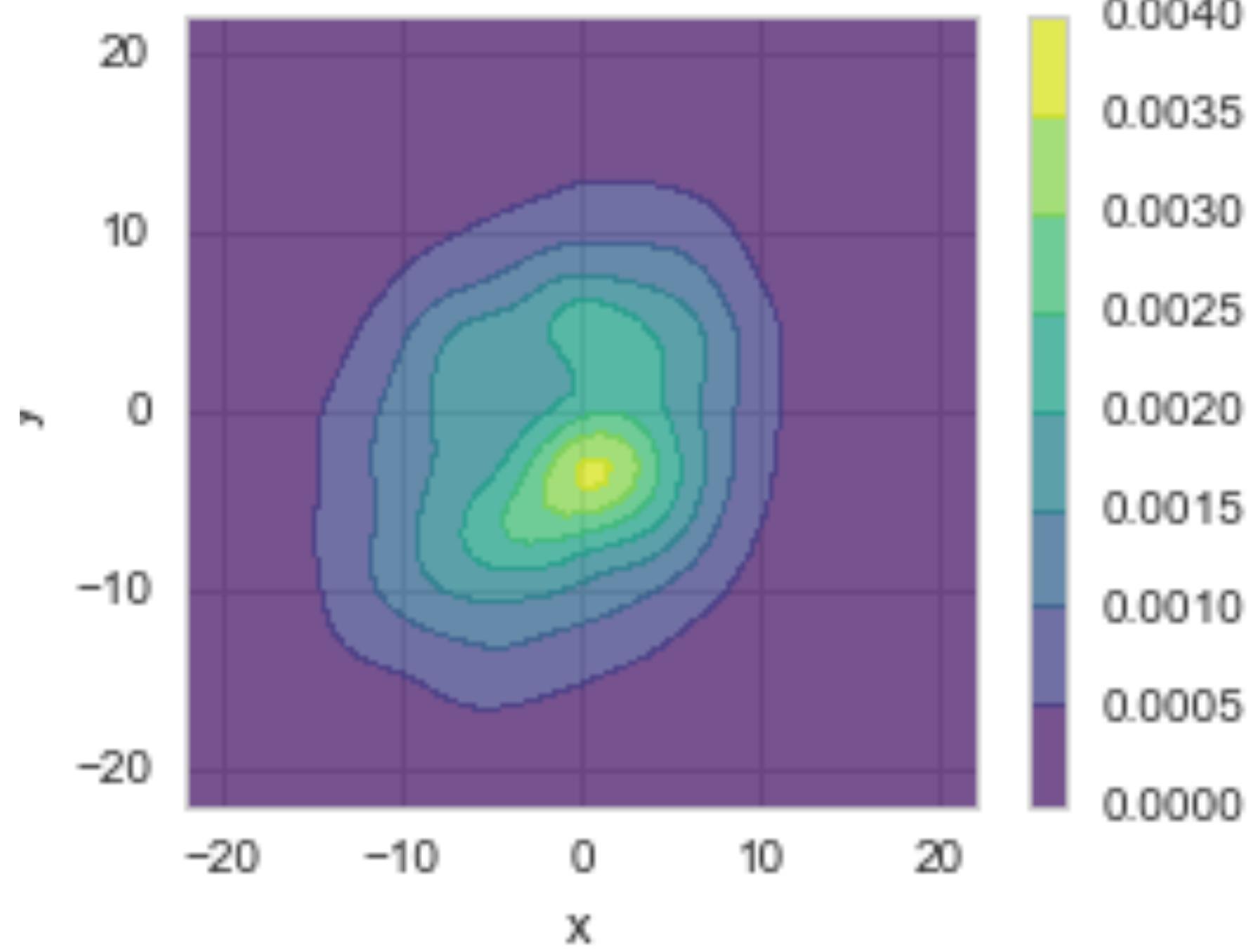
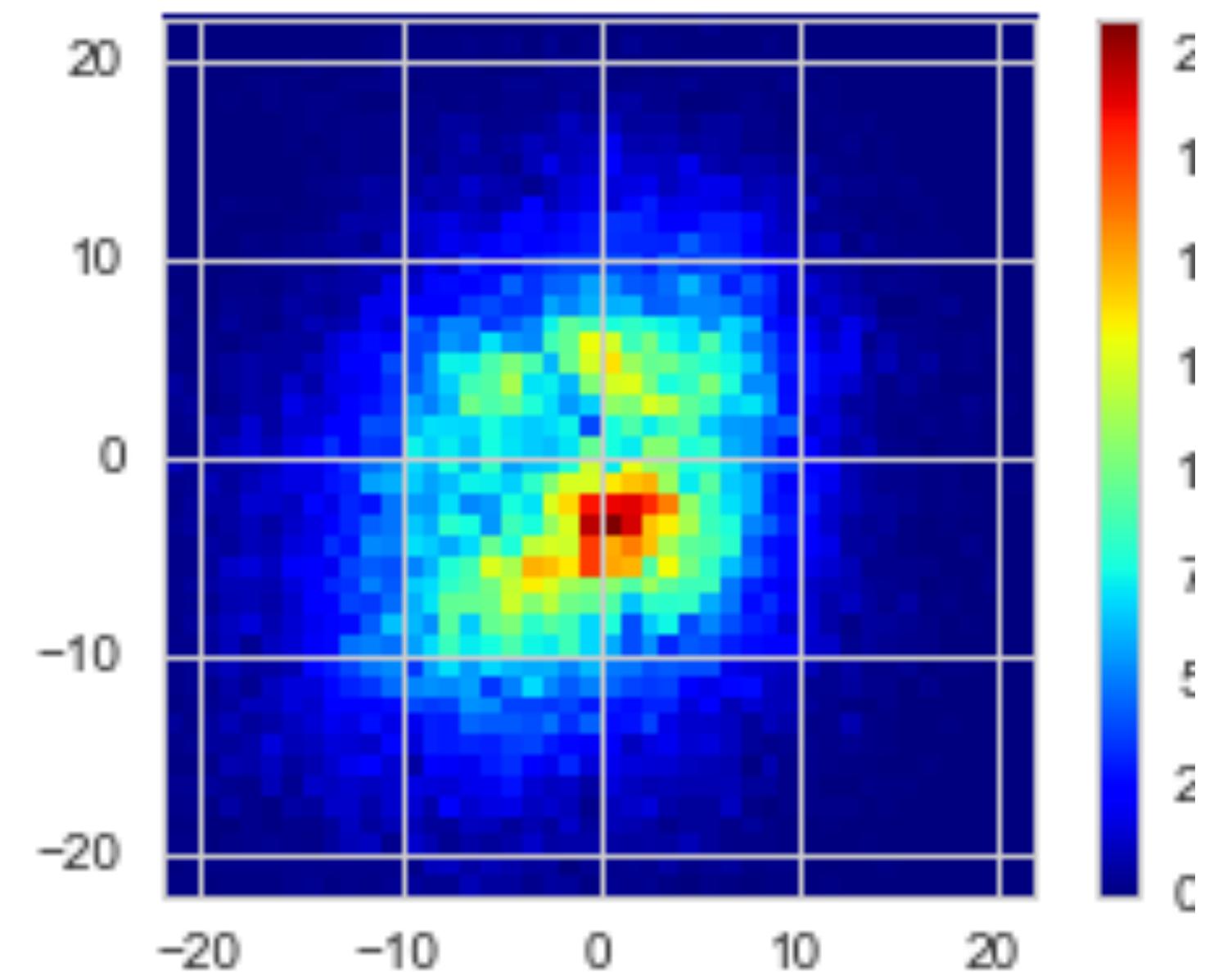
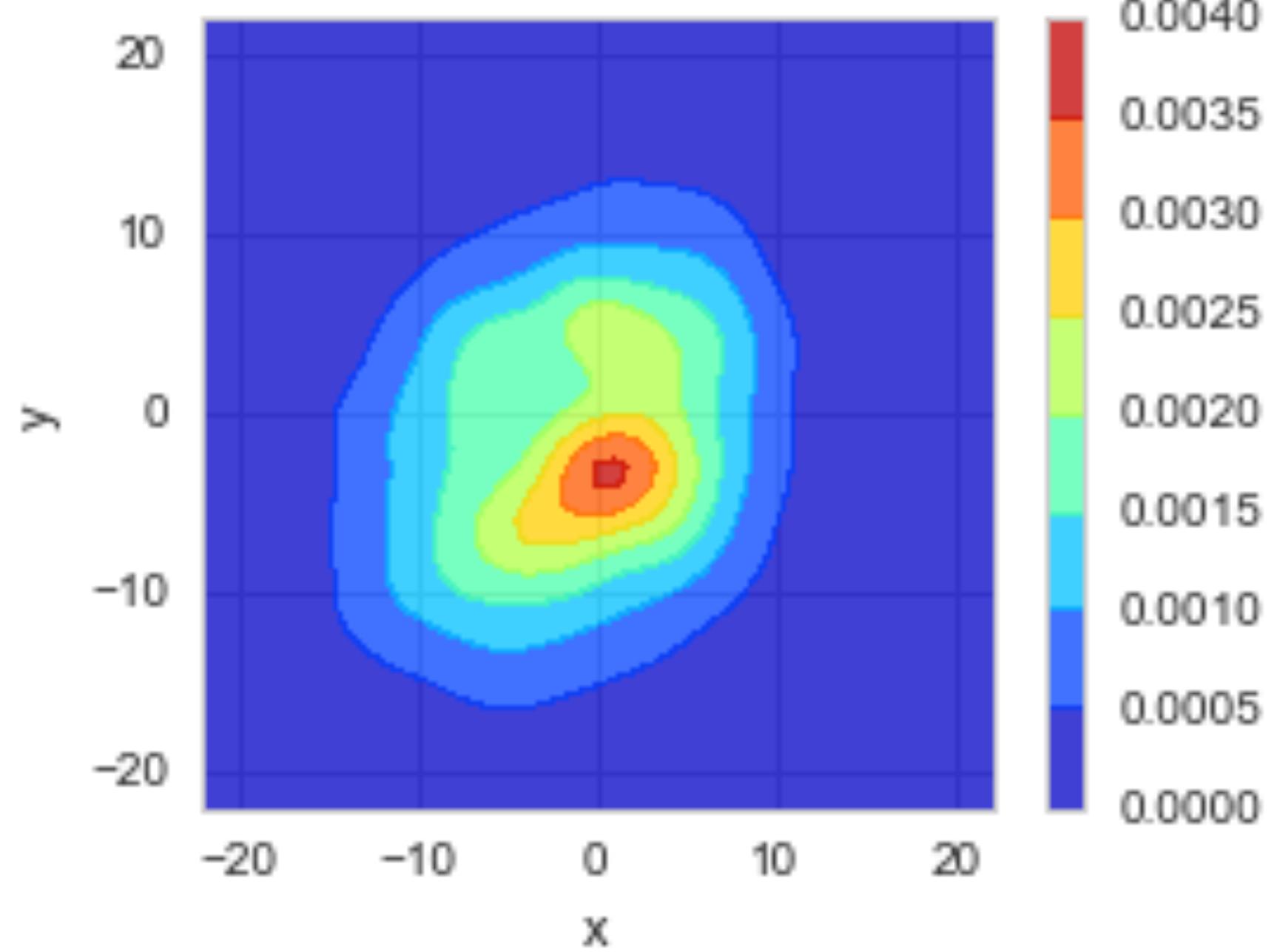


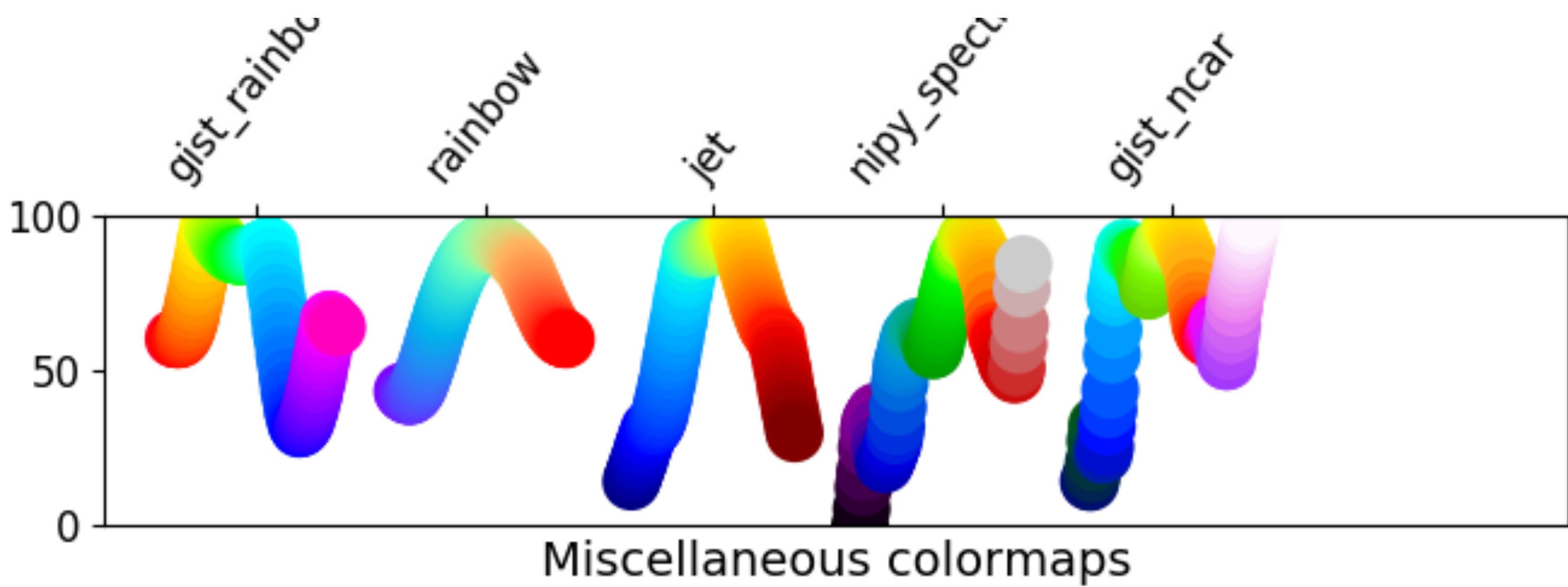
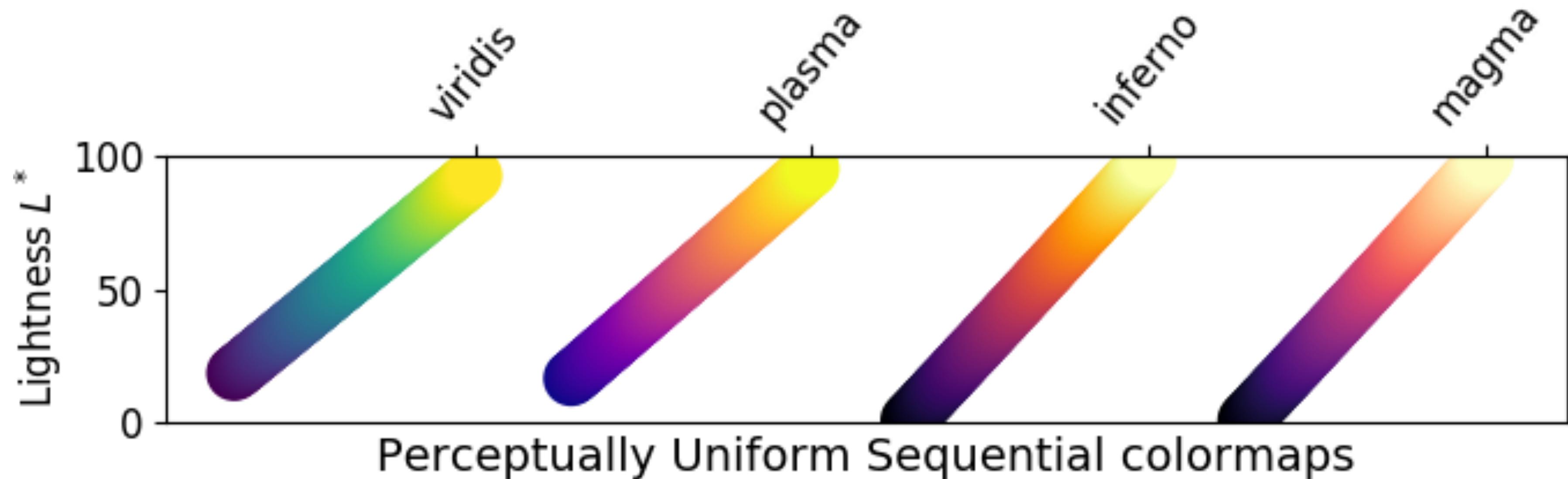
inferno

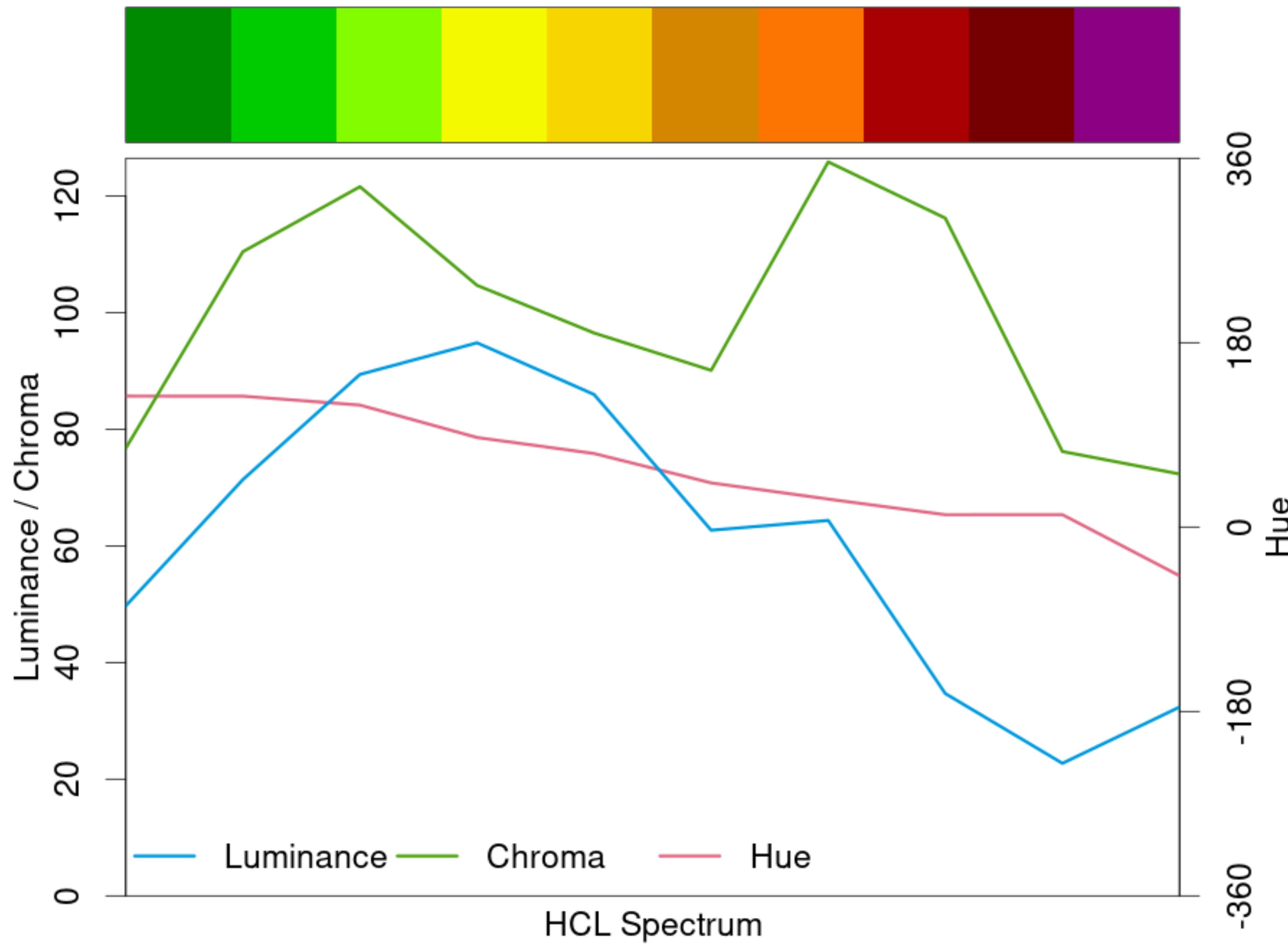


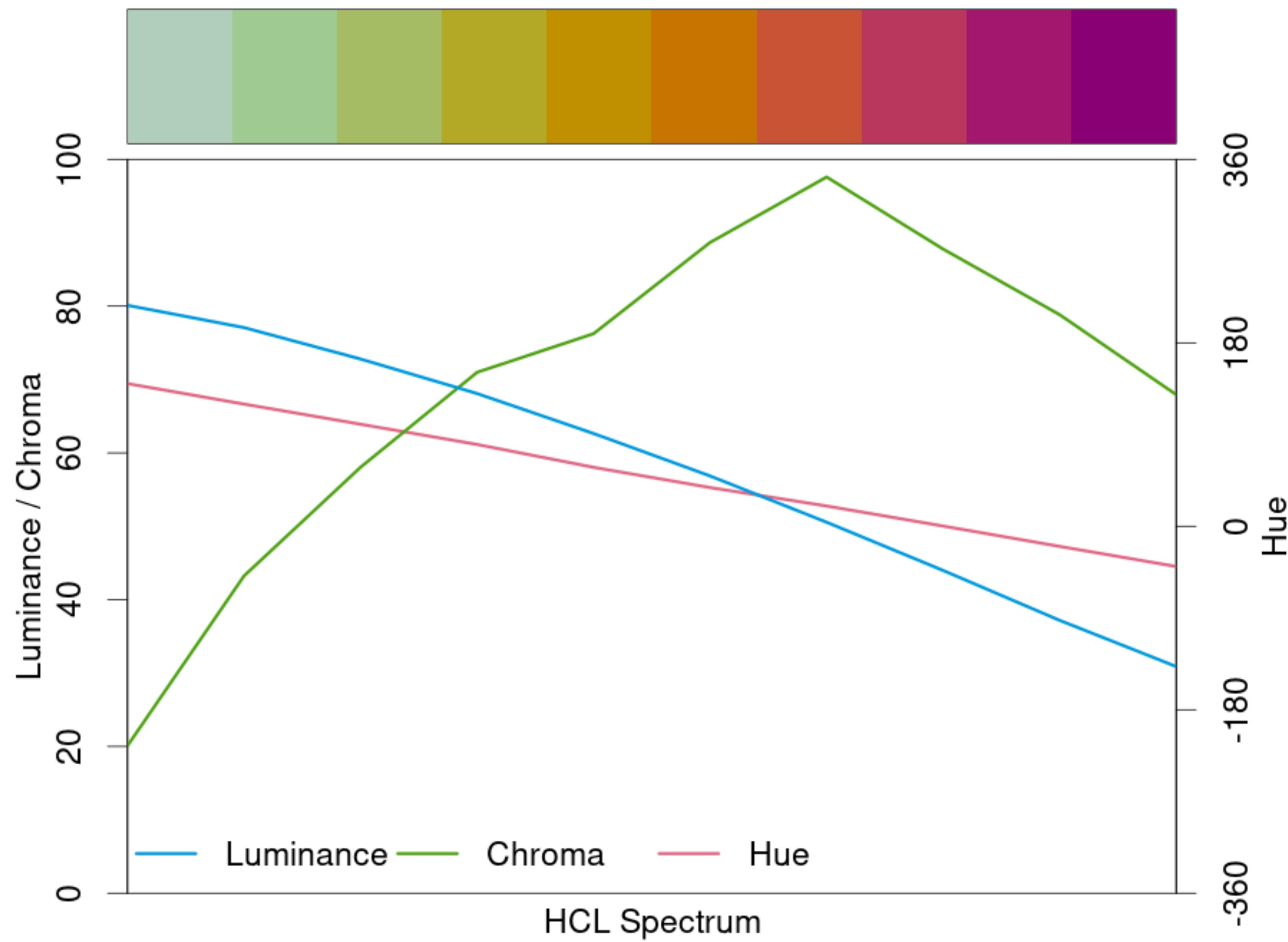
cividis

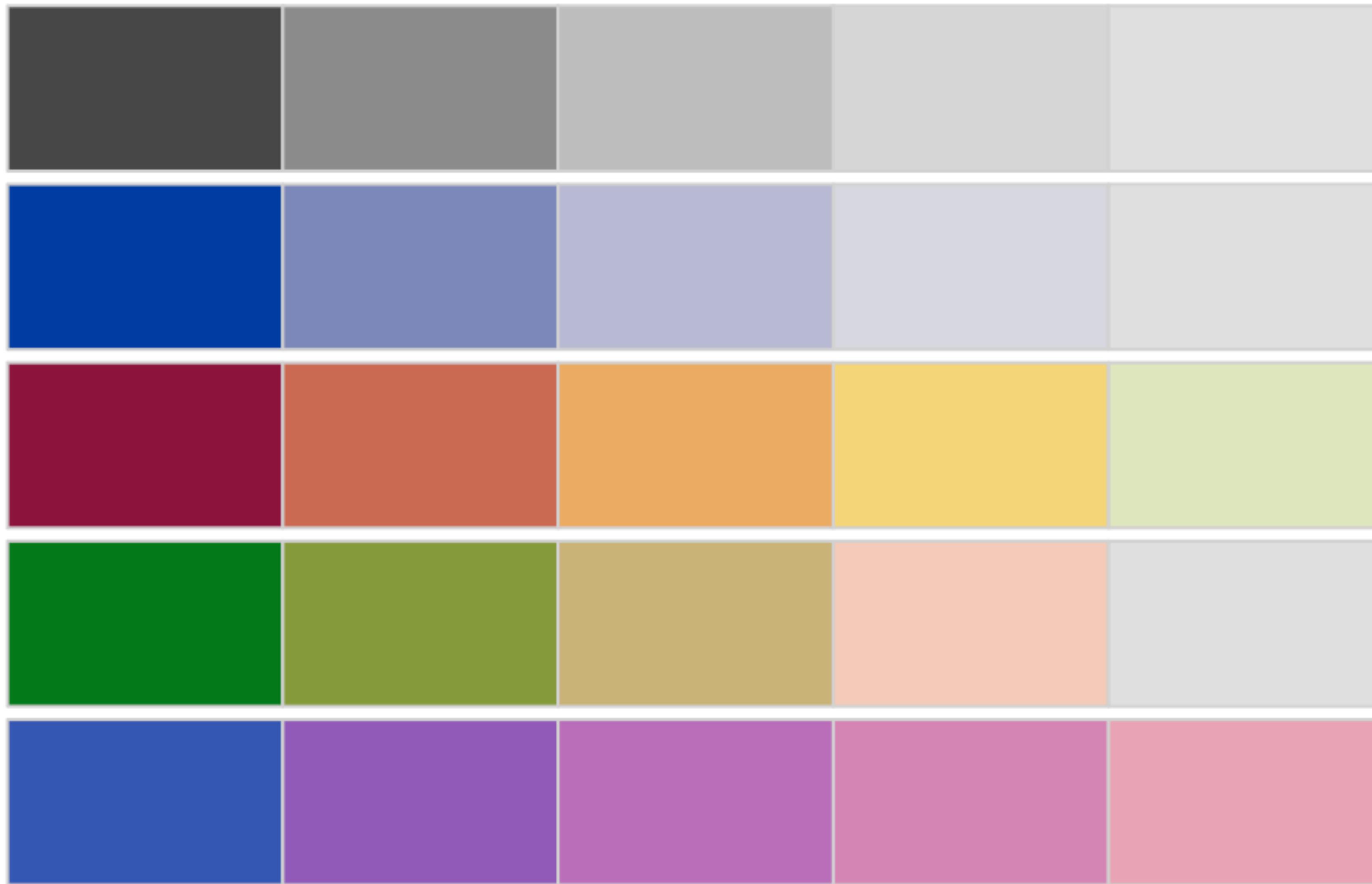


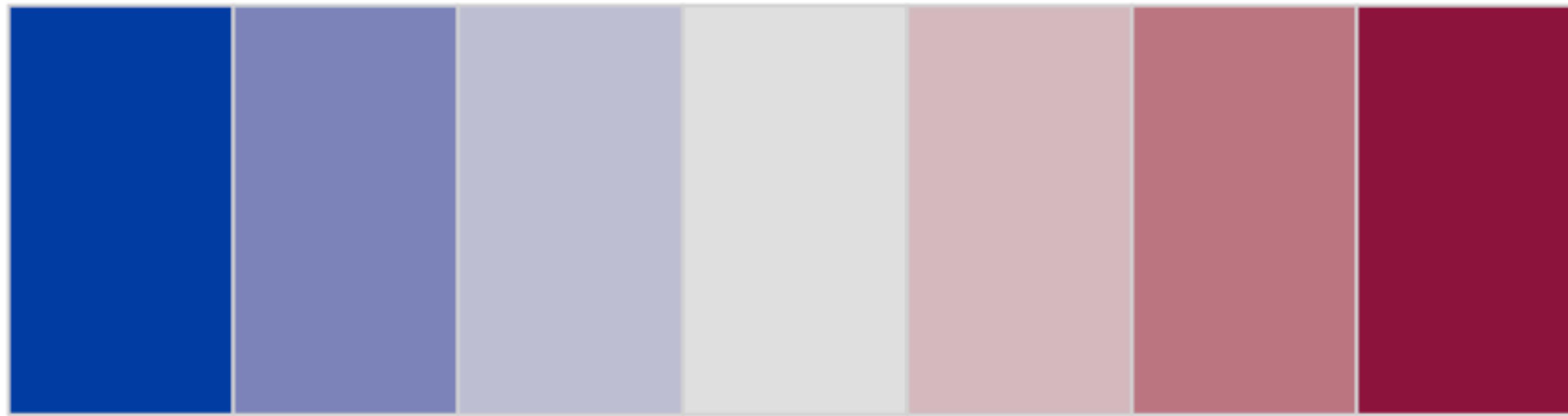








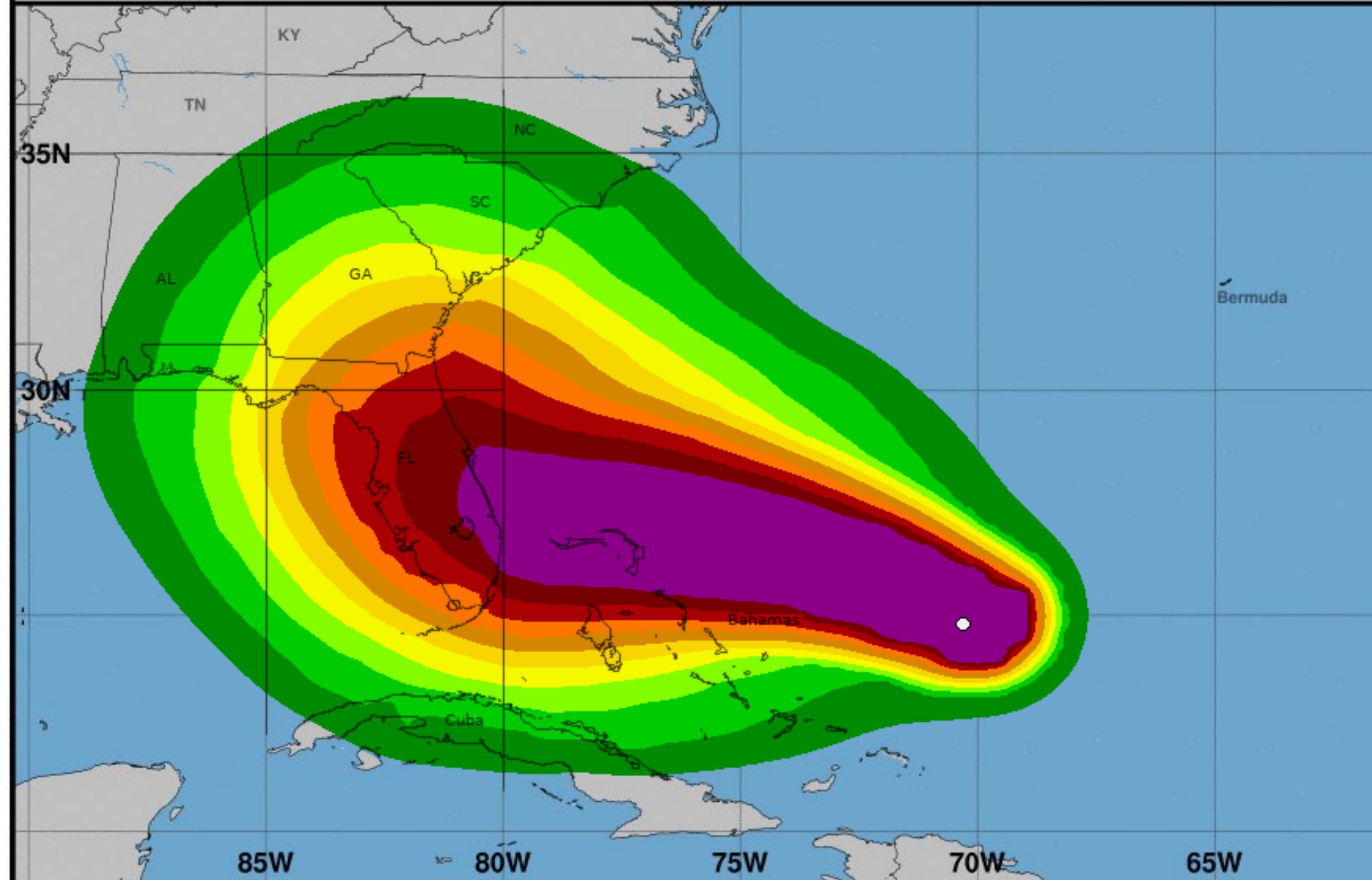






Tropical-Storm-Force Wind Speed Probabilities

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



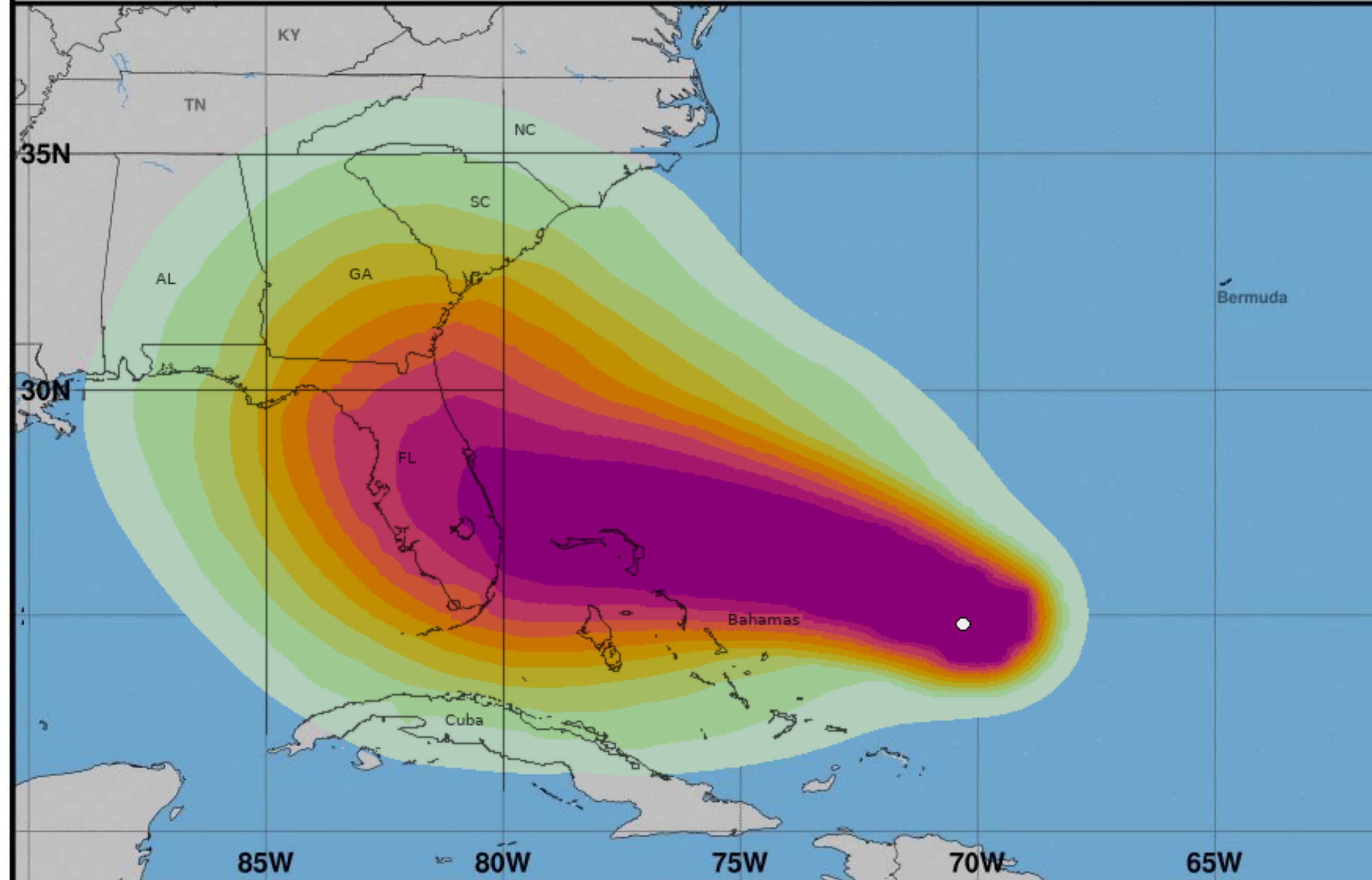
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





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)

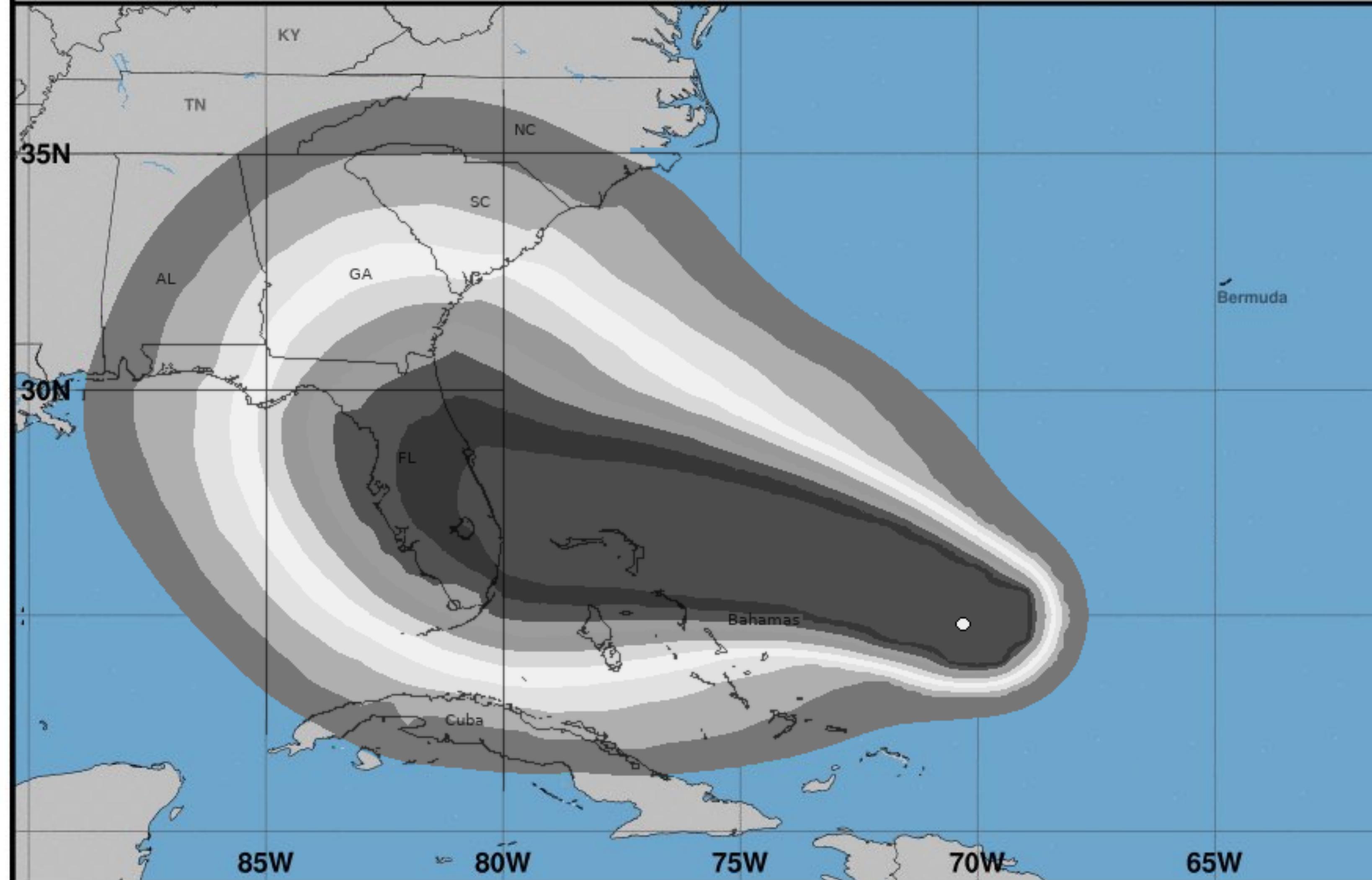


Achim Zeileis
<https://bit.ly/35zBIYL>



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



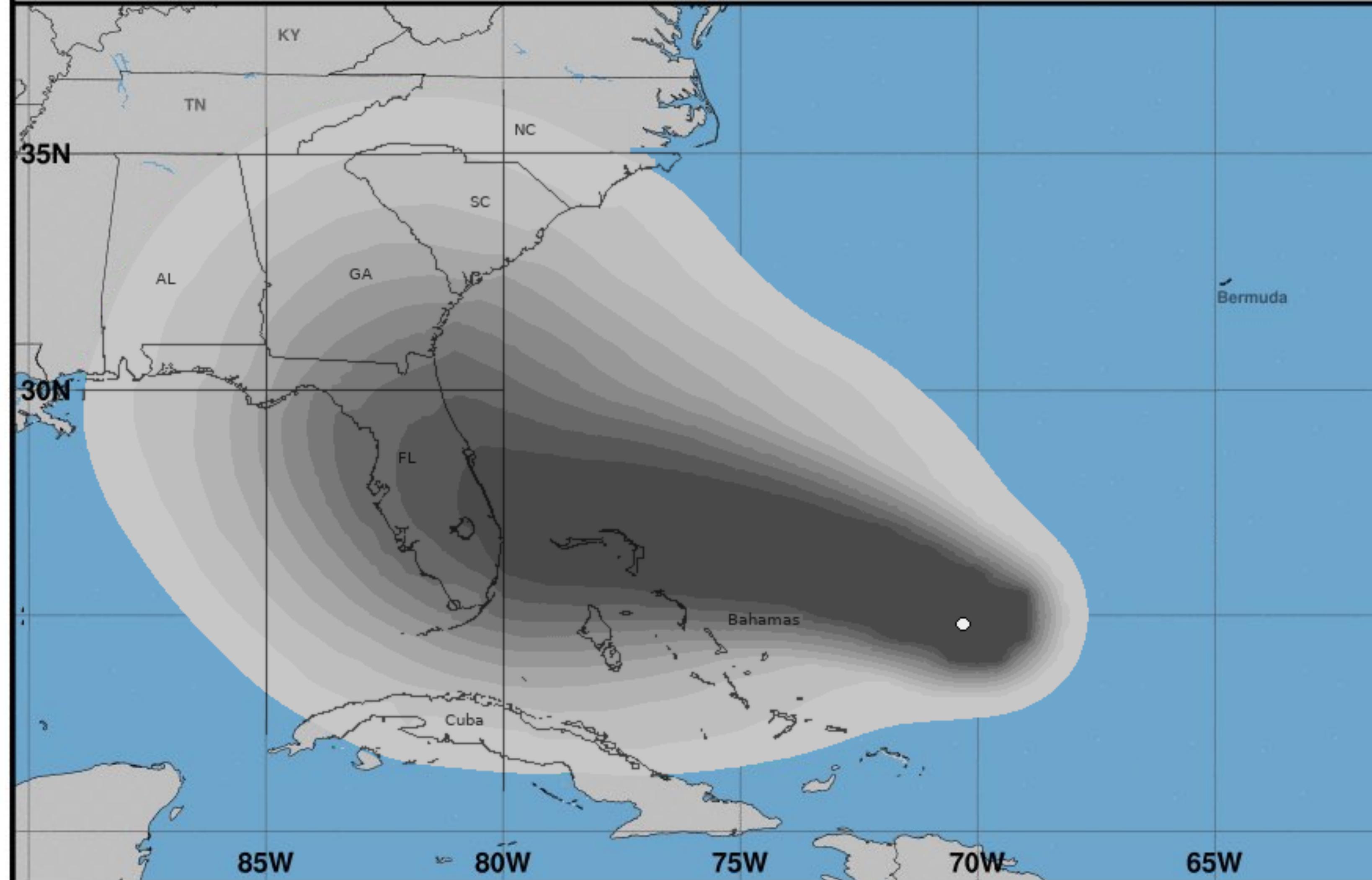
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





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



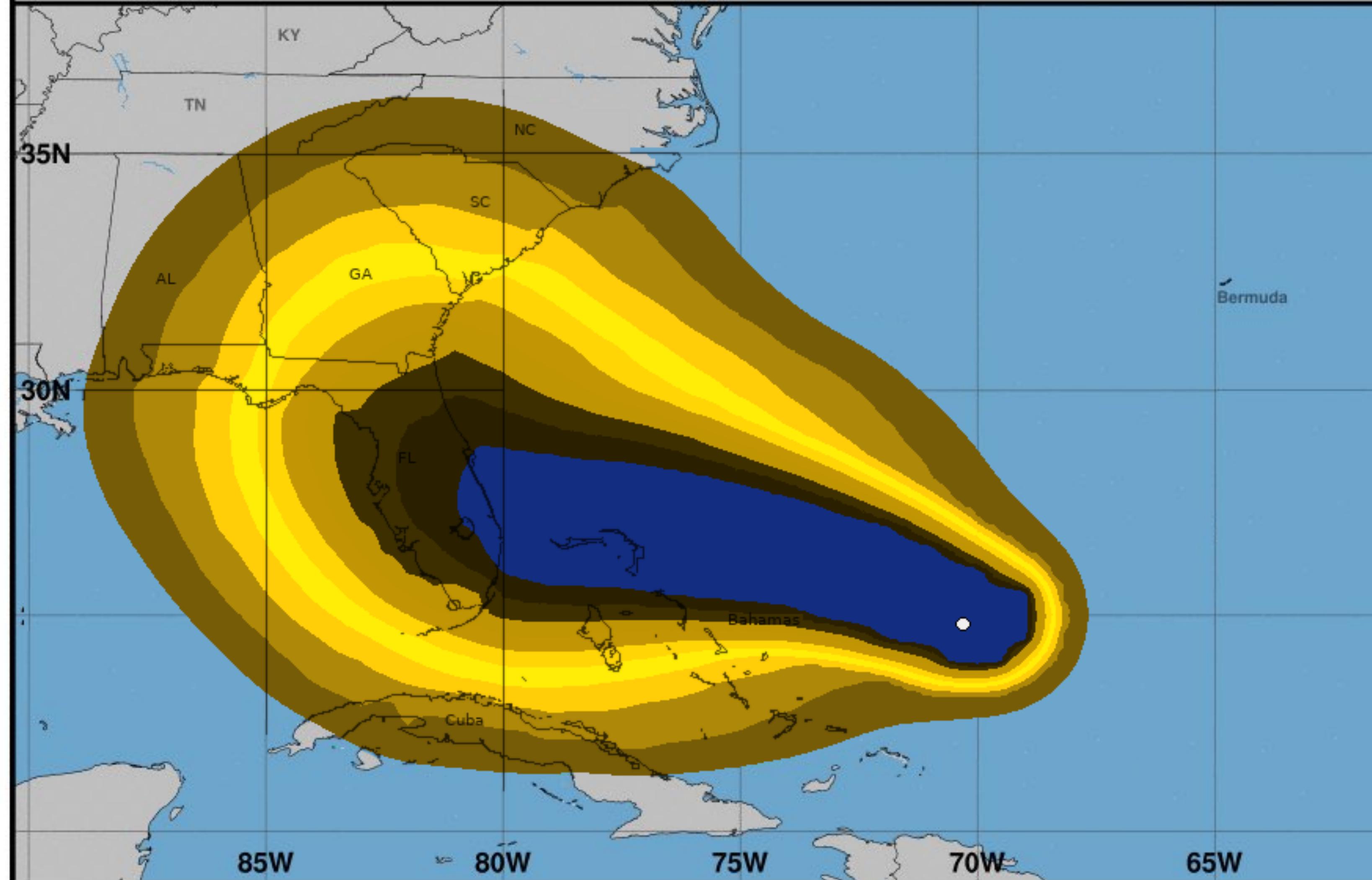
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





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



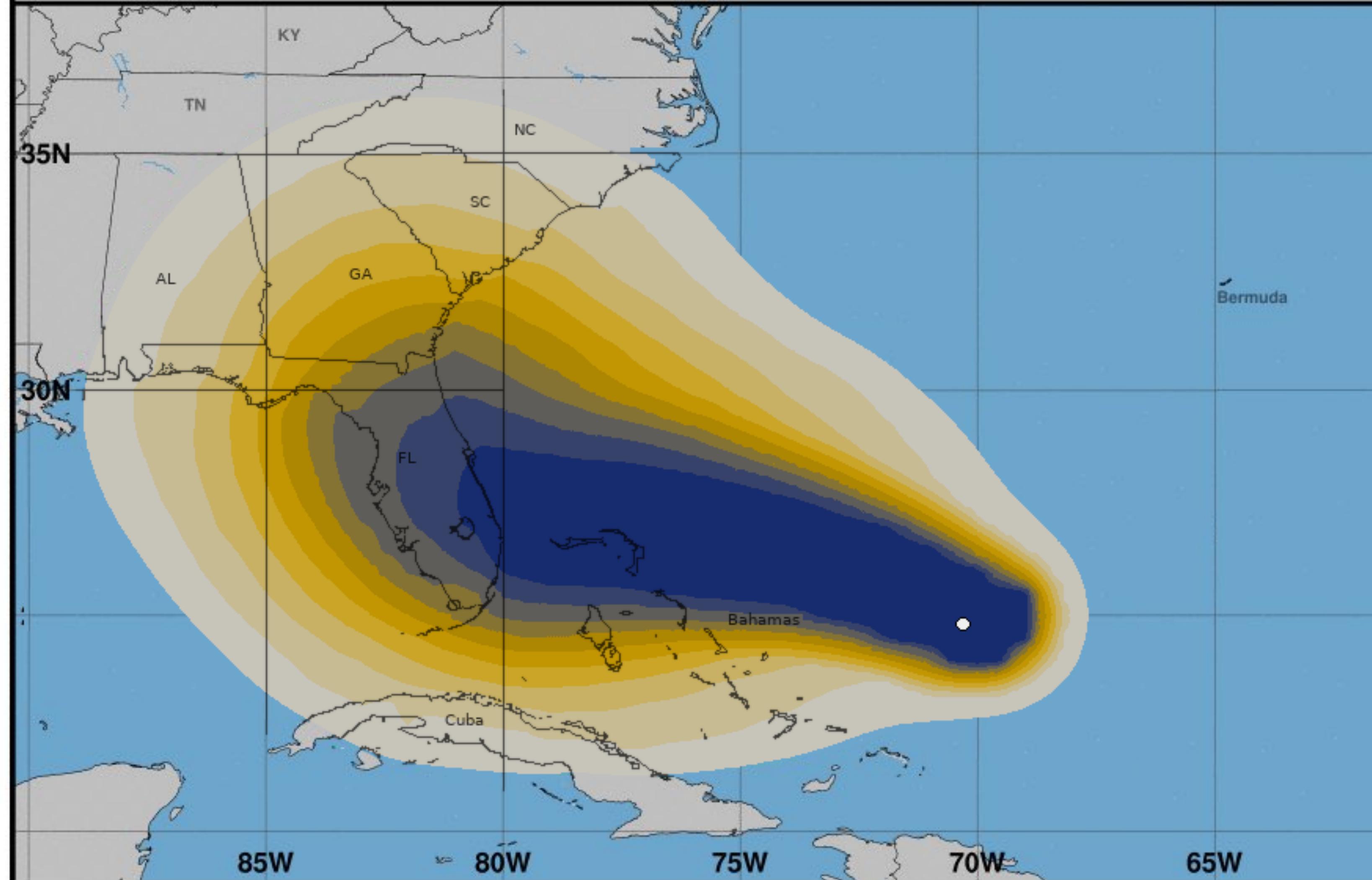
Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)





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

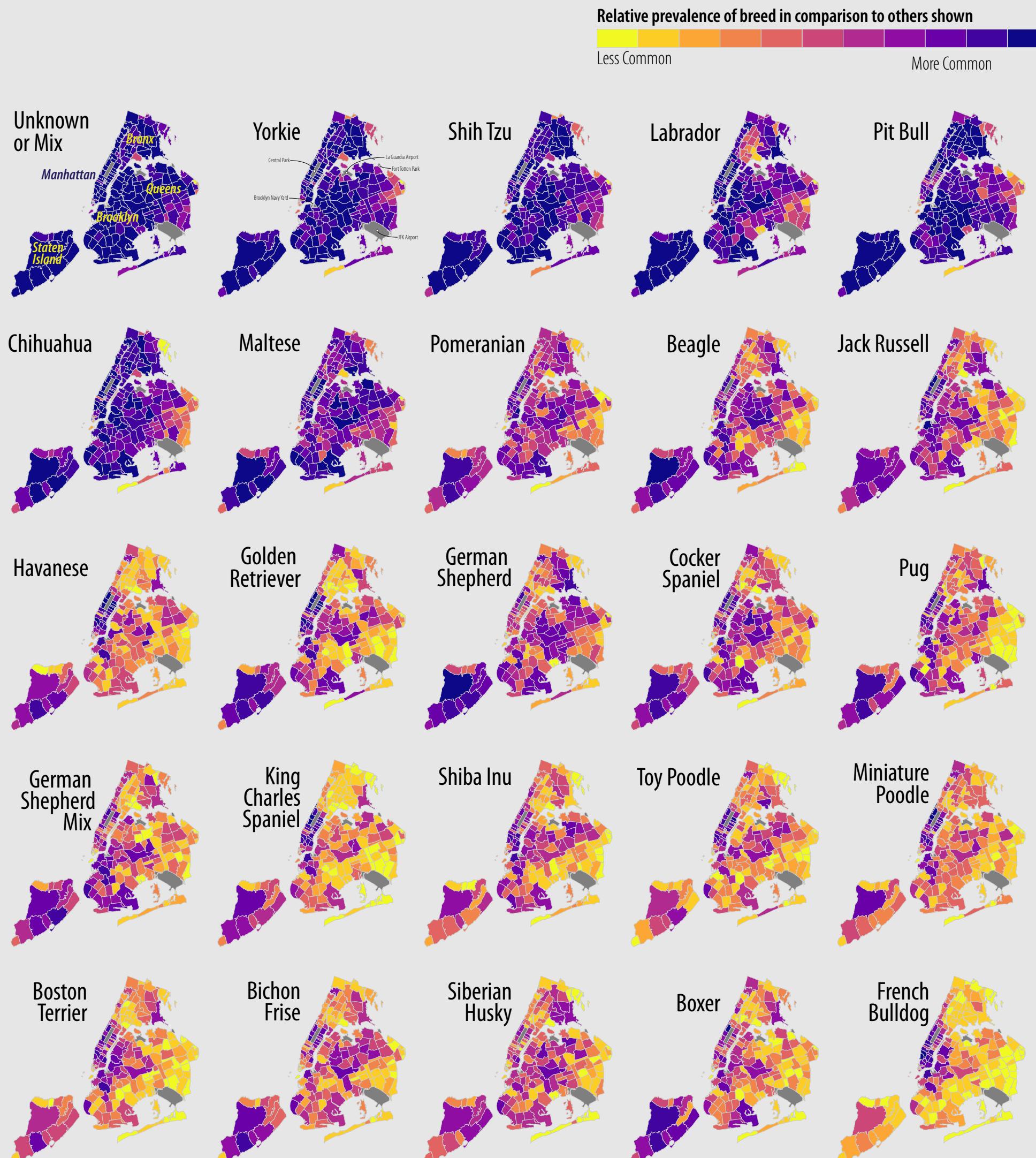


Probability of tropical-storm-force winds (1-minute average ≥ 39 mph) from all tropical cyclones
○ indicates Hurricane Dorian center location at 2 PM EDT FRI AUG 30, 2019 (Forecast/Advisory #26)



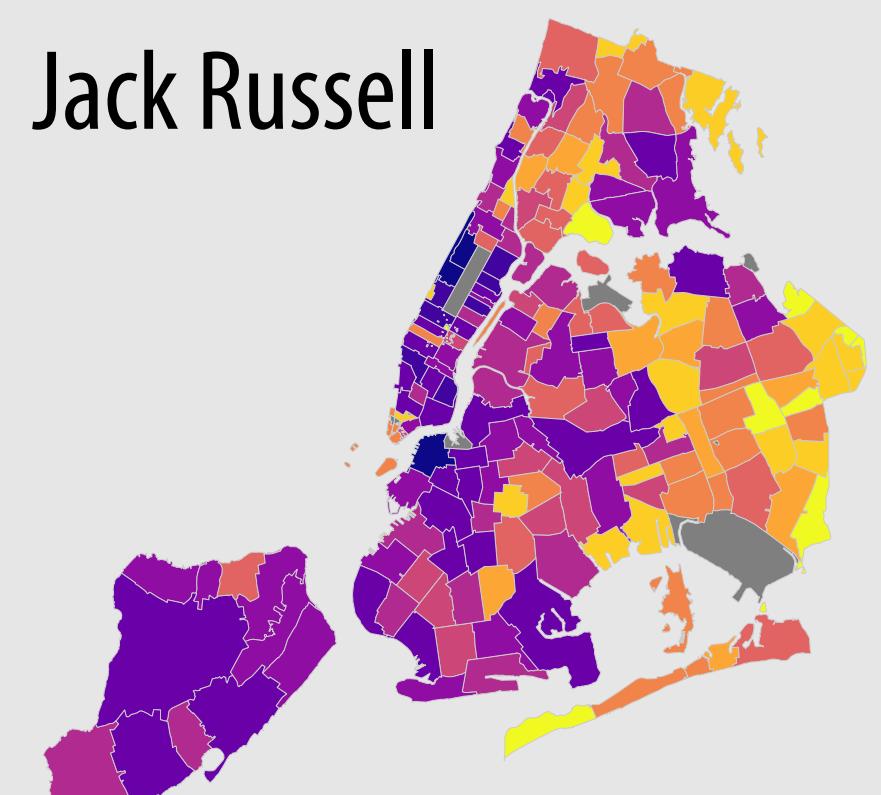
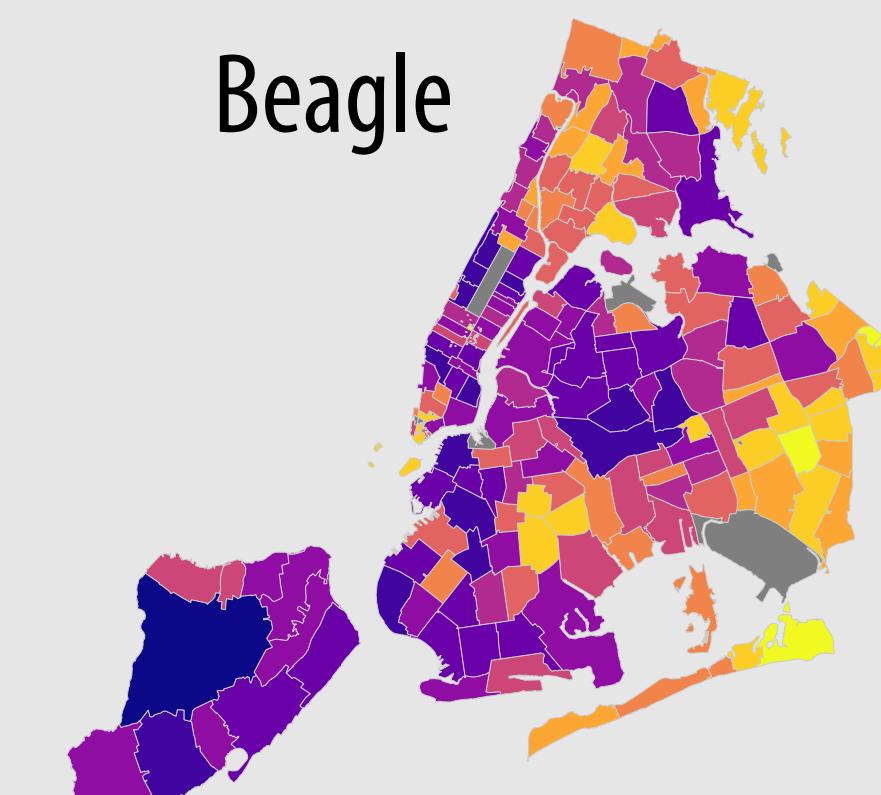
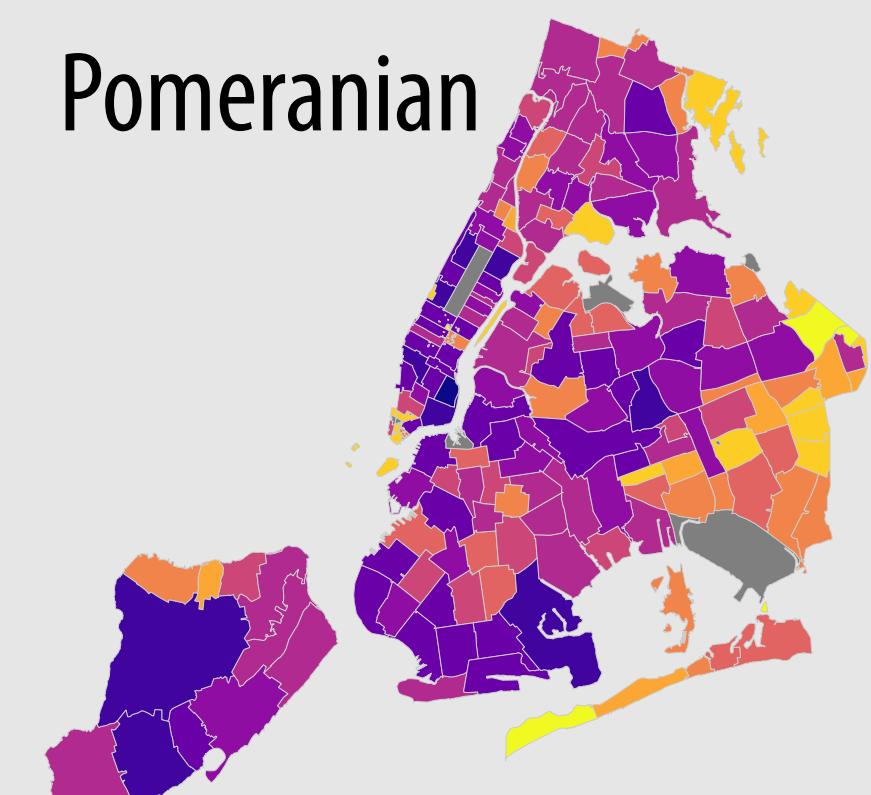
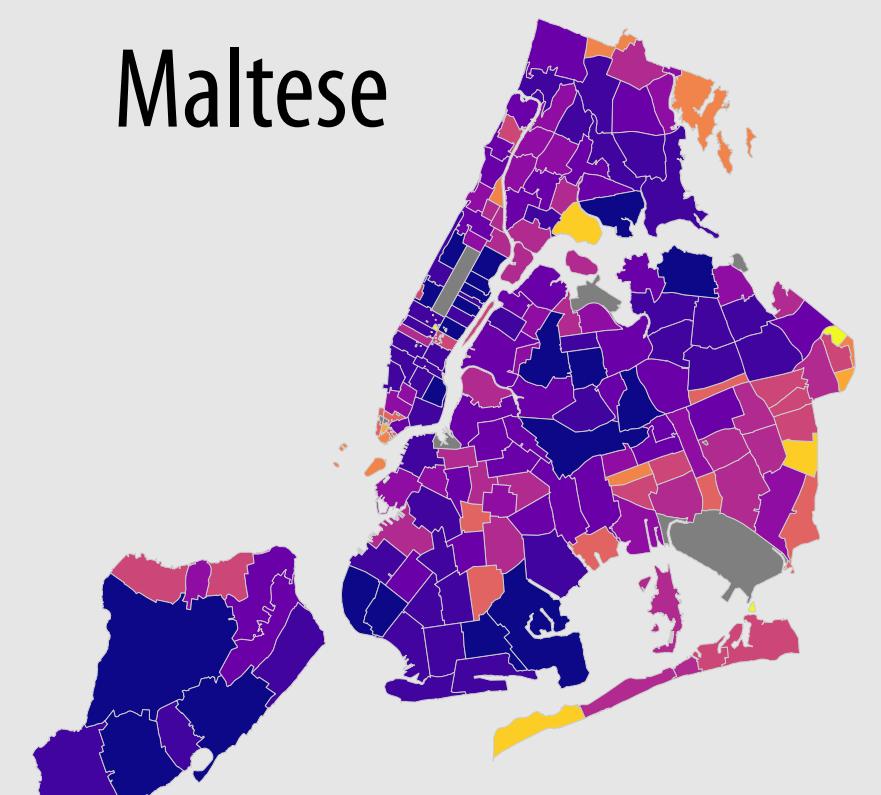
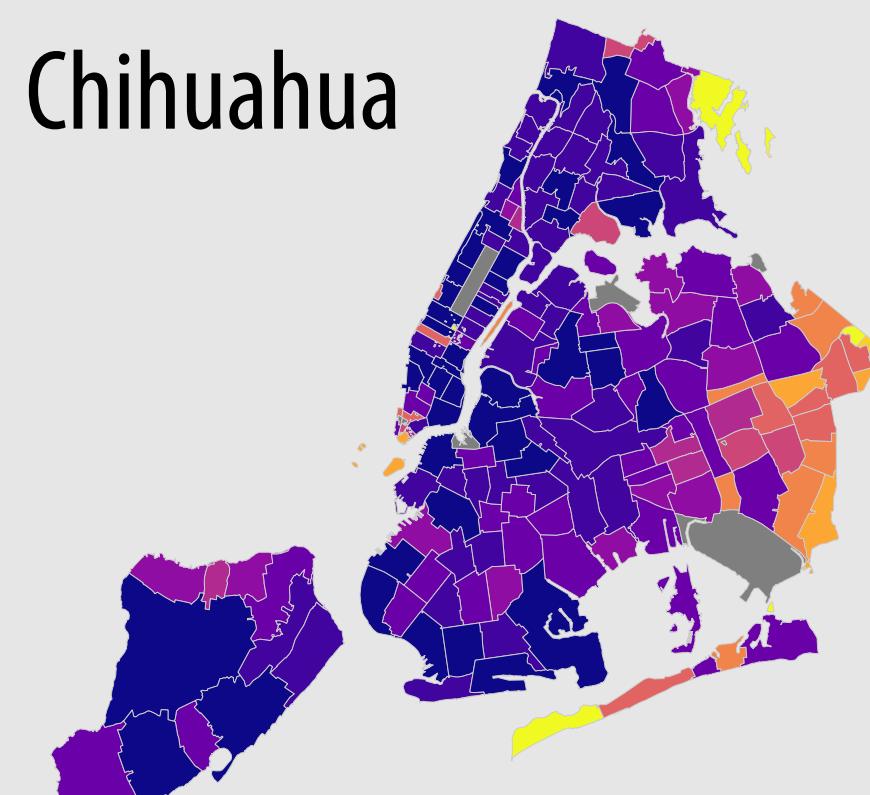
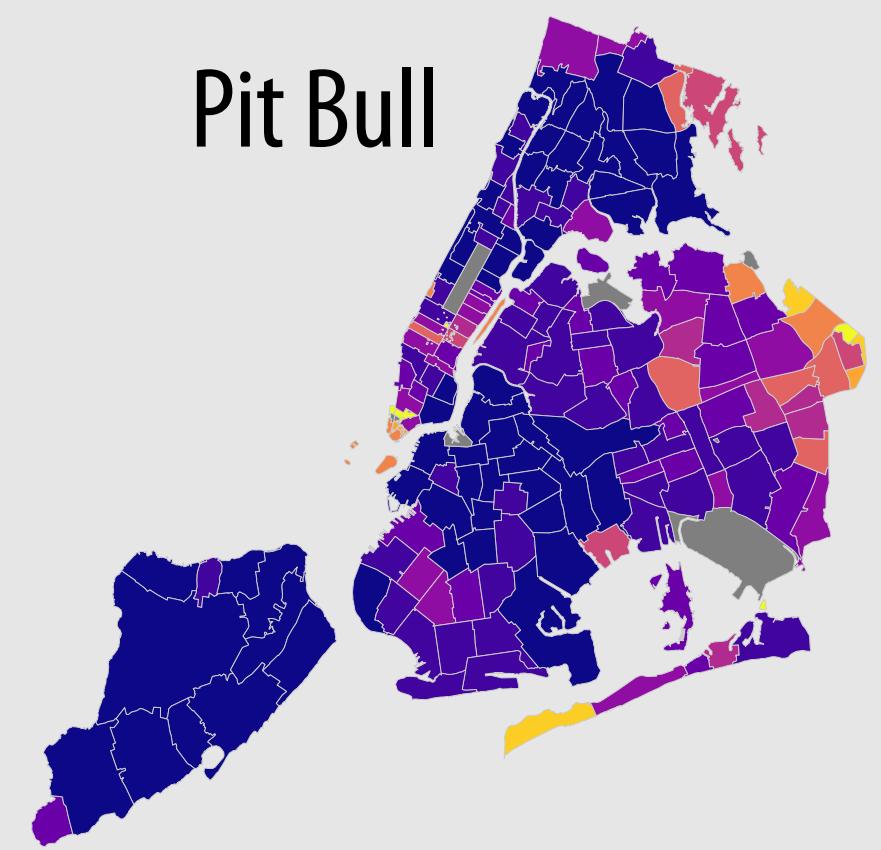
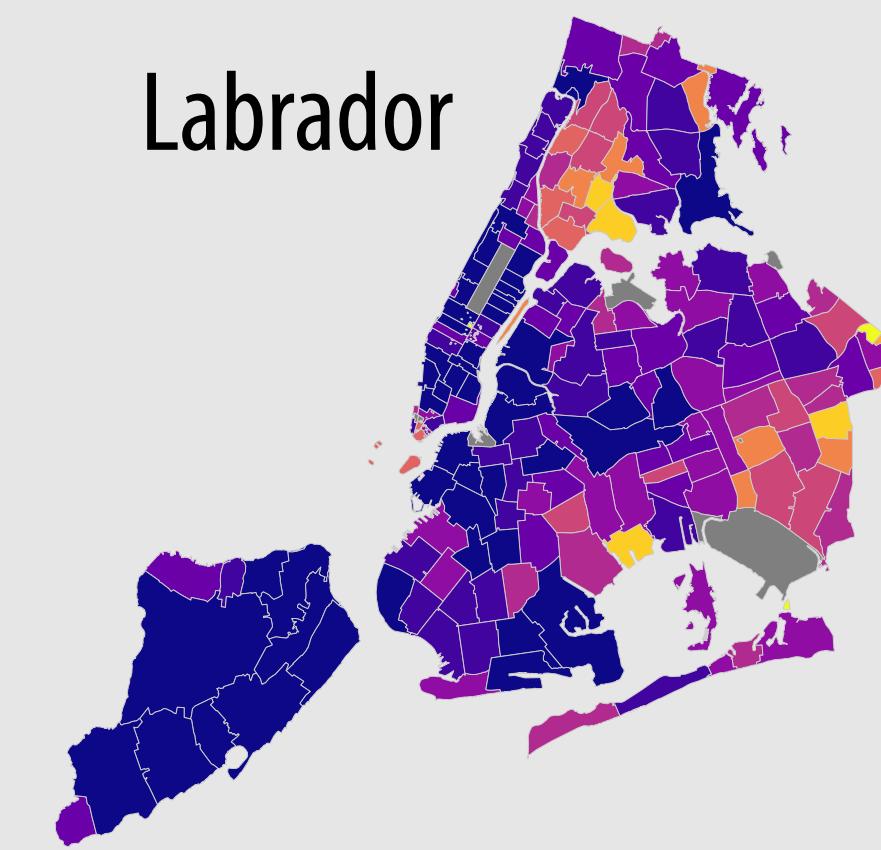
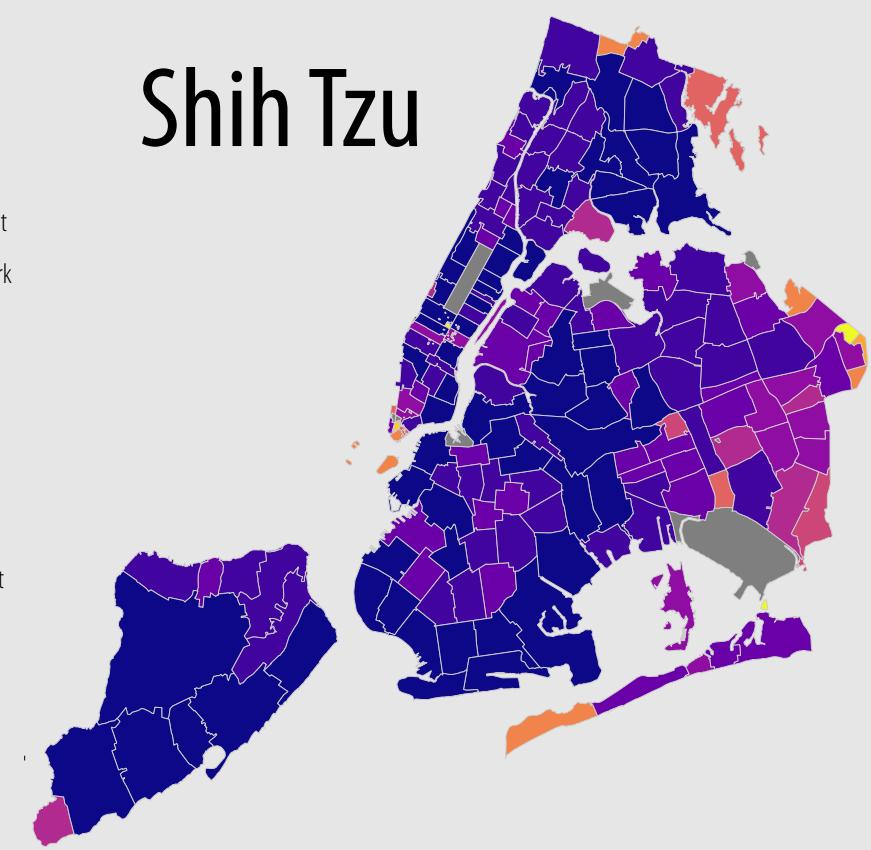
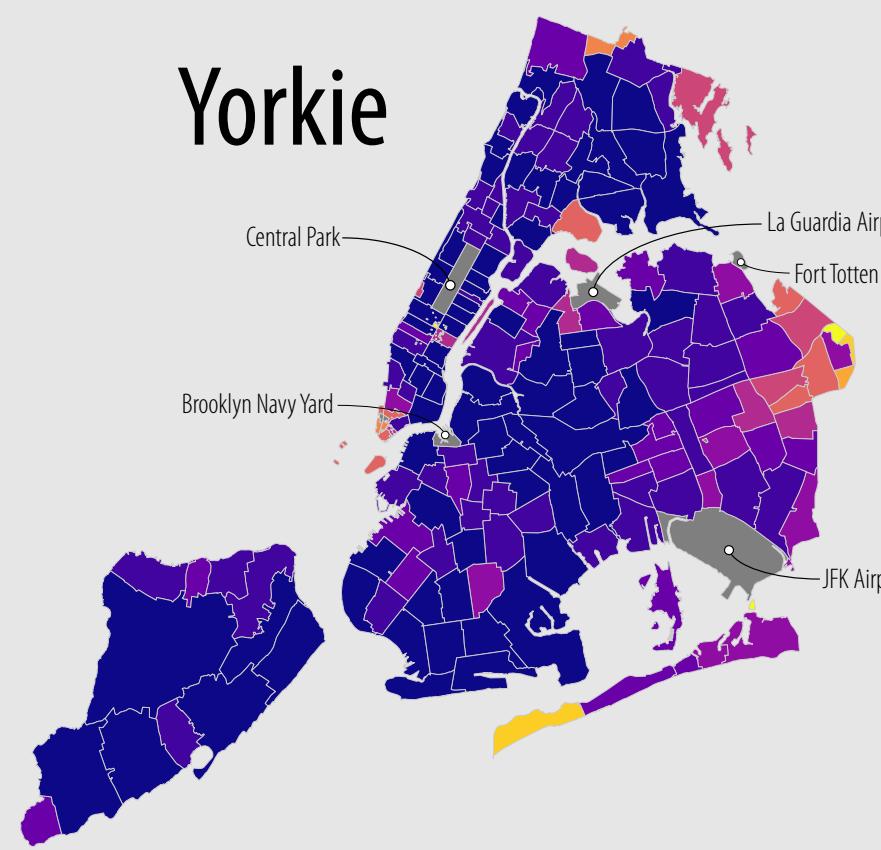
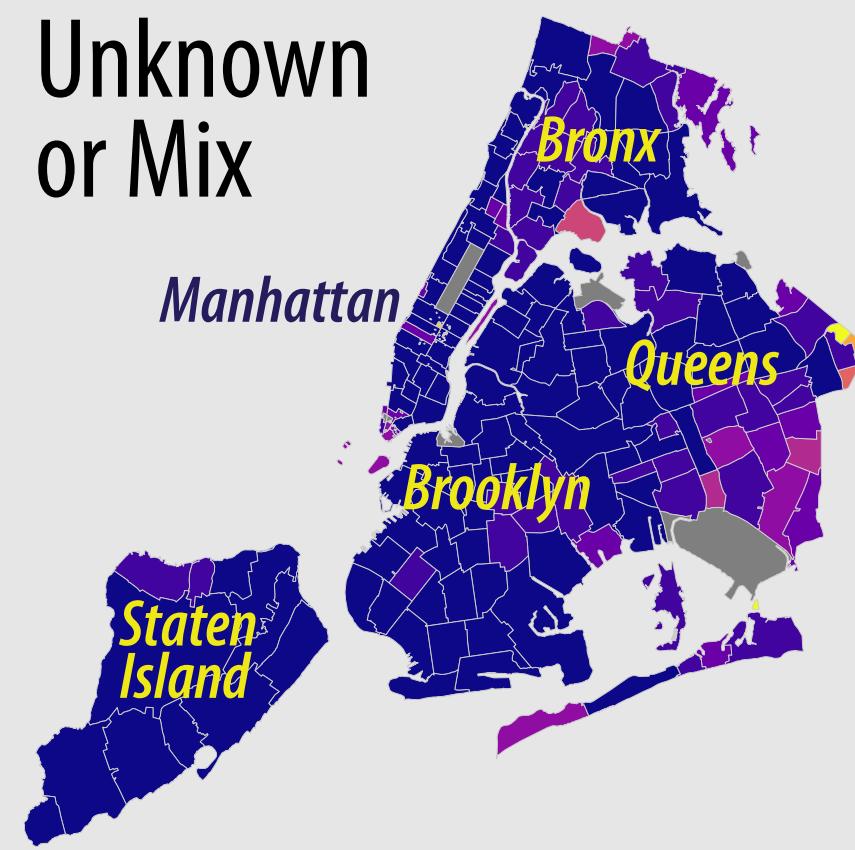
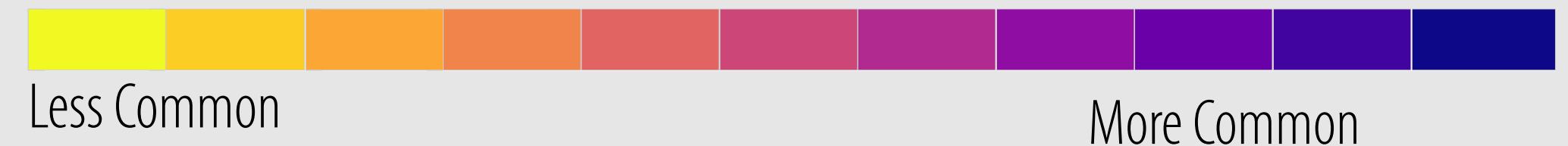
Dogs of New York

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.



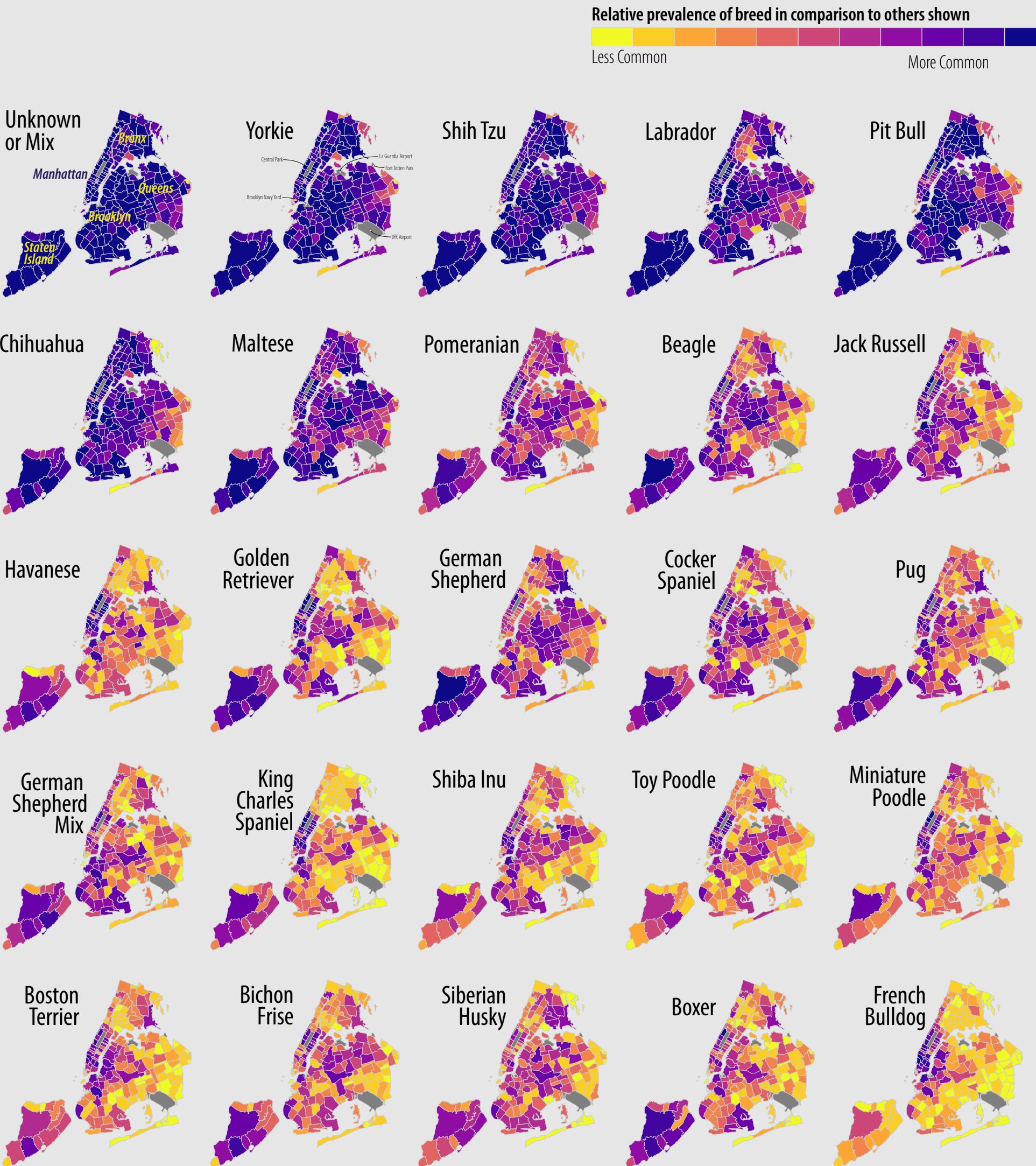
Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.

Relative prevalence of breed in comparison to others shown



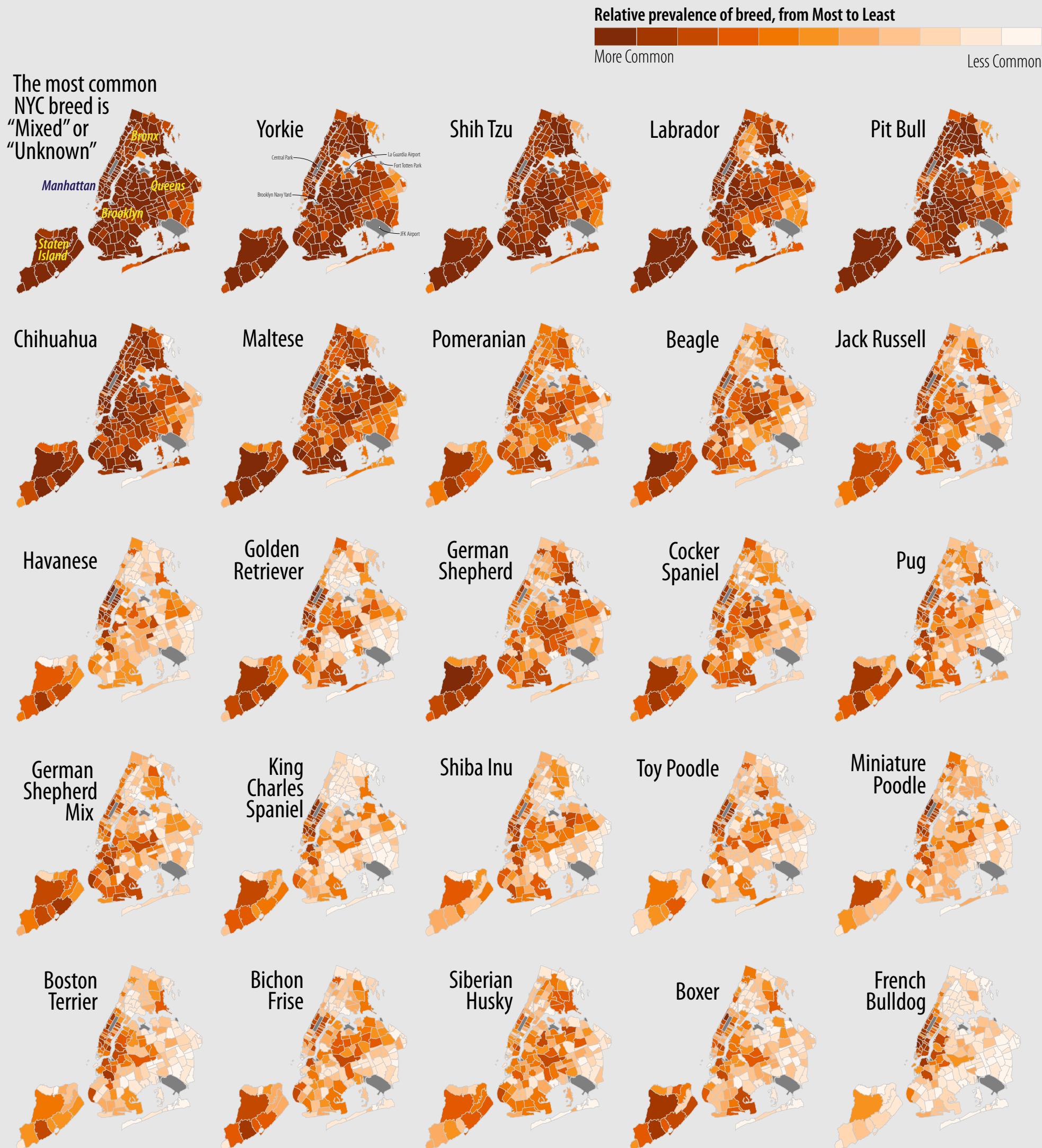
Dogs of New York

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

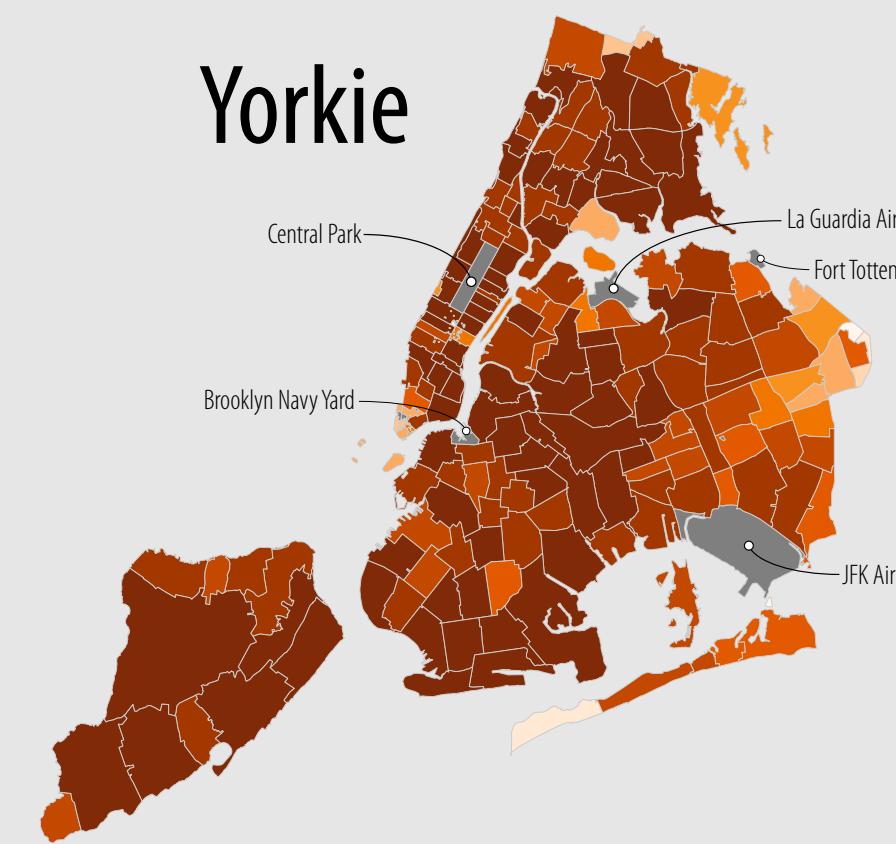
Based on data from New York City's Dog Licensing System, these maps show the relative prevalence of the twenty five most common breeds of dog, by zip code.



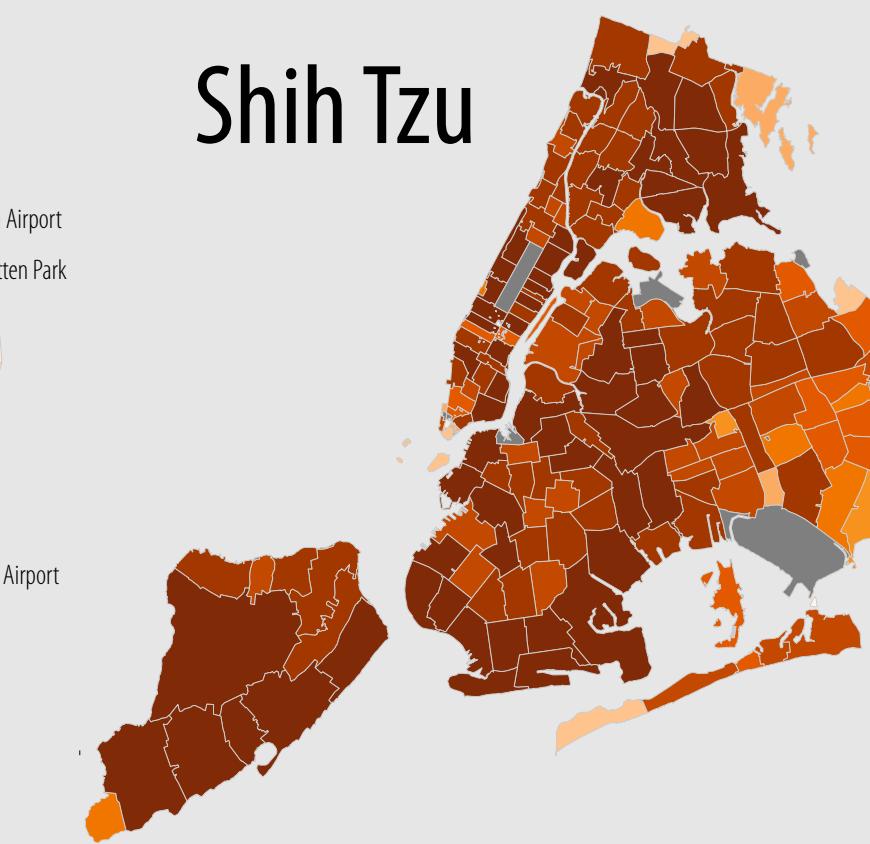
The most common
NYC breed is
“Mixed” or
“Unknown”



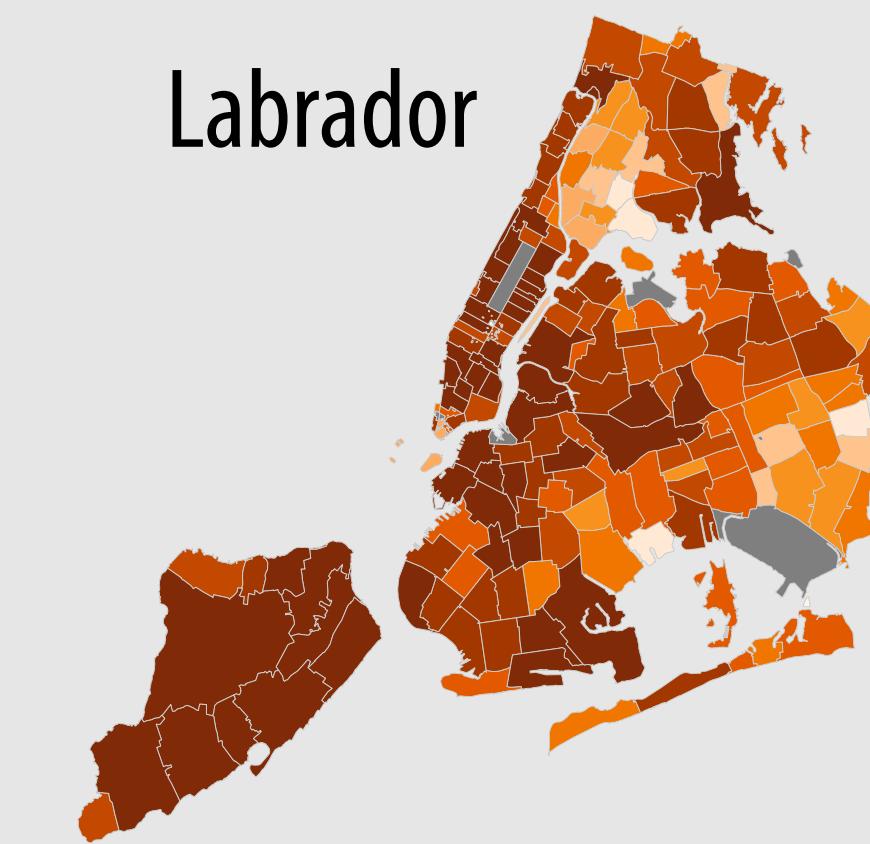
Yorkie



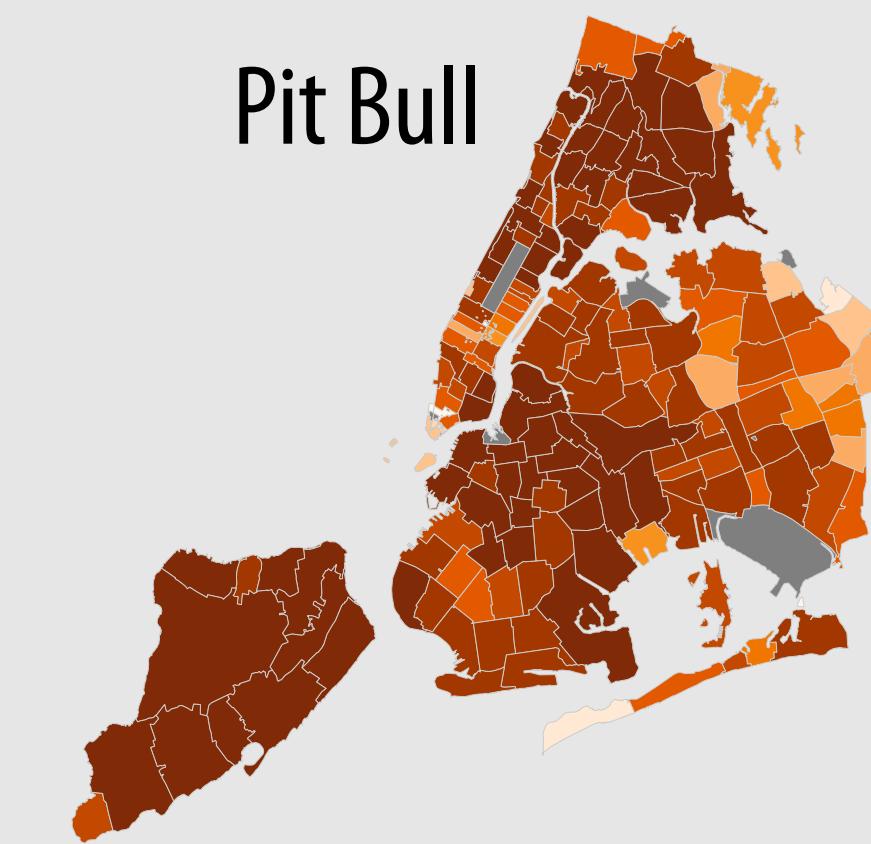
Shih Tzu



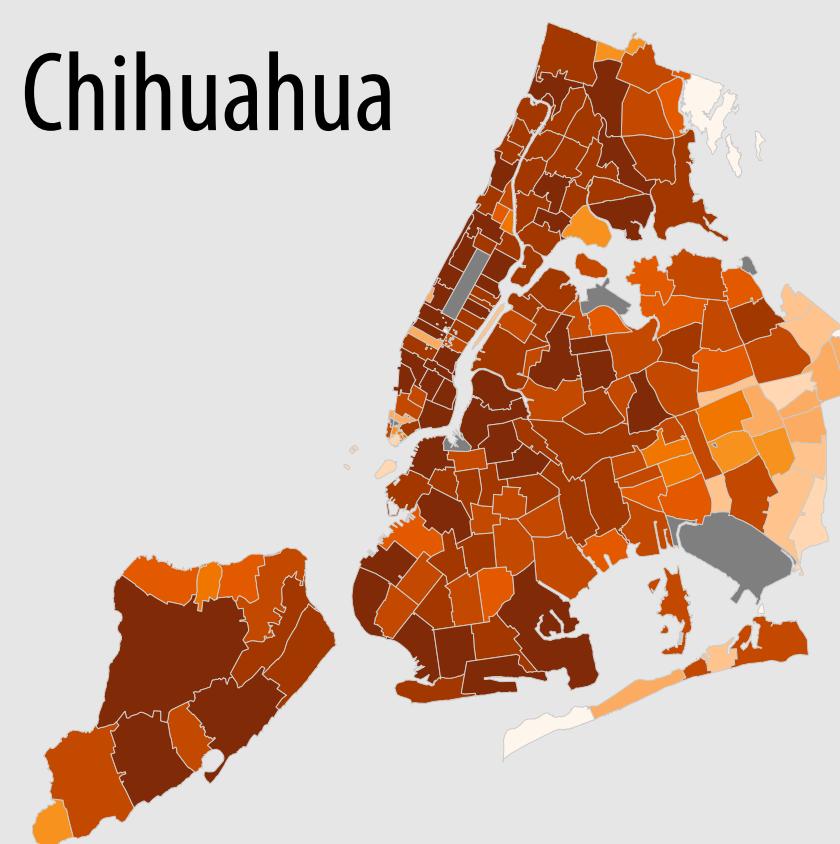
Labrador



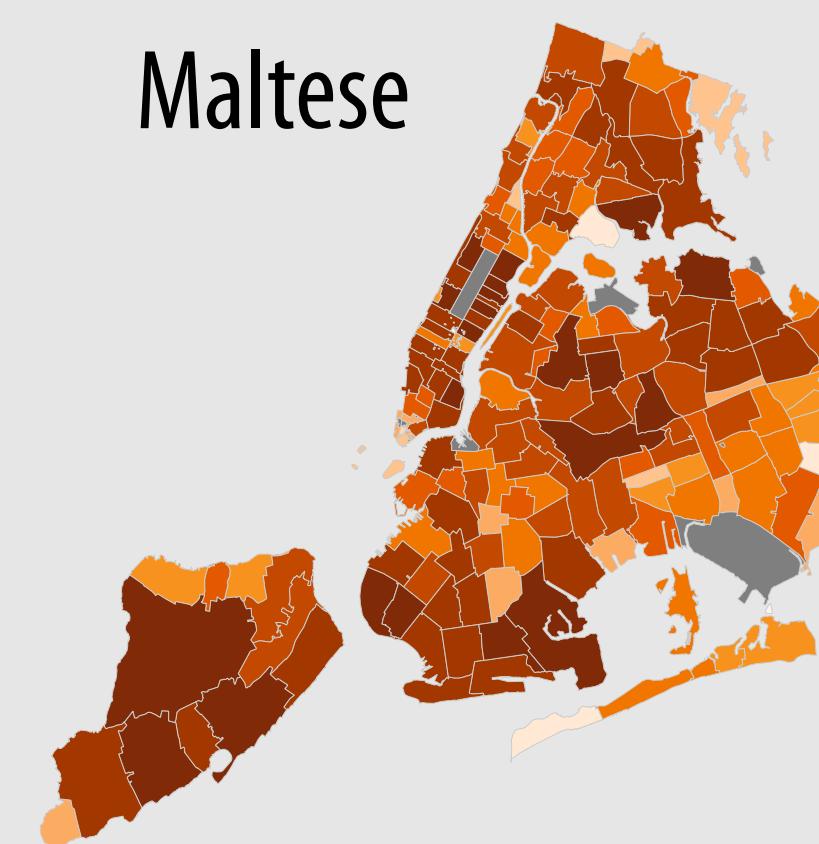
Pit Bull



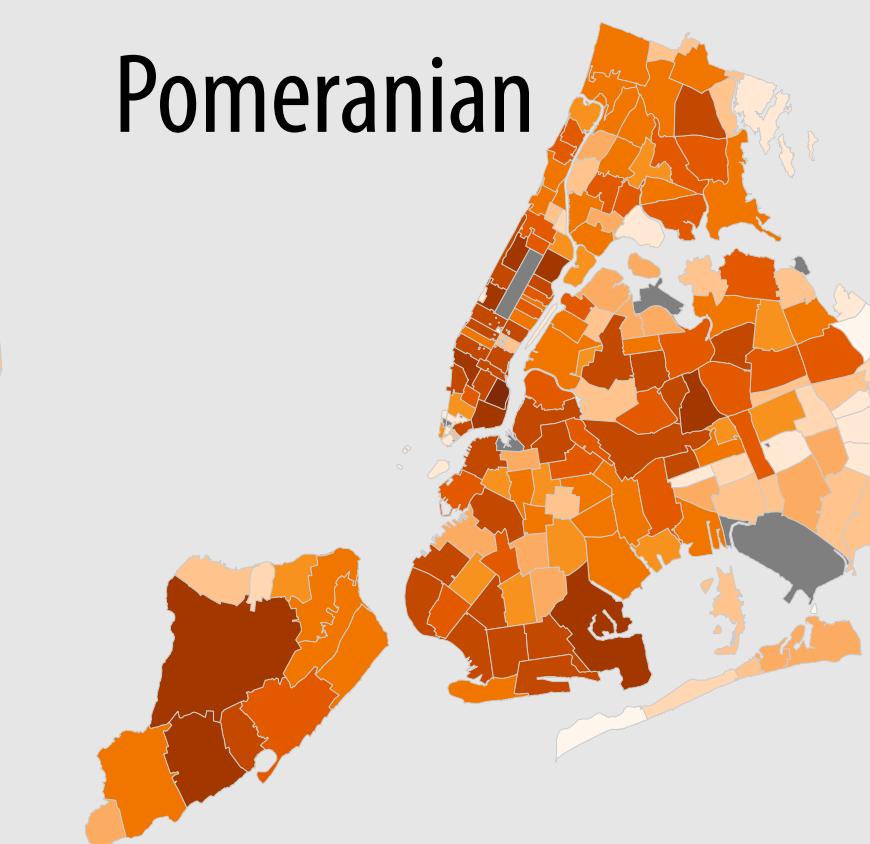
Chihuahua



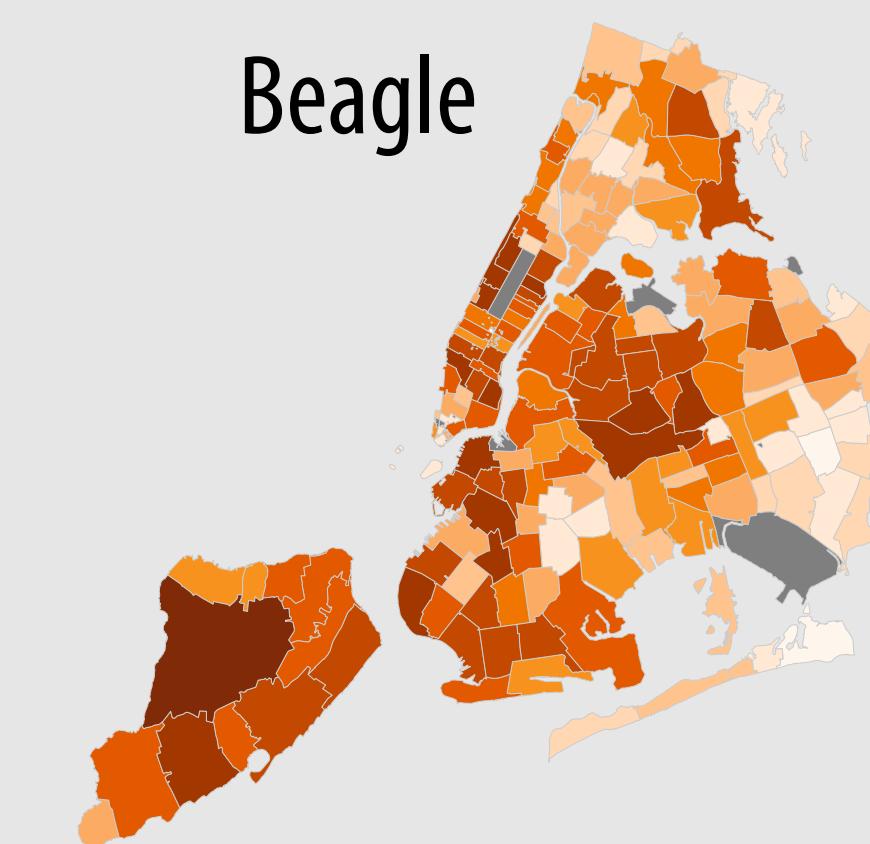
Maltese



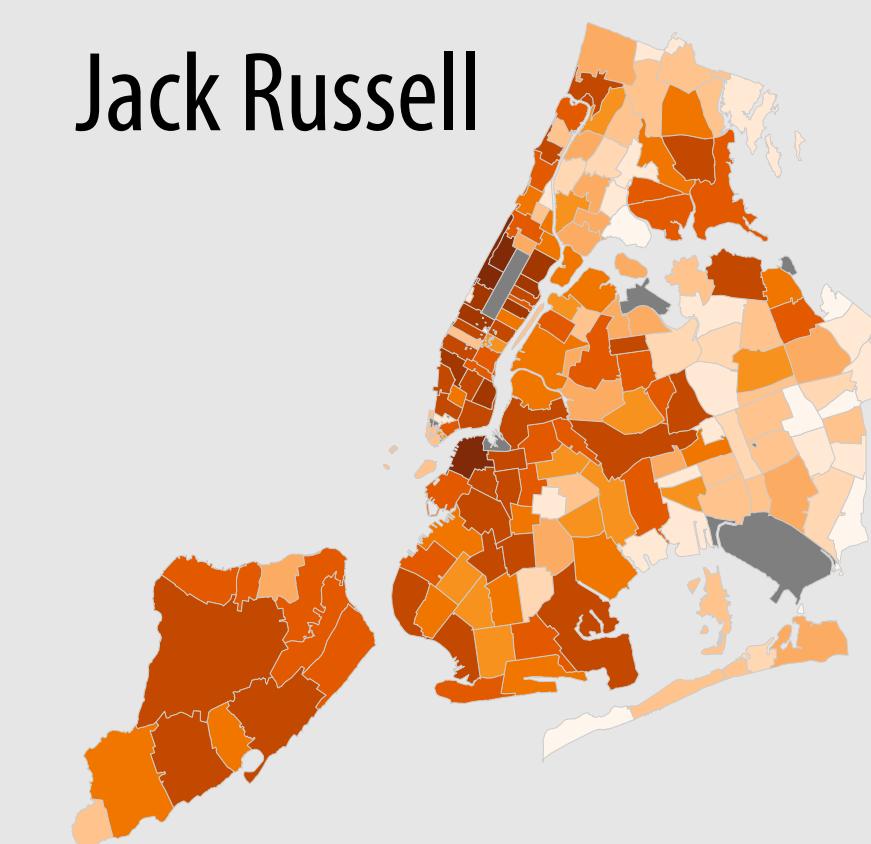
Pomeranian



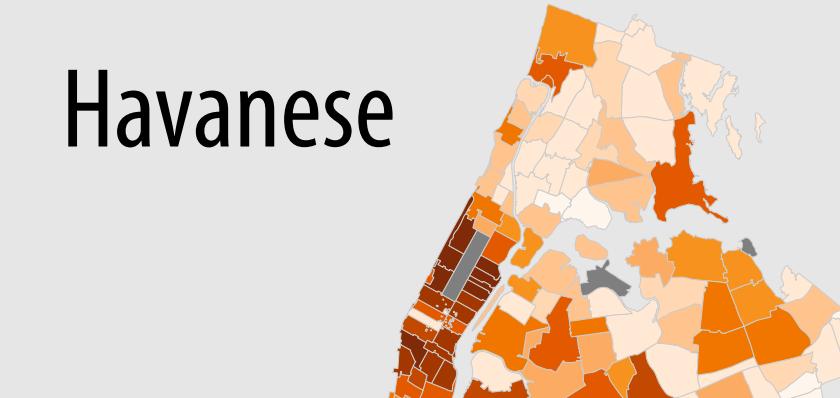
Beagle



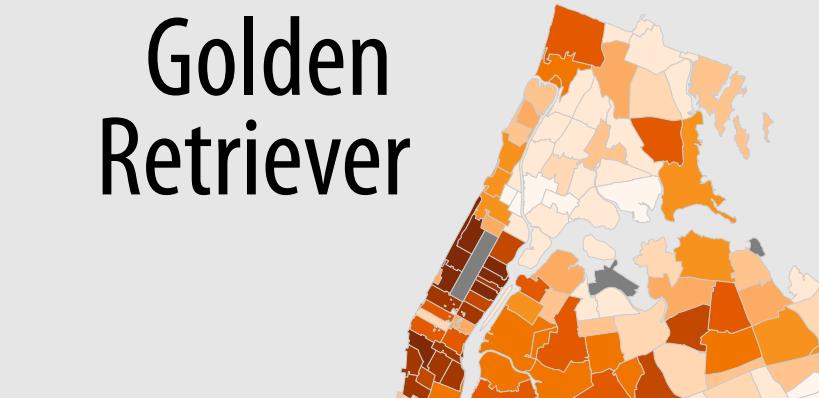
Jack Russell



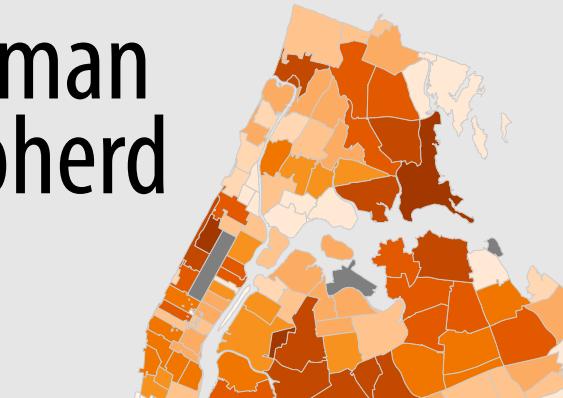
Havanese



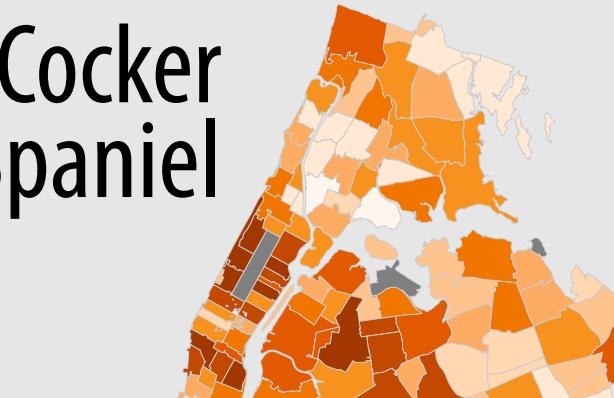
Golden Retriever



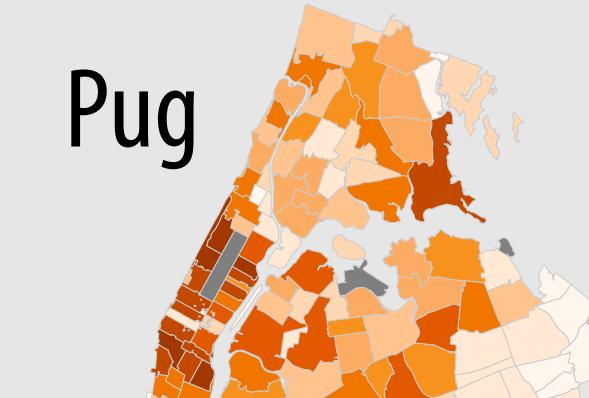
German Shepherd



Cocker Spaniel



Pug



Relative prevalence of breed, from Most to Least

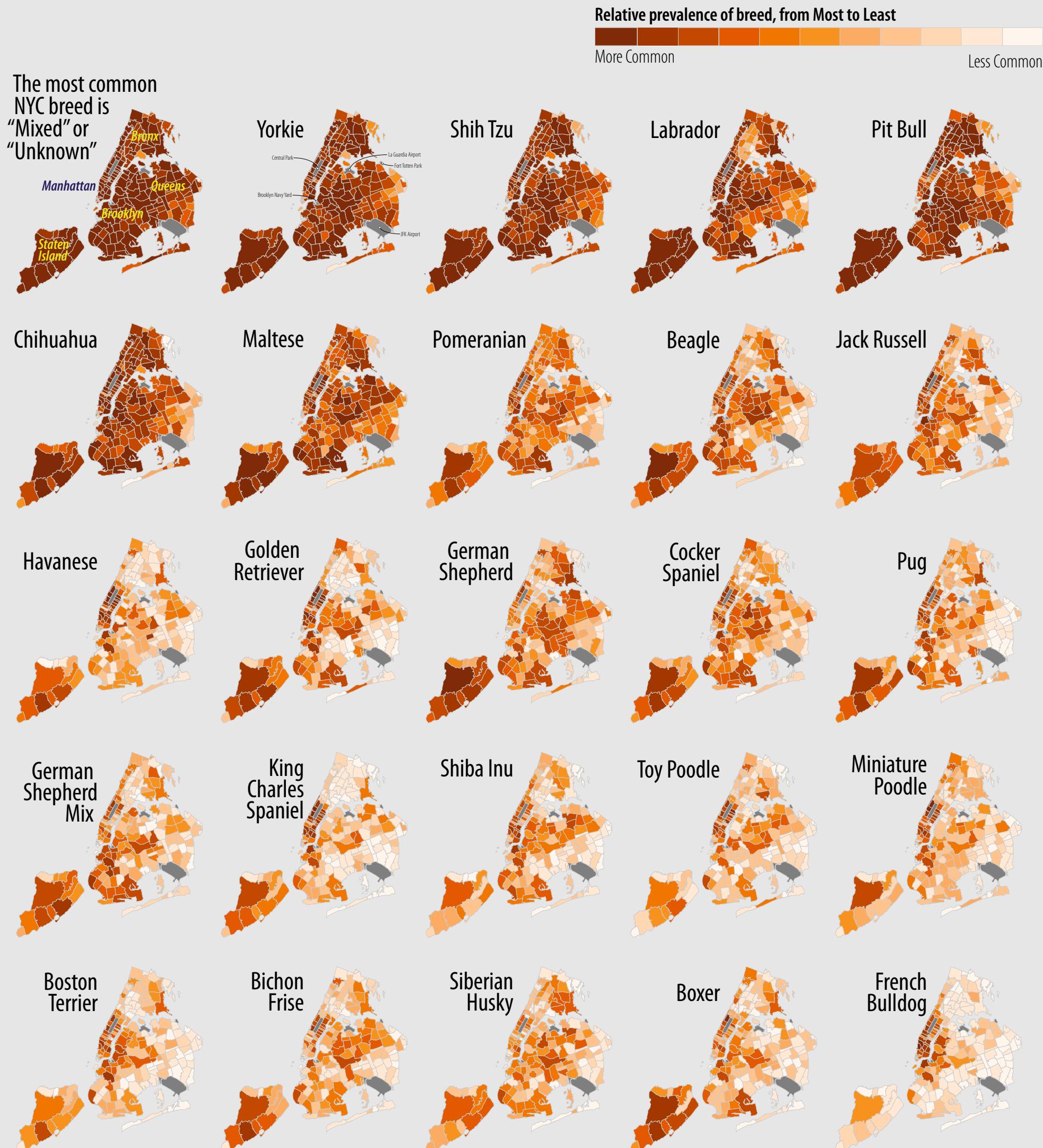


More Common

Less Common

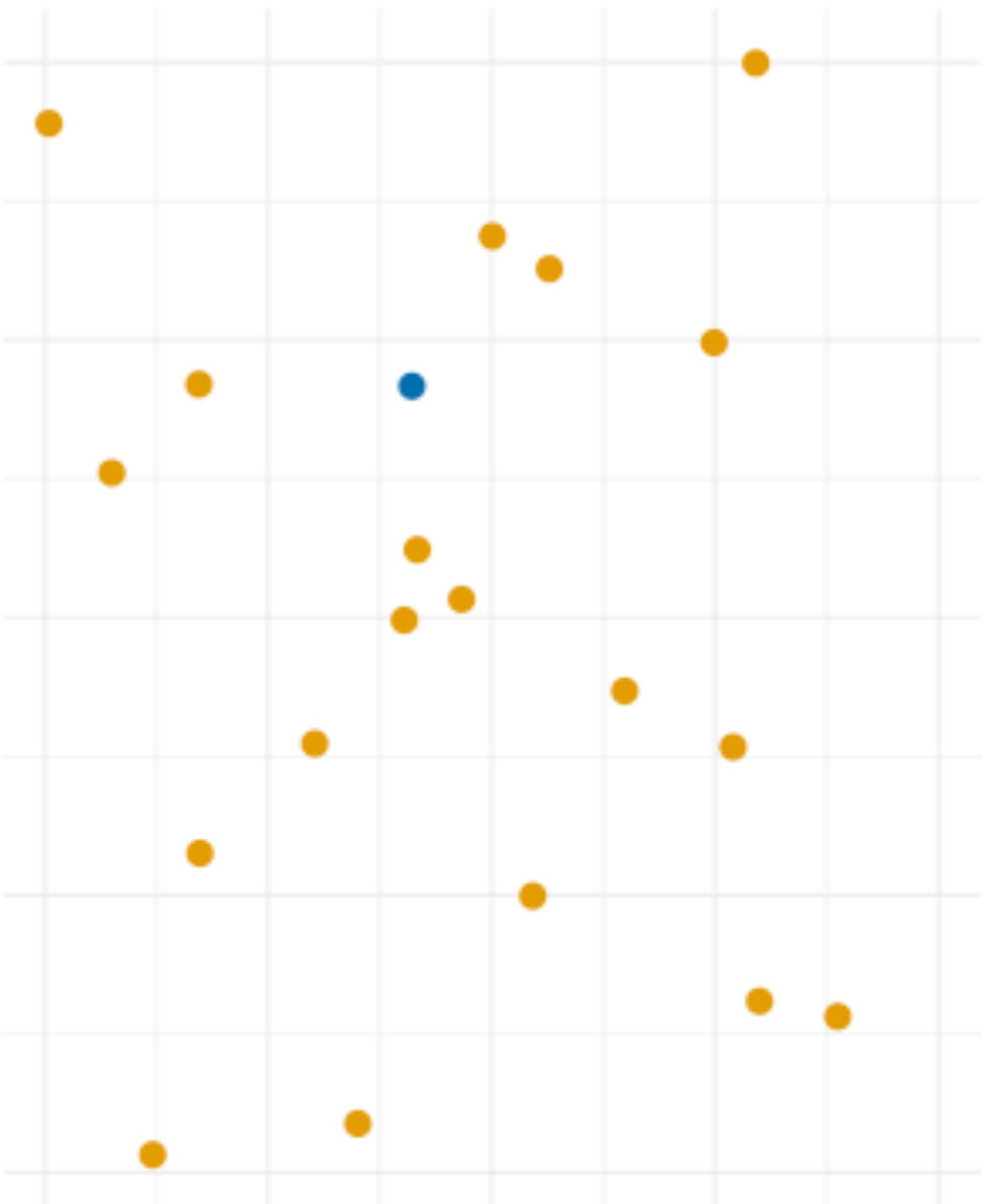
Dogs of New York

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.

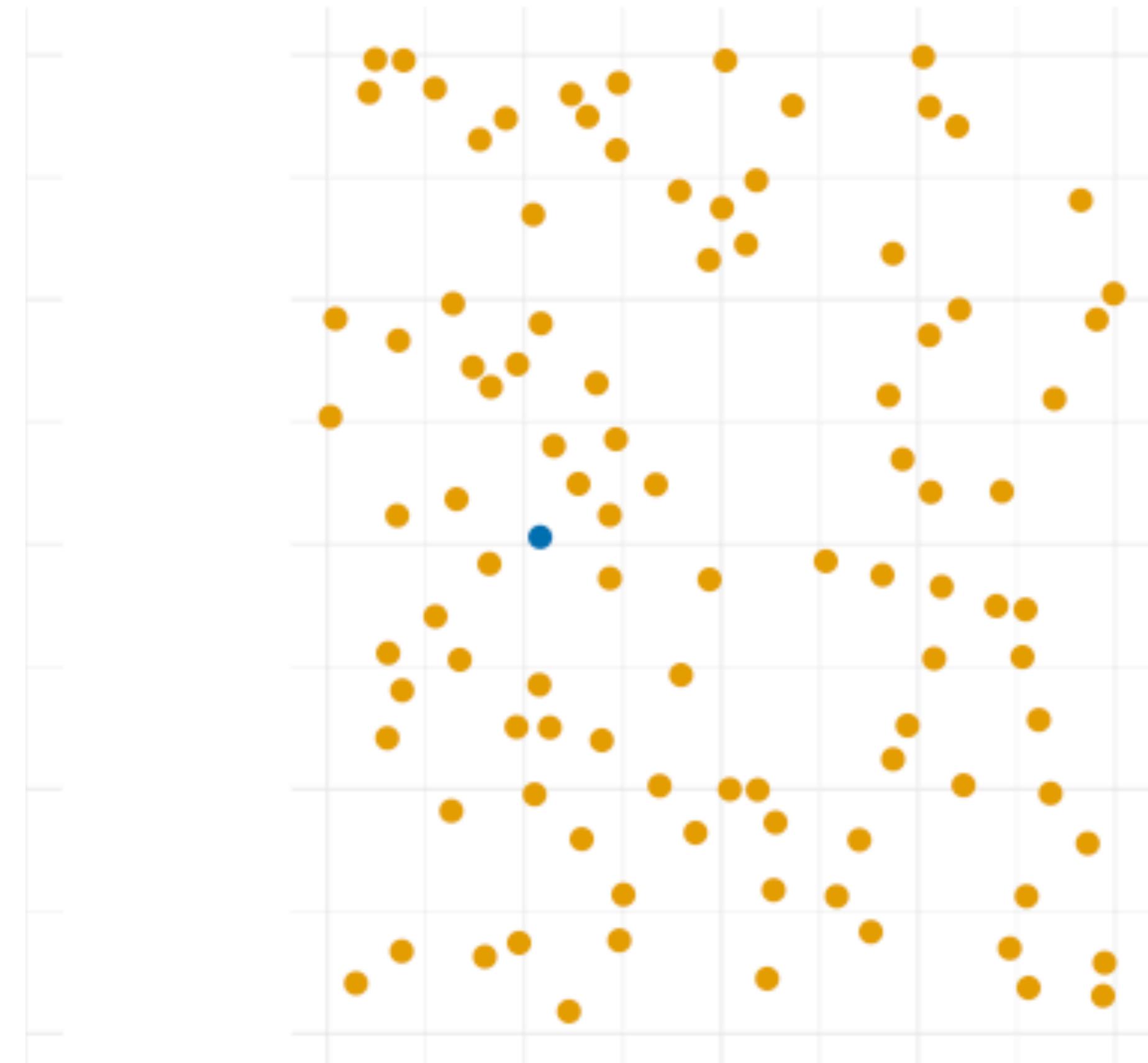


PRE-ATTENTIVE PROCESSING

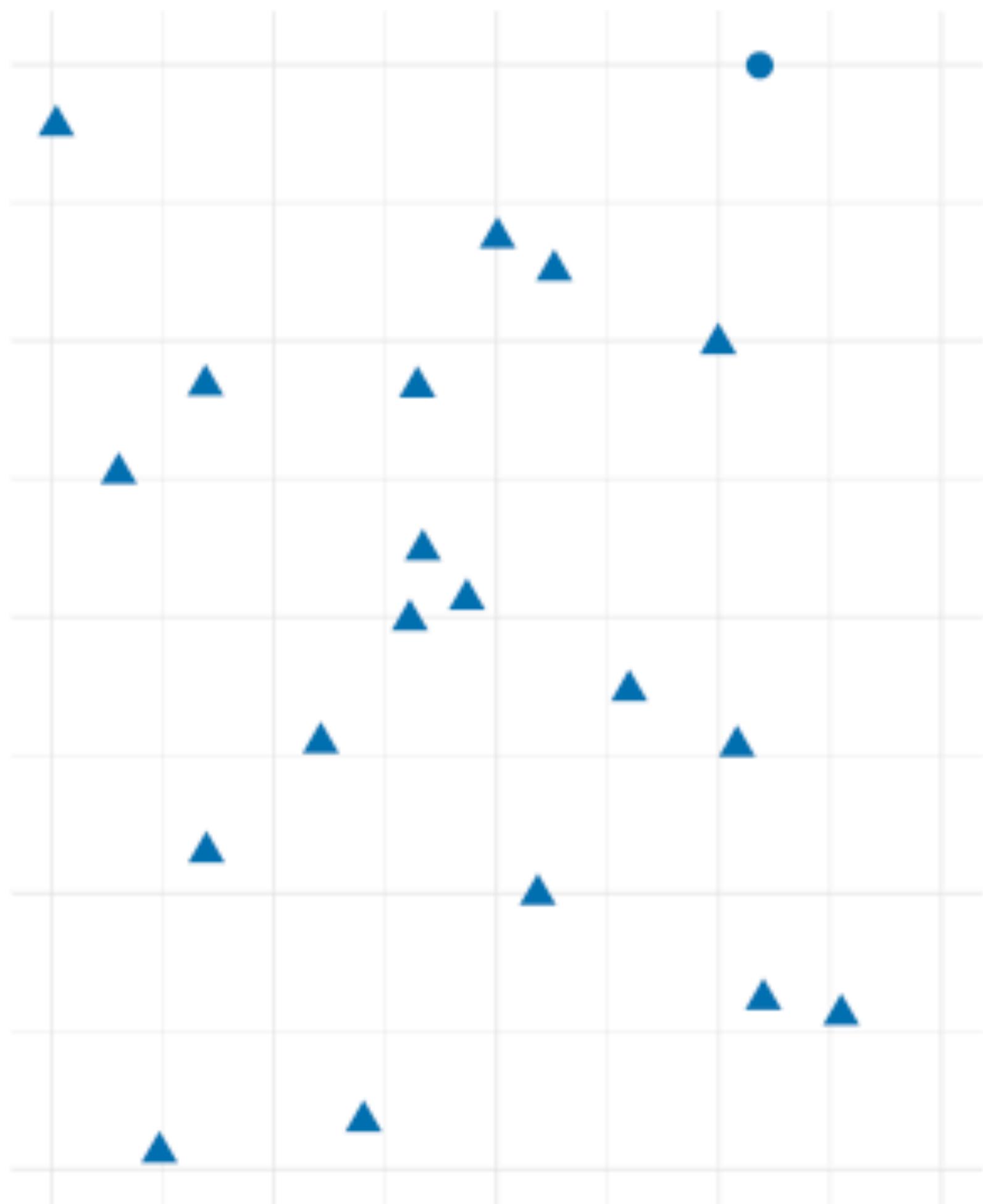
Color Only, N=20



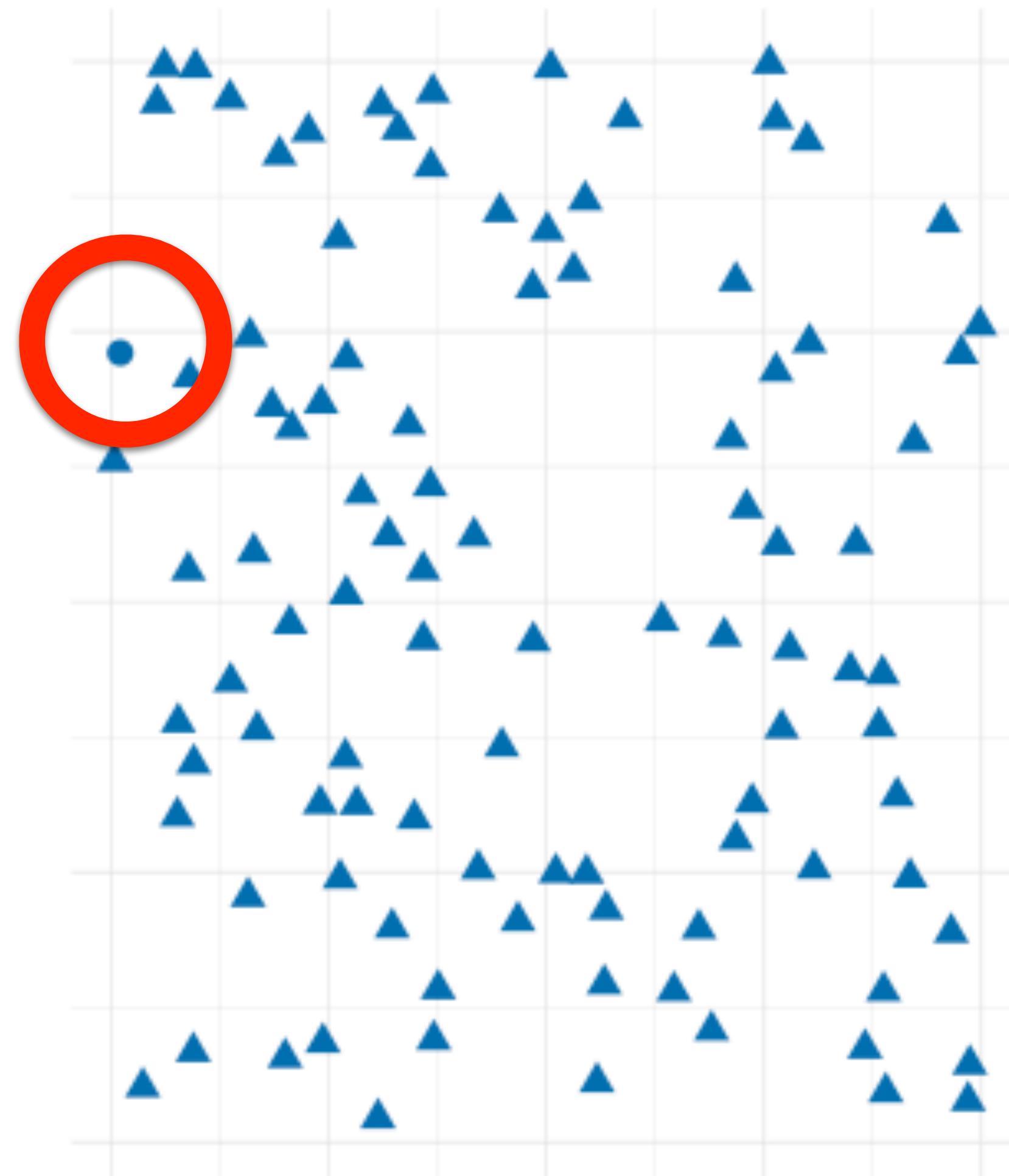
Color Only, N=100



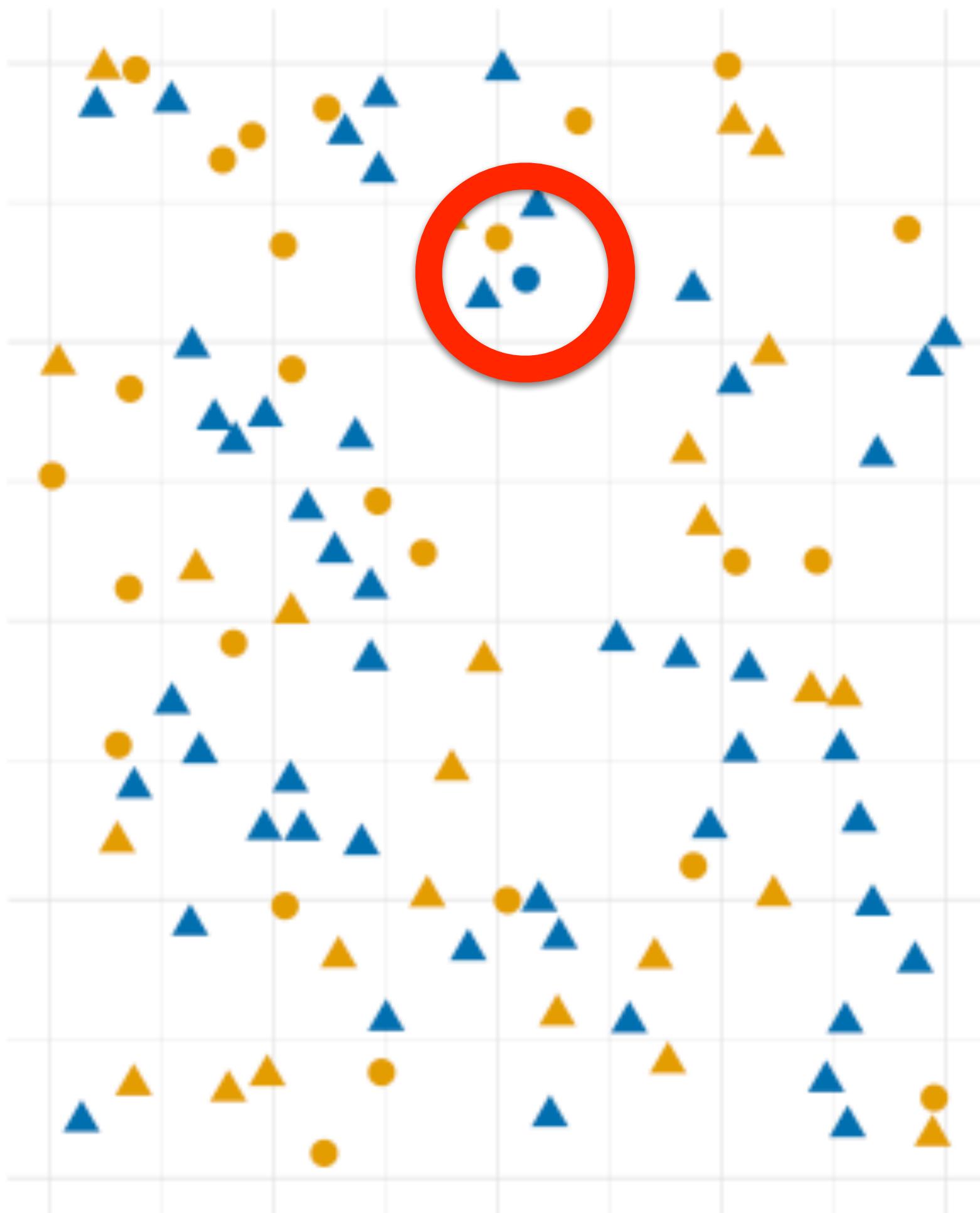
Shape Only, N=20

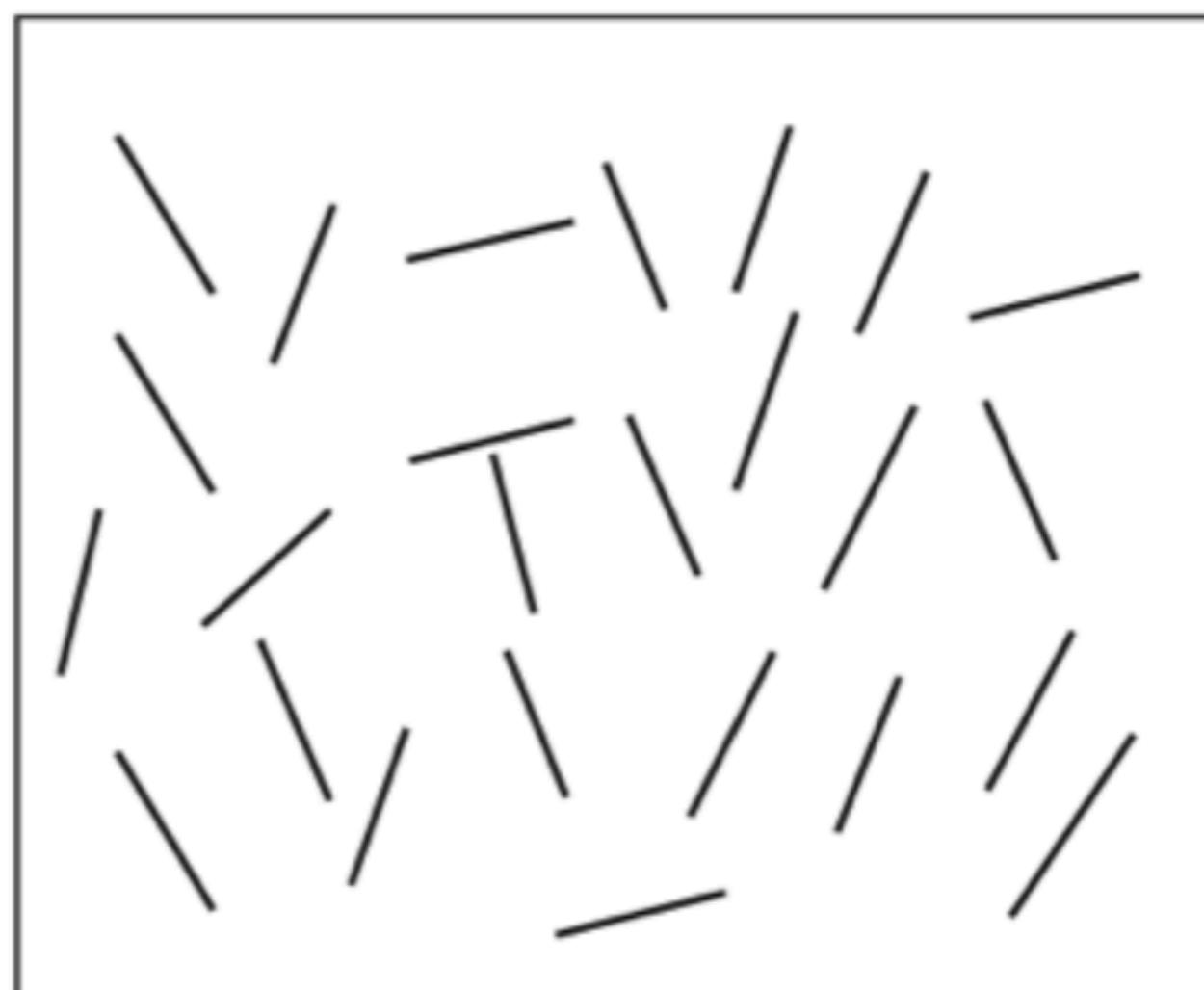
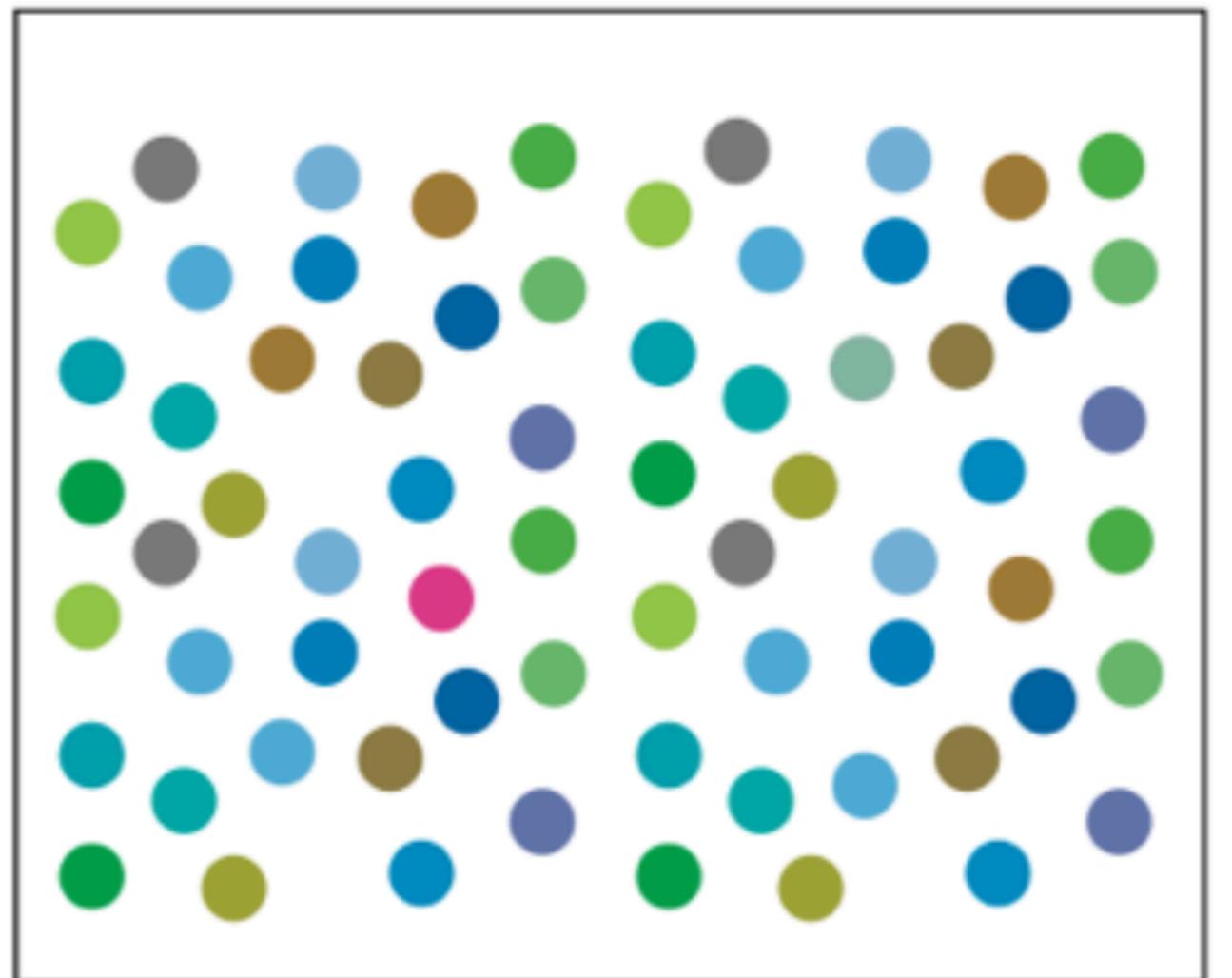
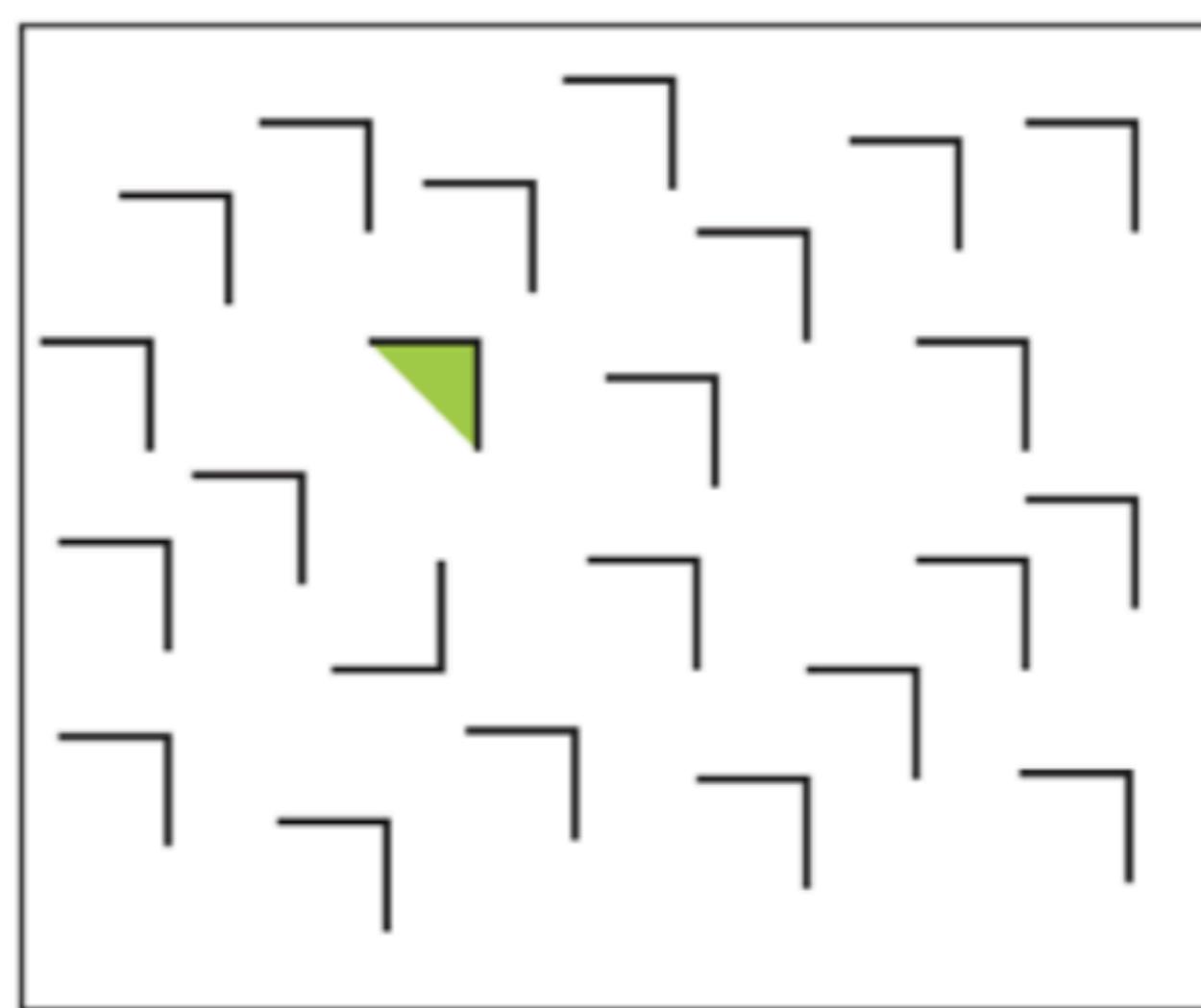
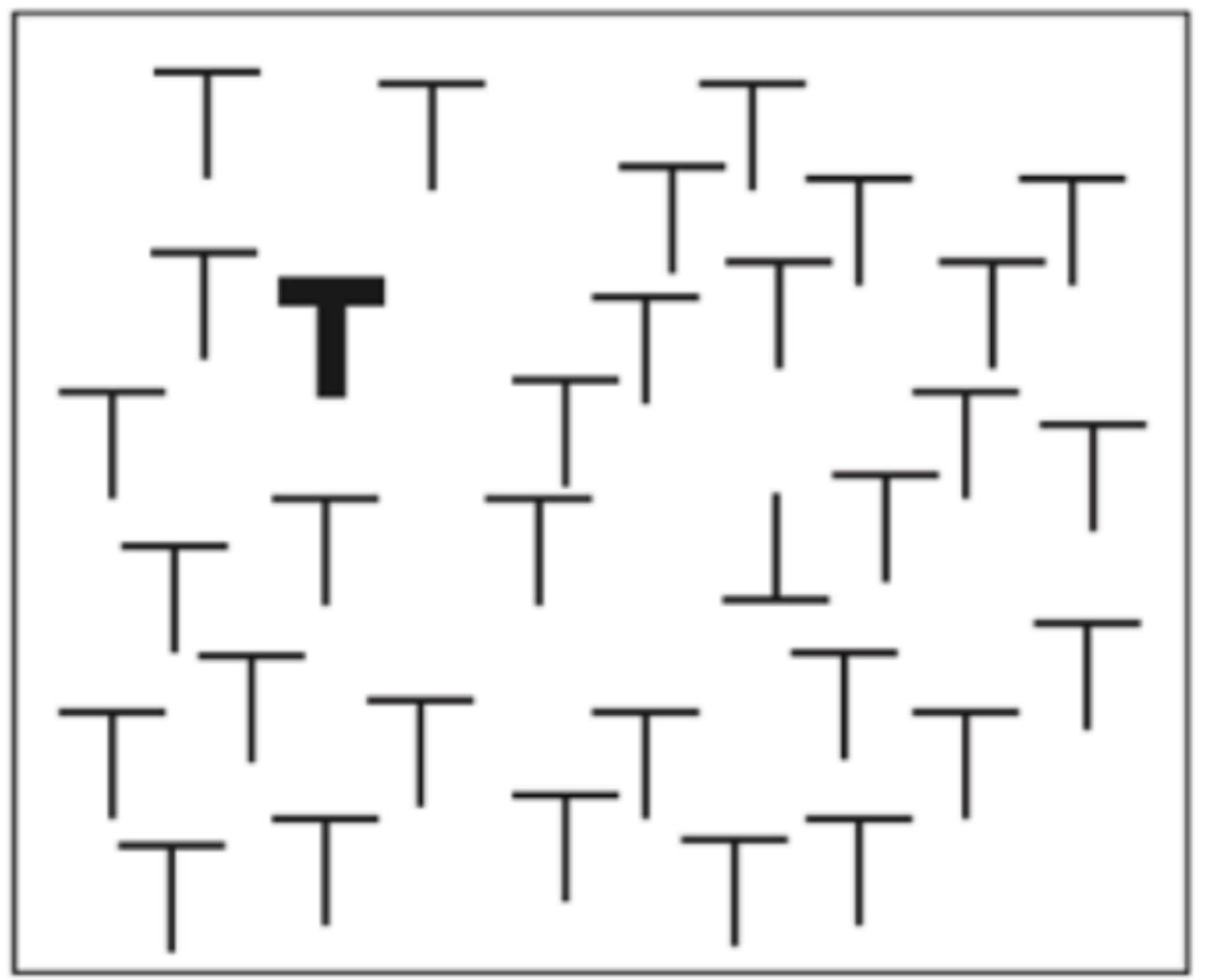


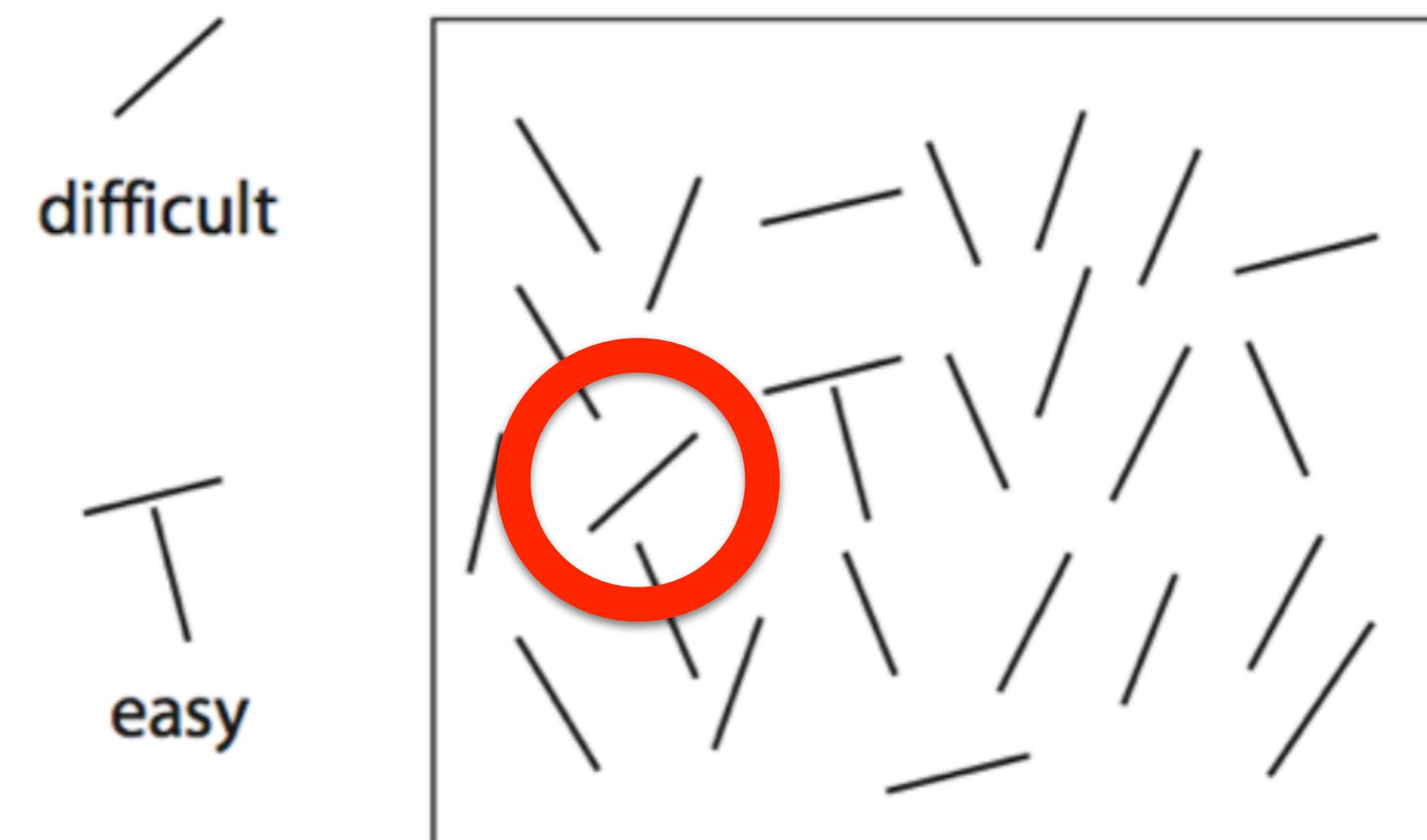
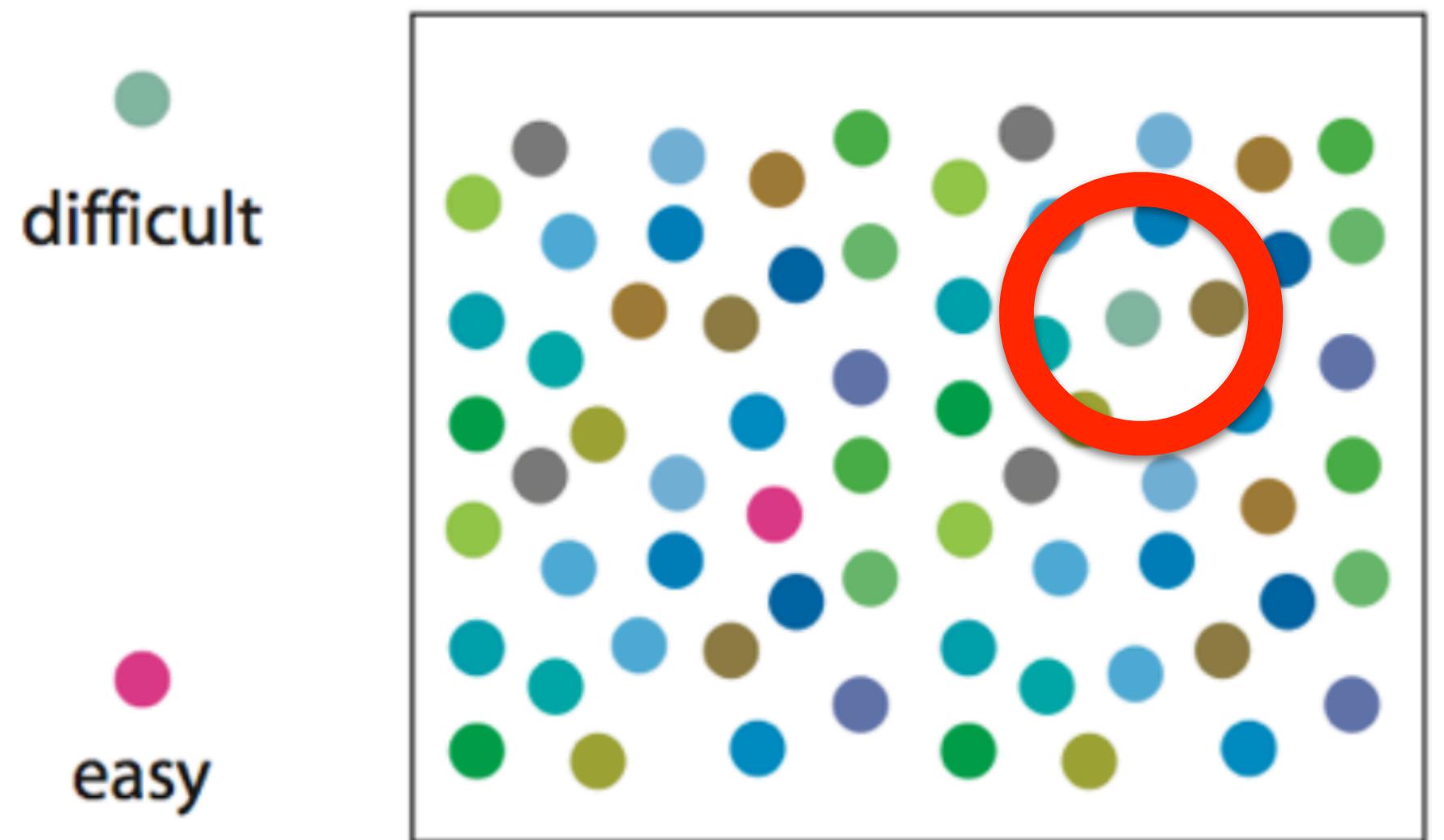
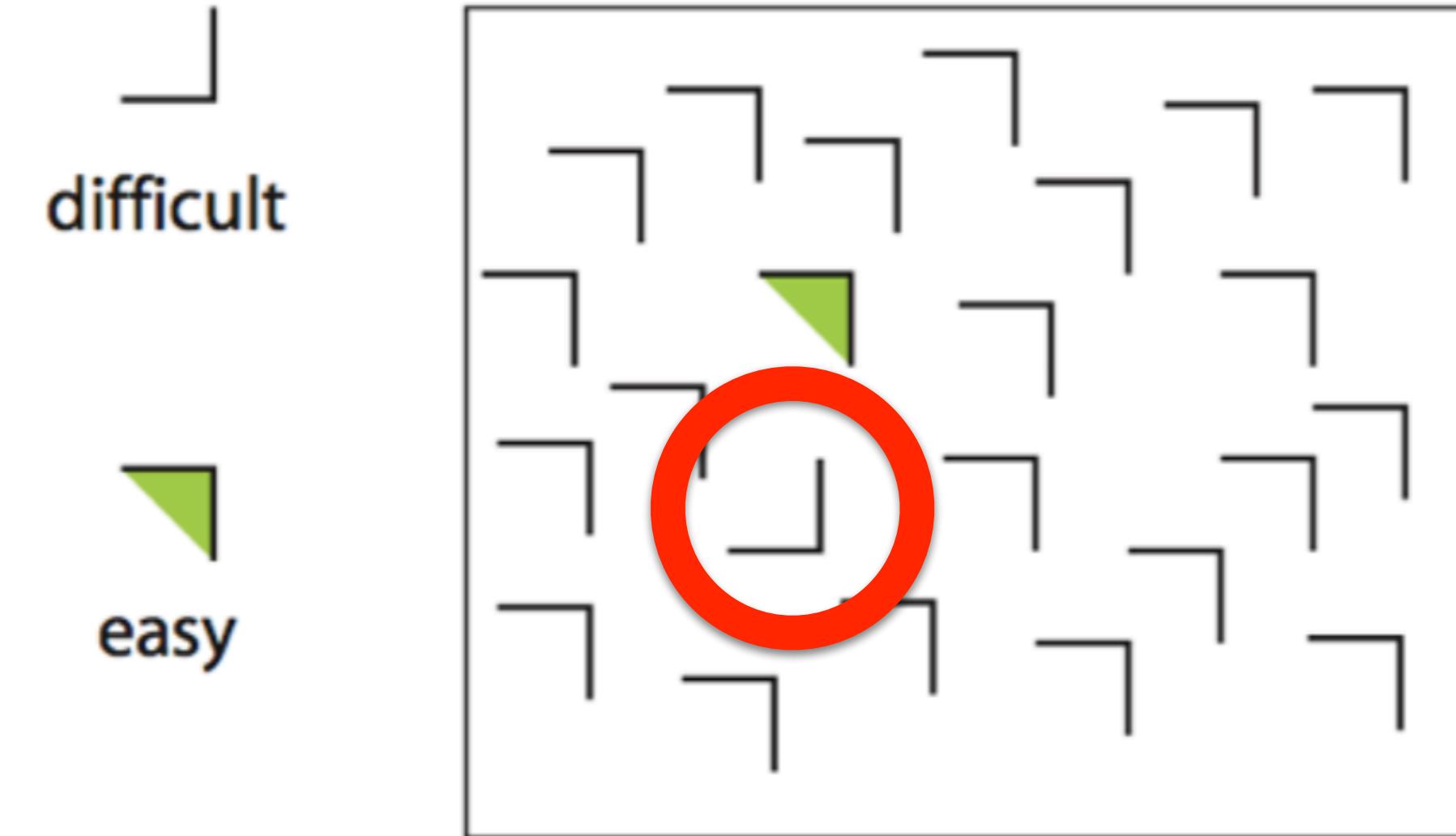
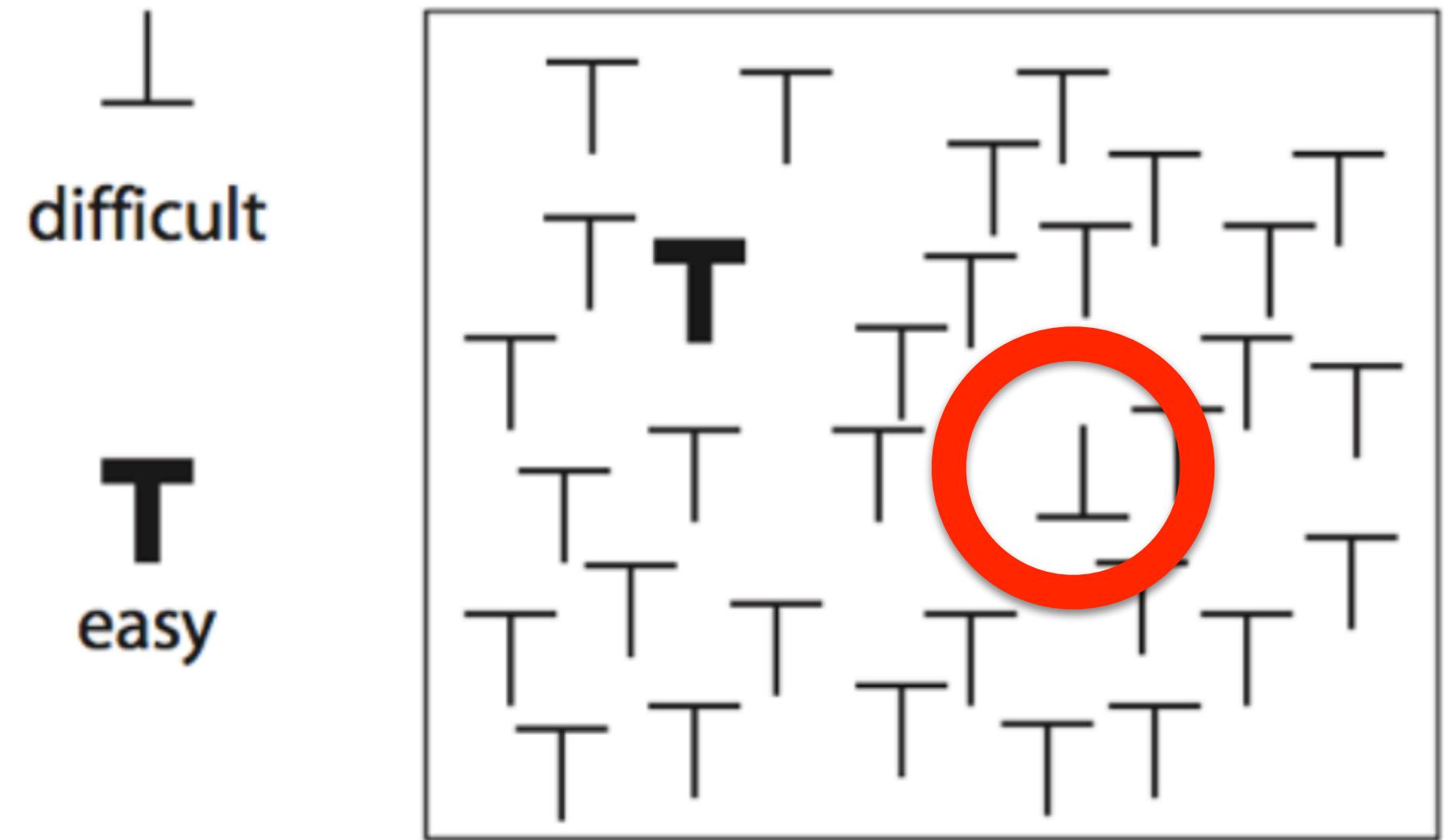
Shape Only, N=100

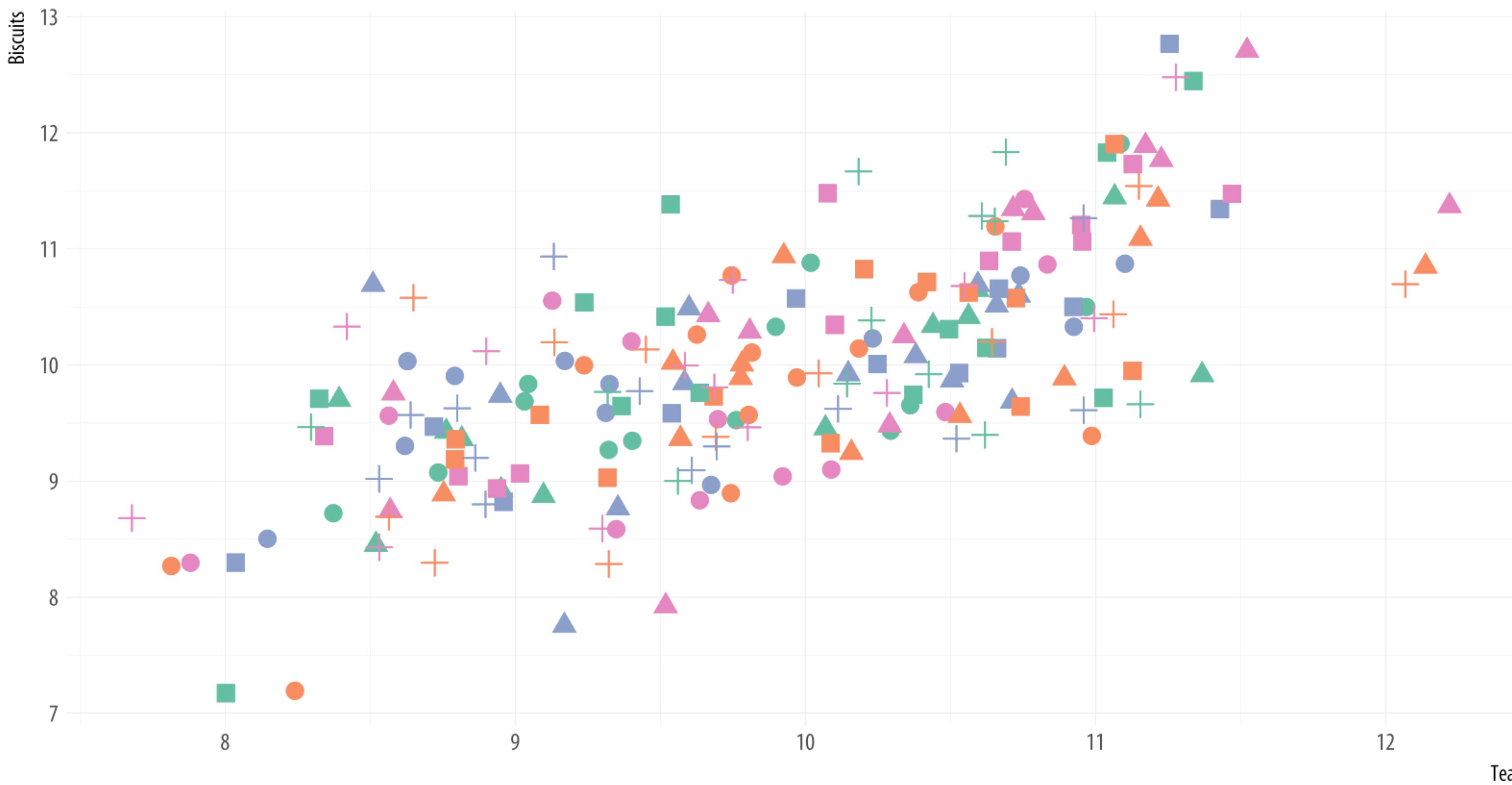


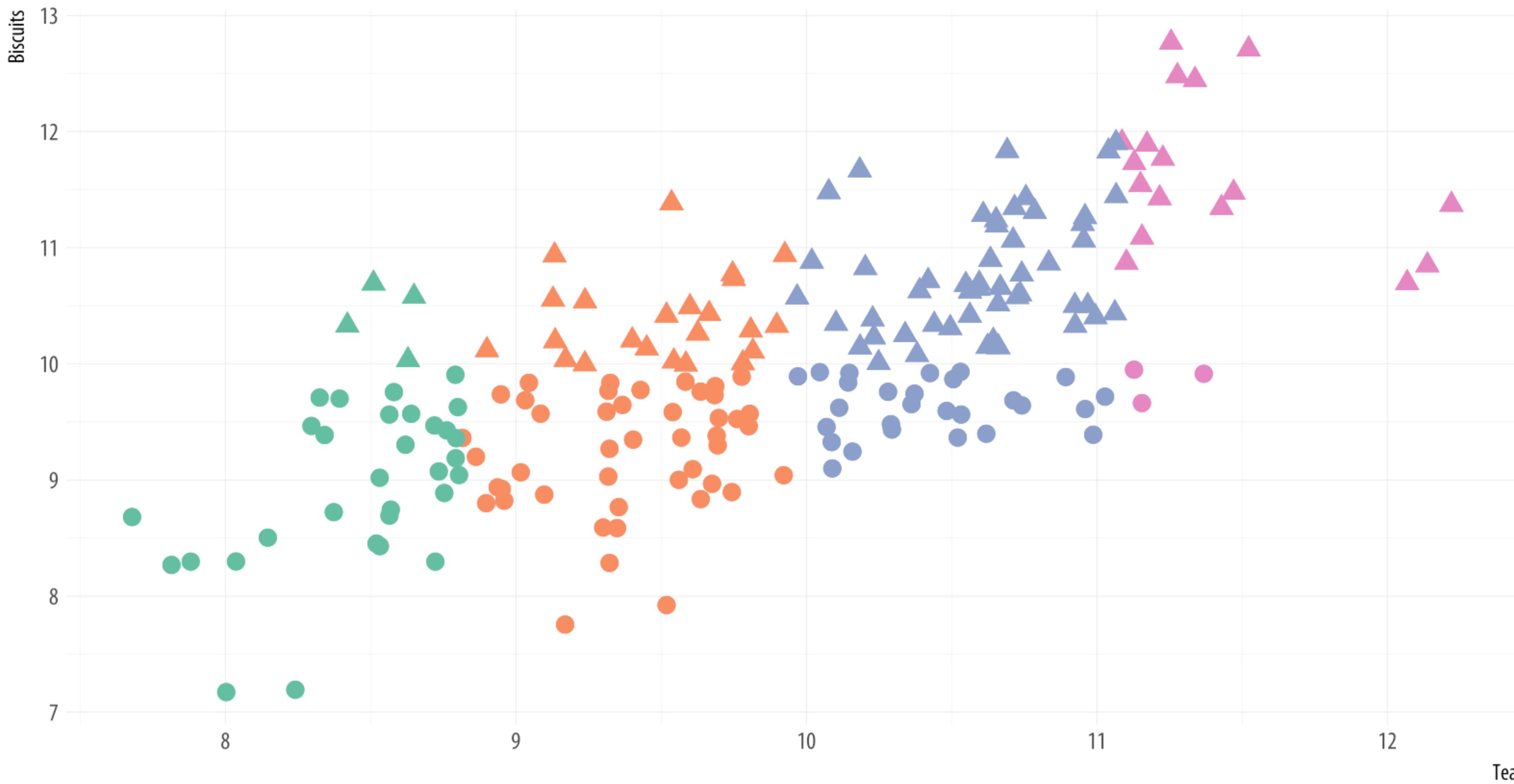
Color & Shape, N=100



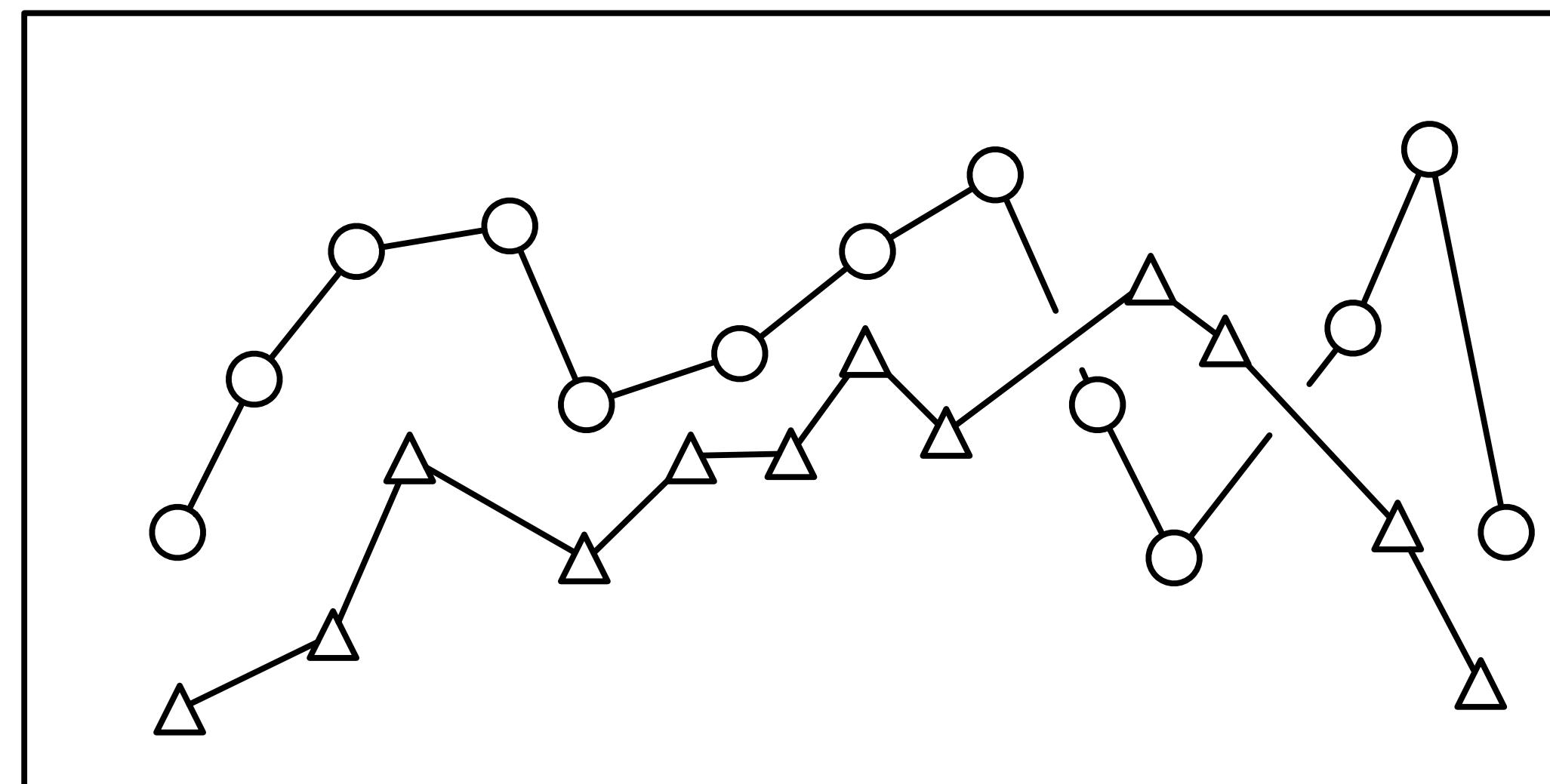
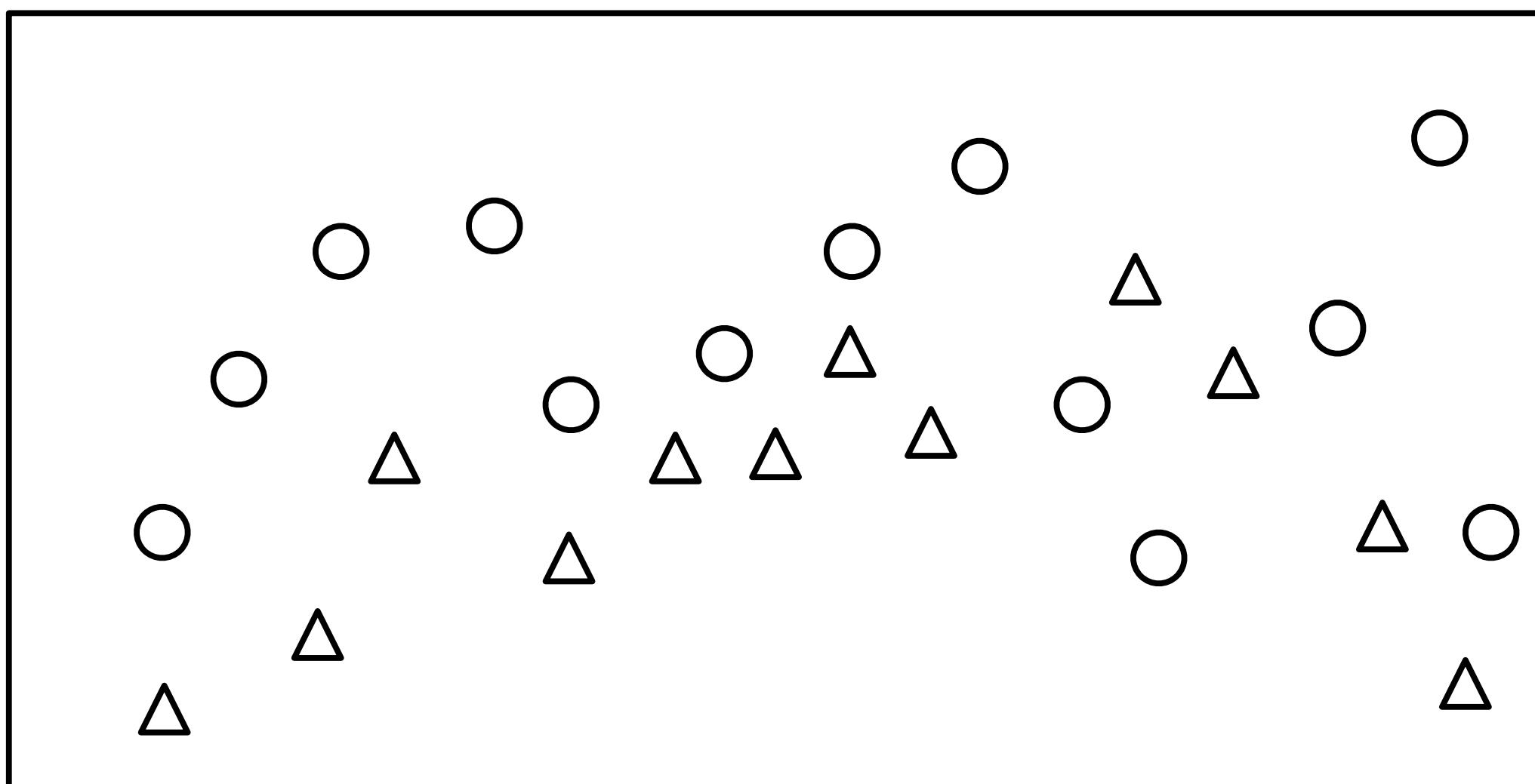
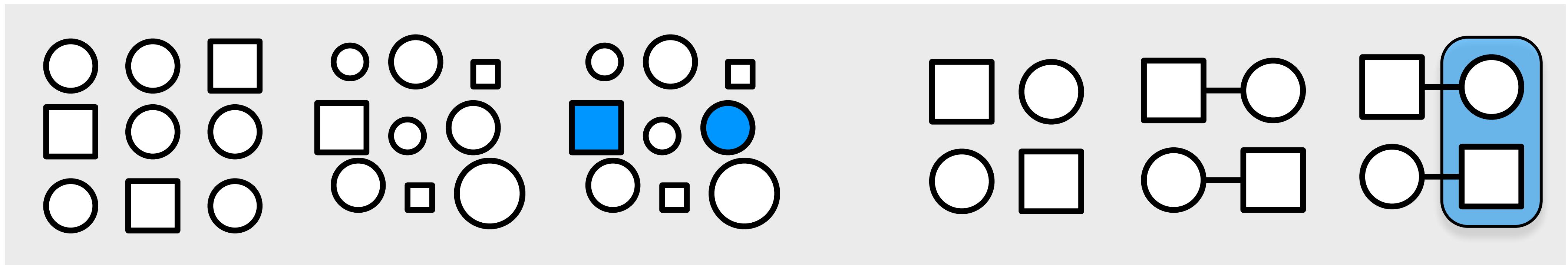
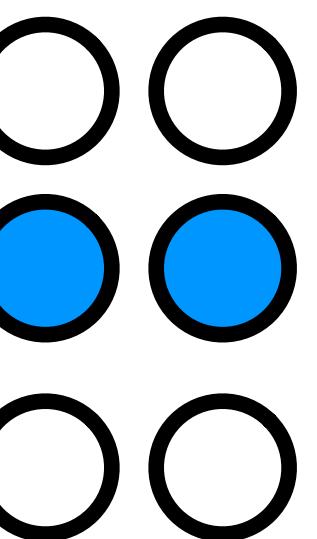
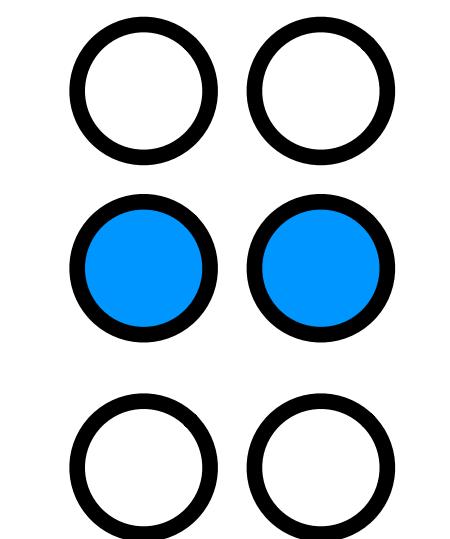
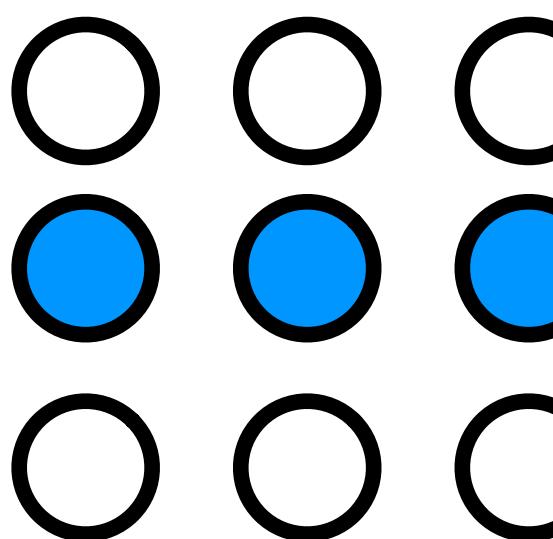
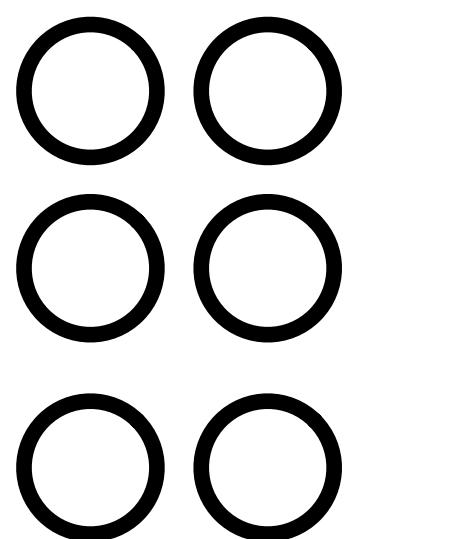
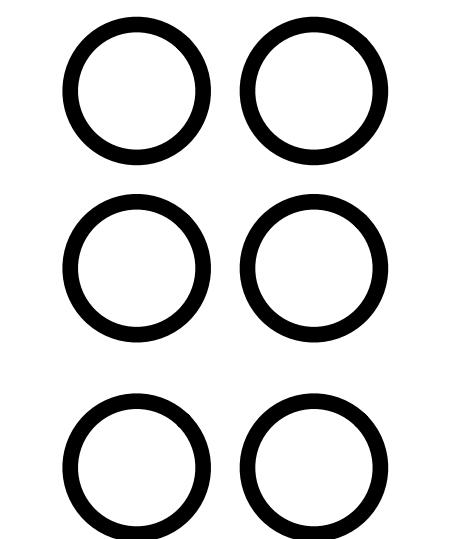
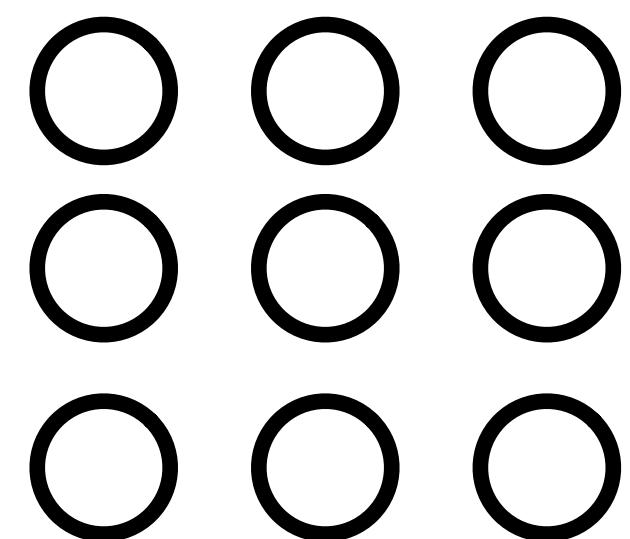




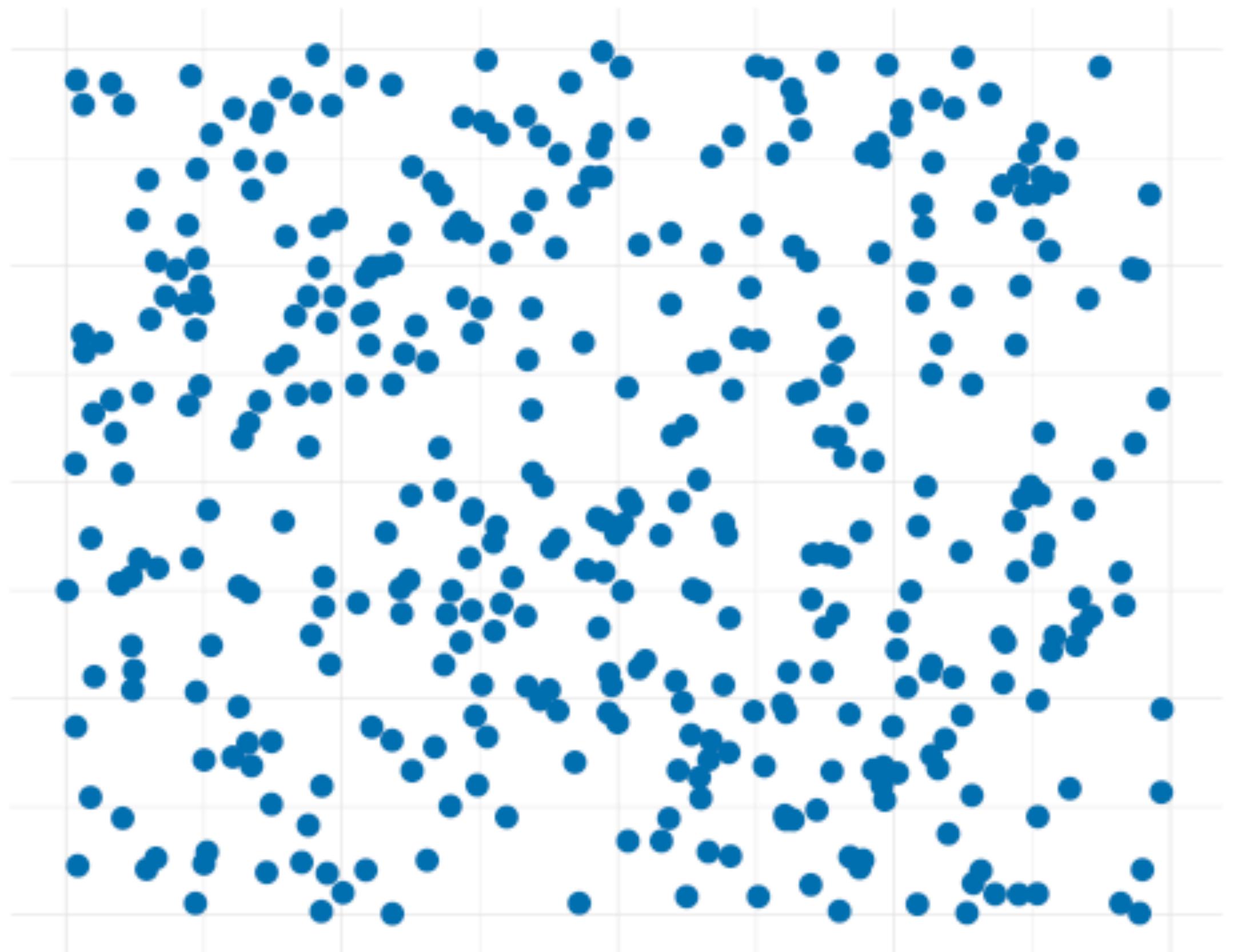




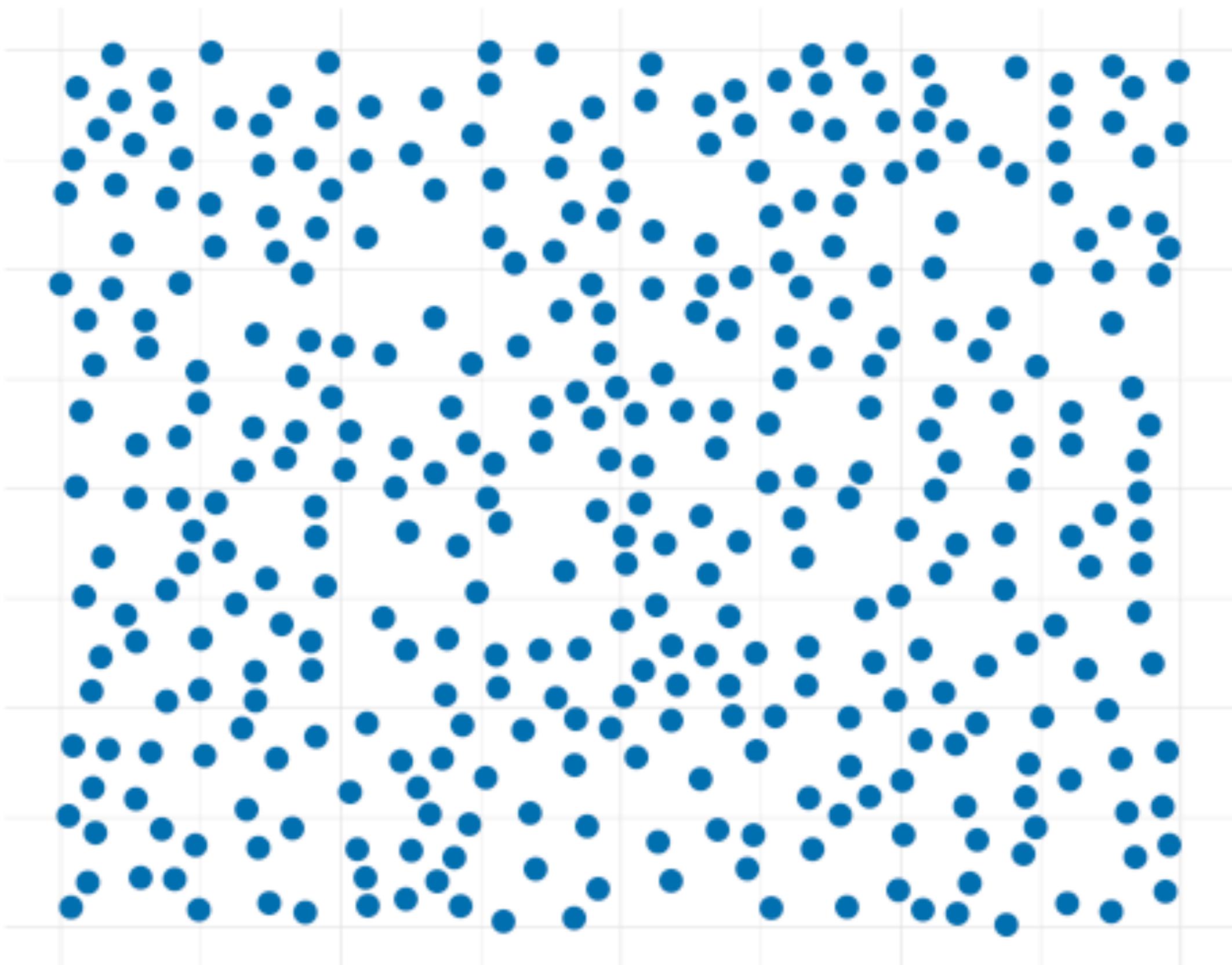
GESTALT INFERENCES



Poisson



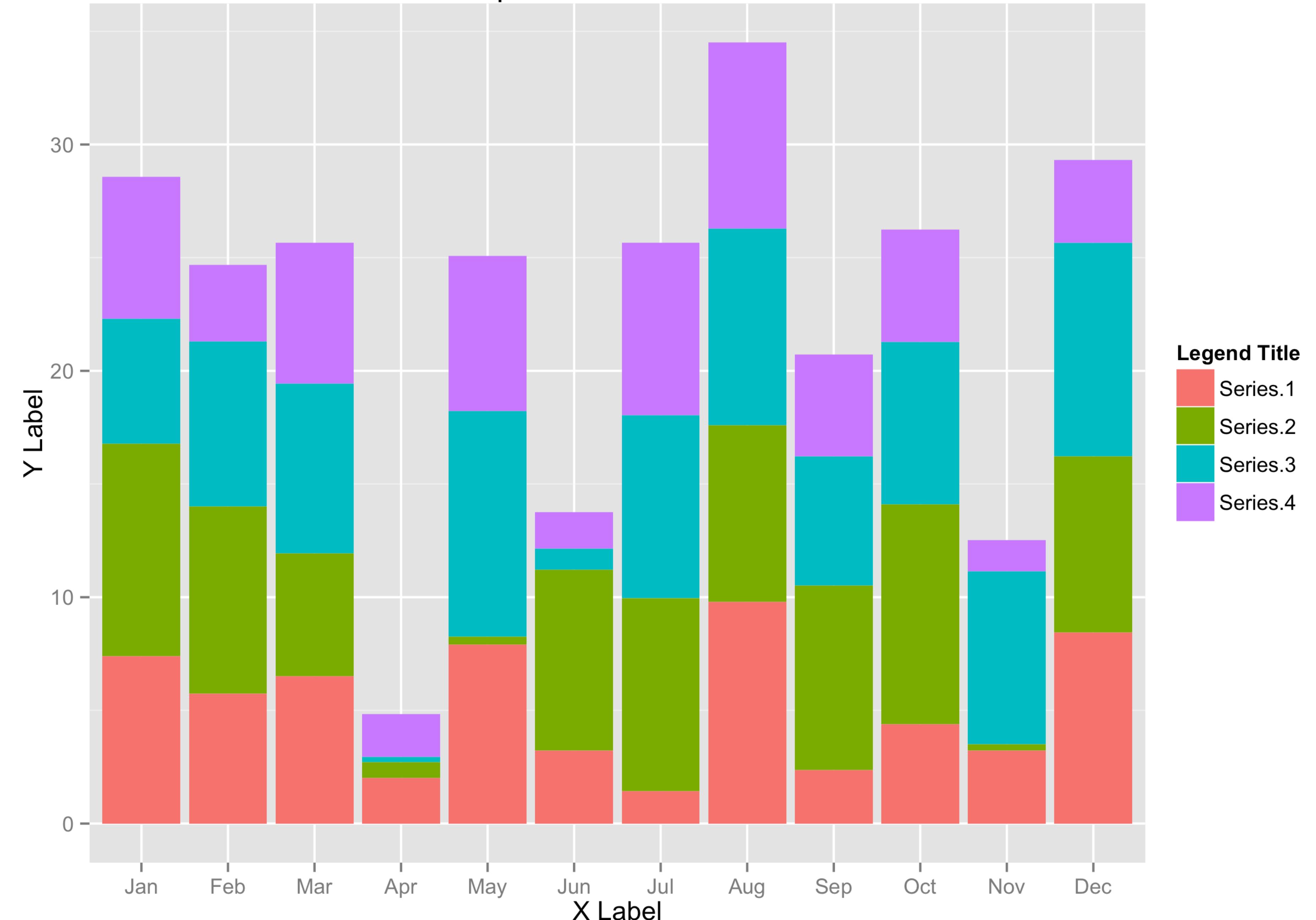
Matérn

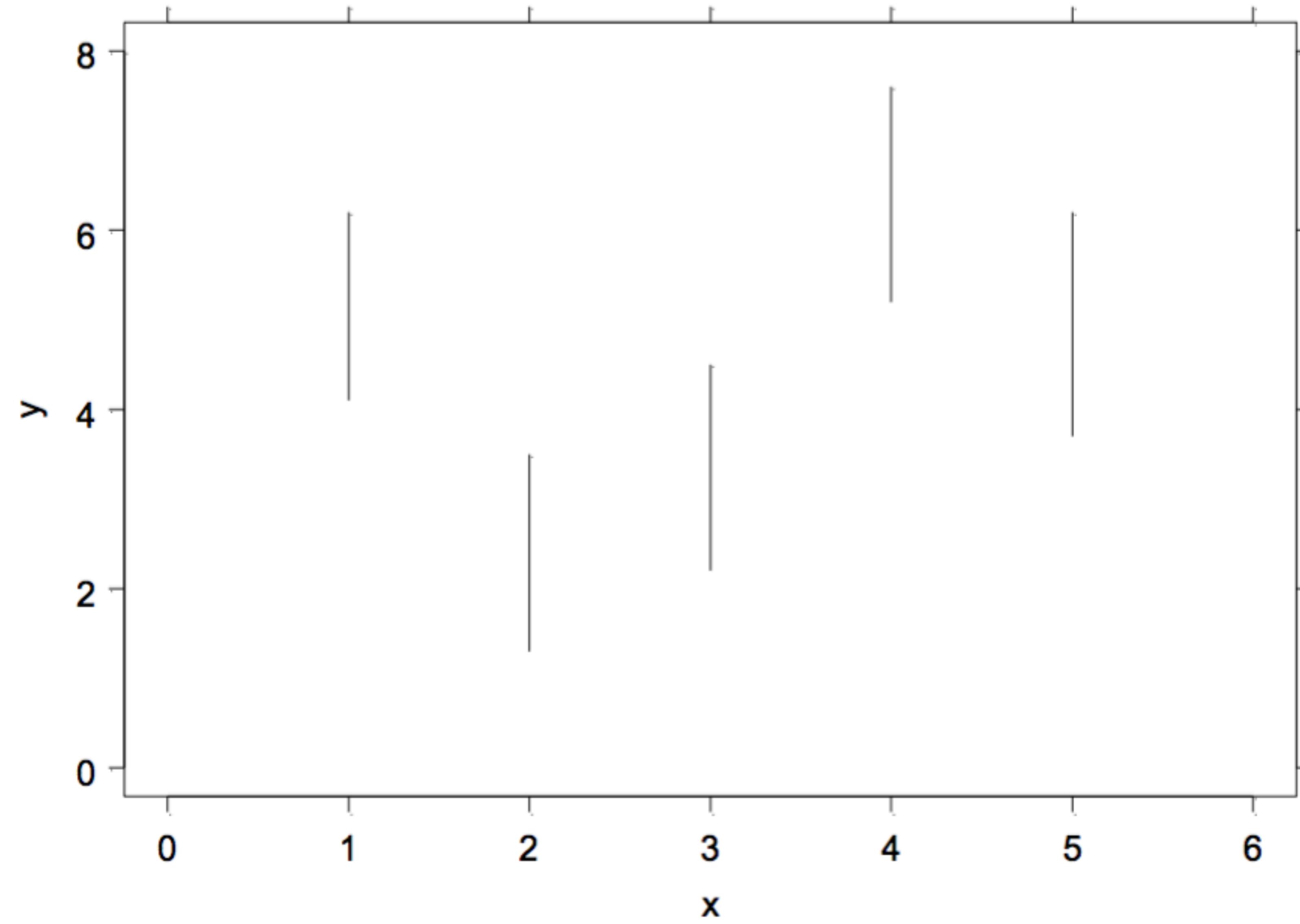


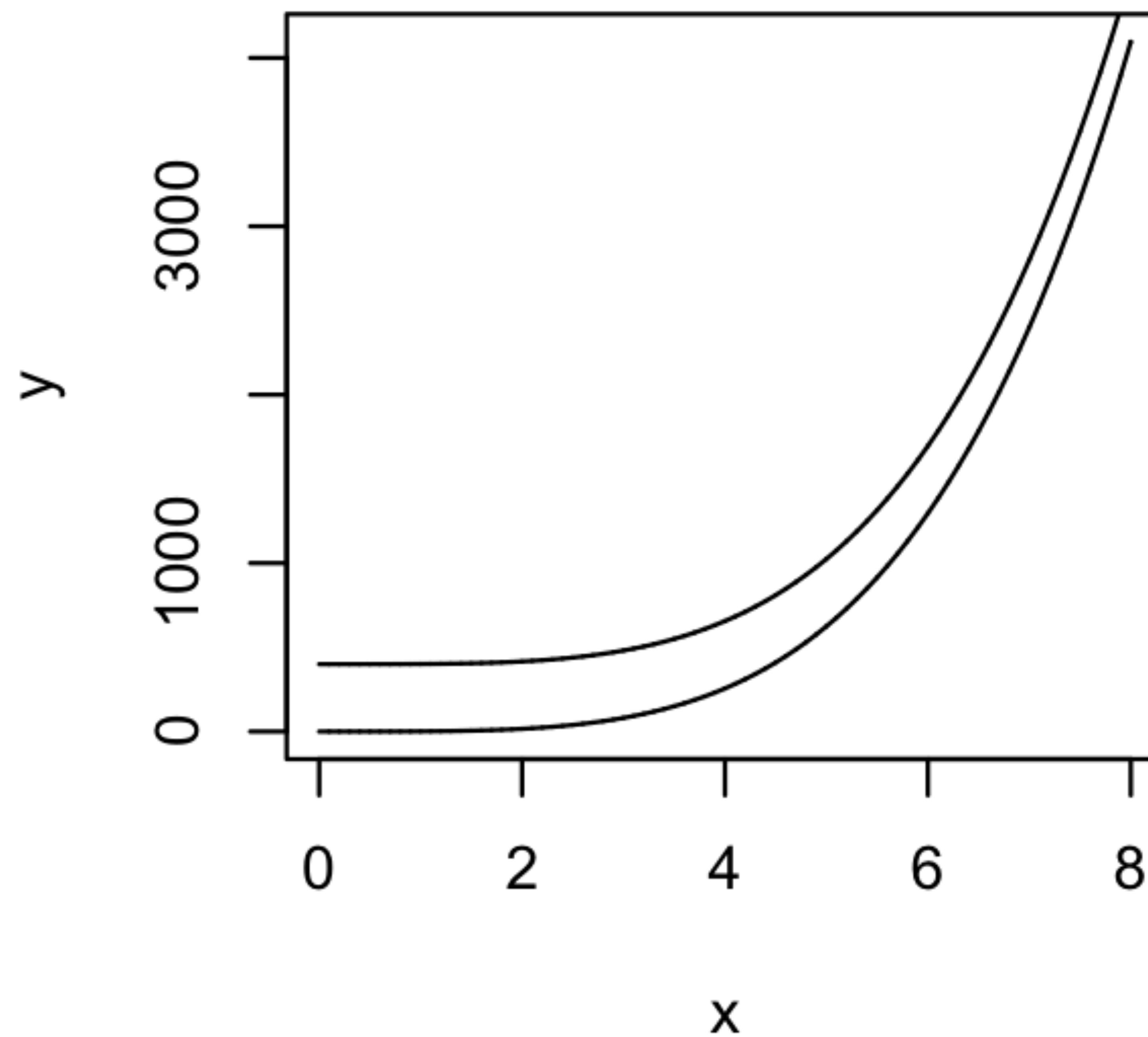
MAPPINGS FOR DATA

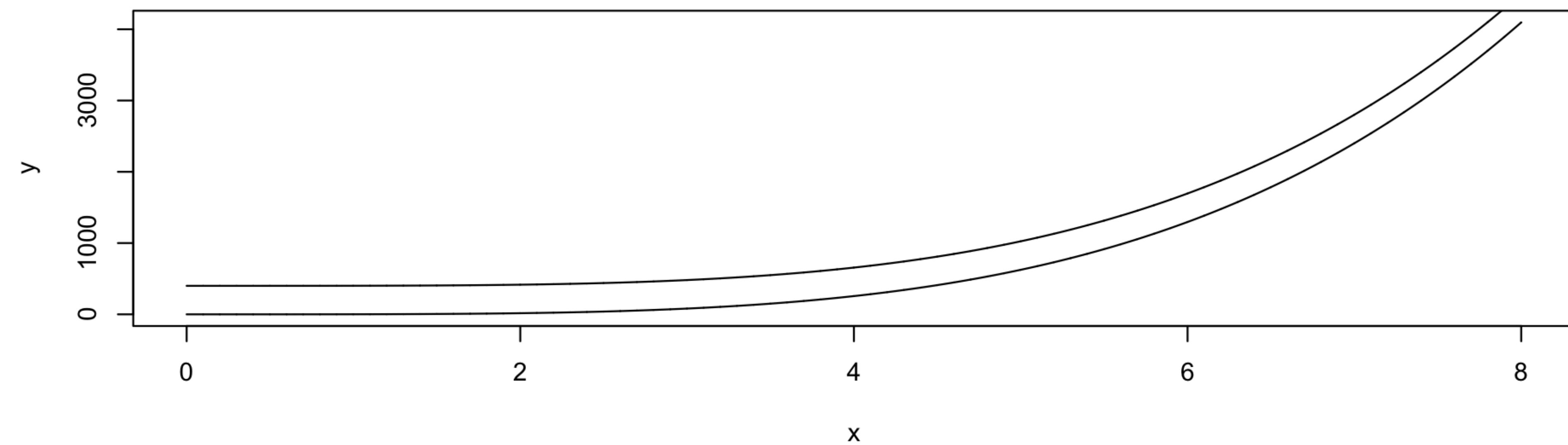
Visual Tasks in Decoding Graphs

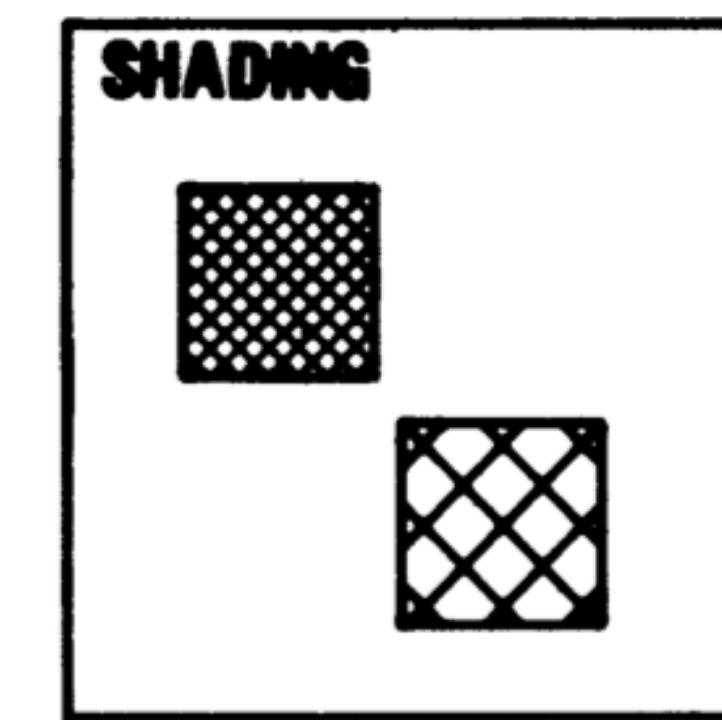
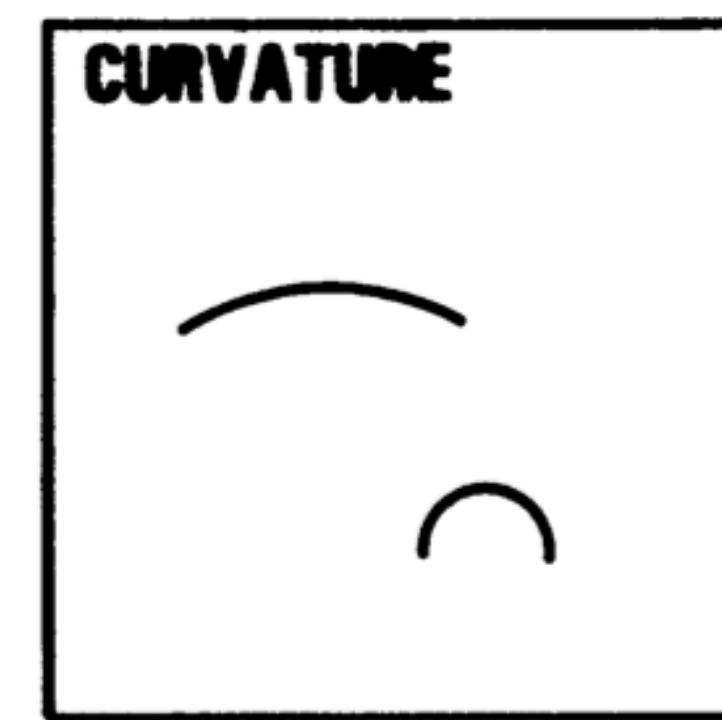
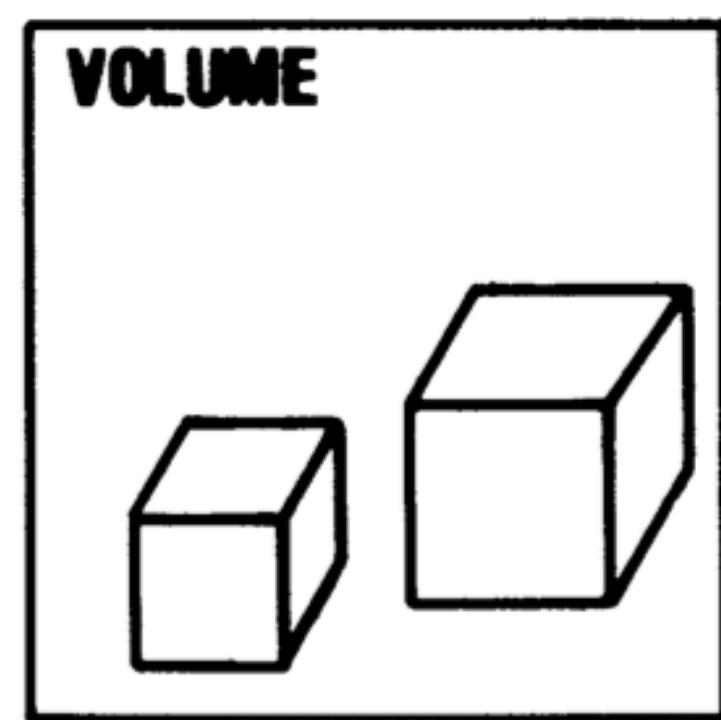
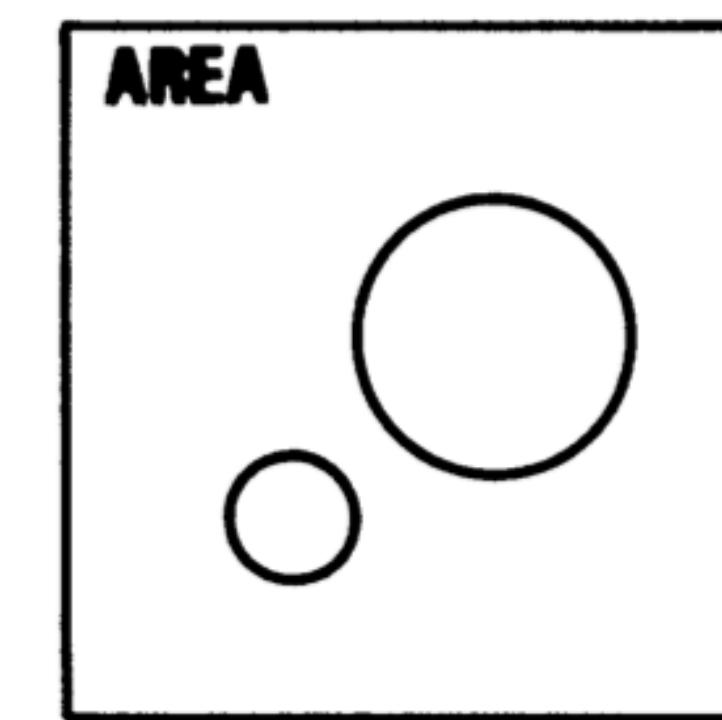
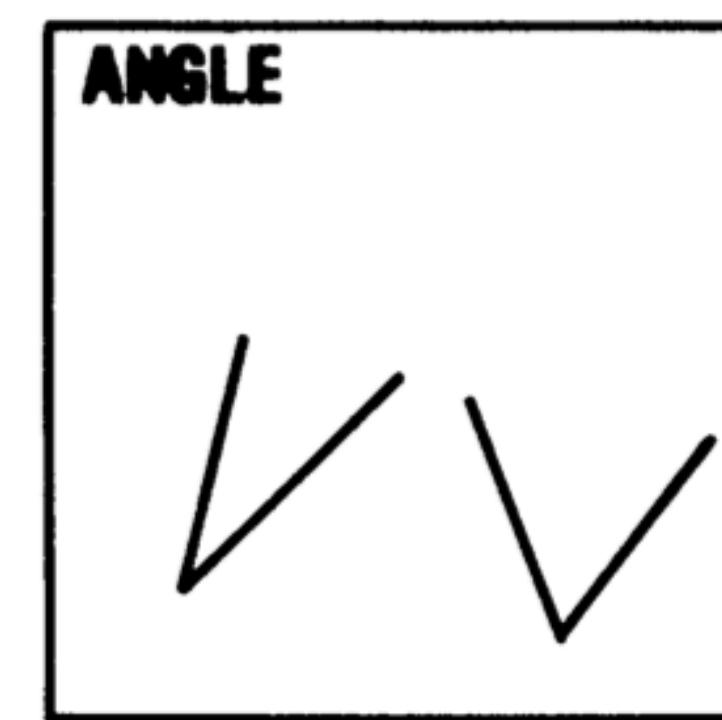
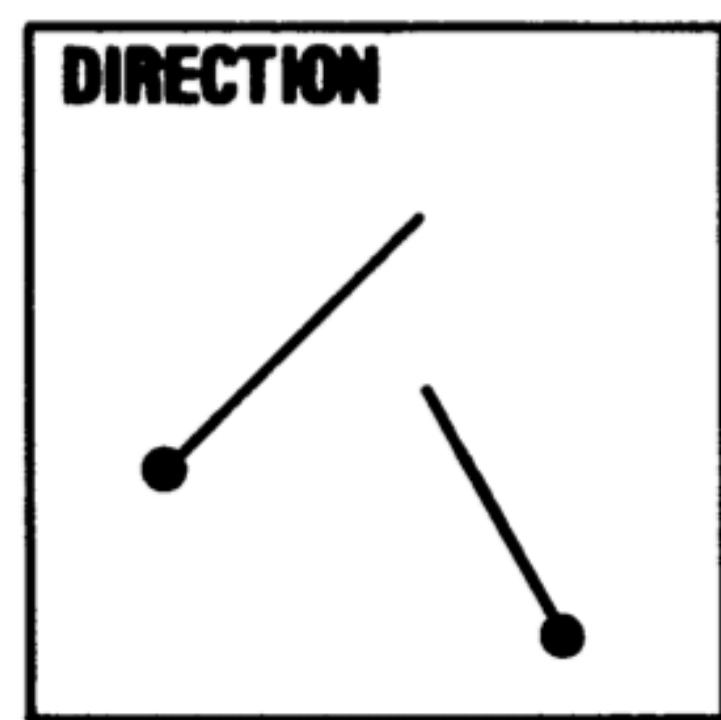
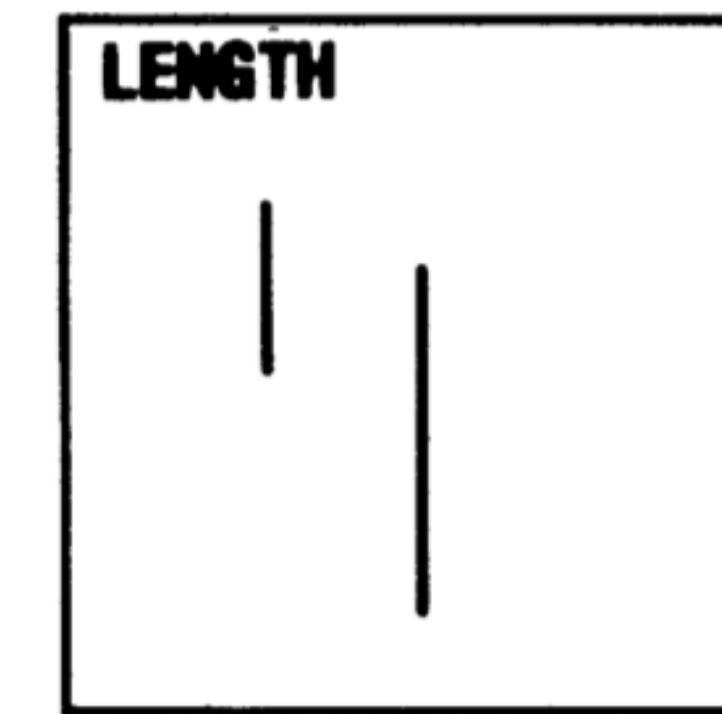
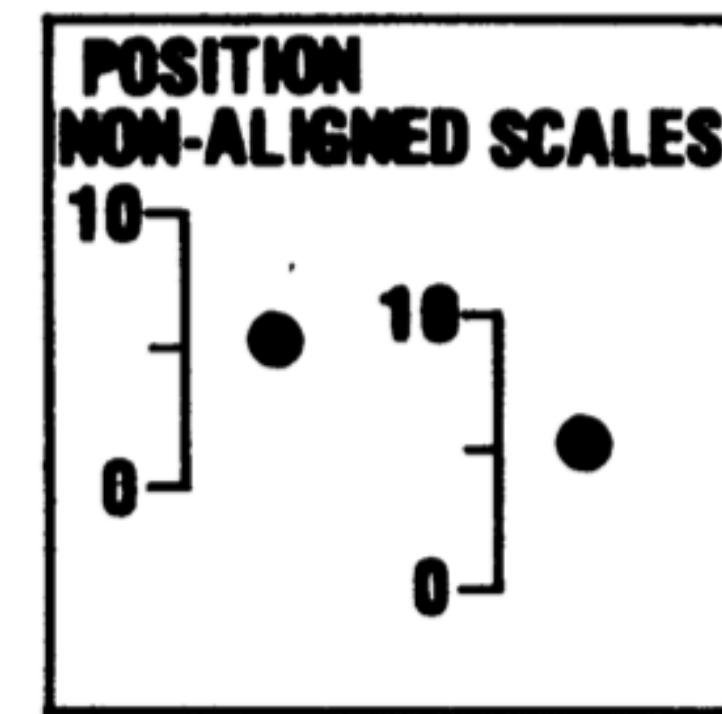
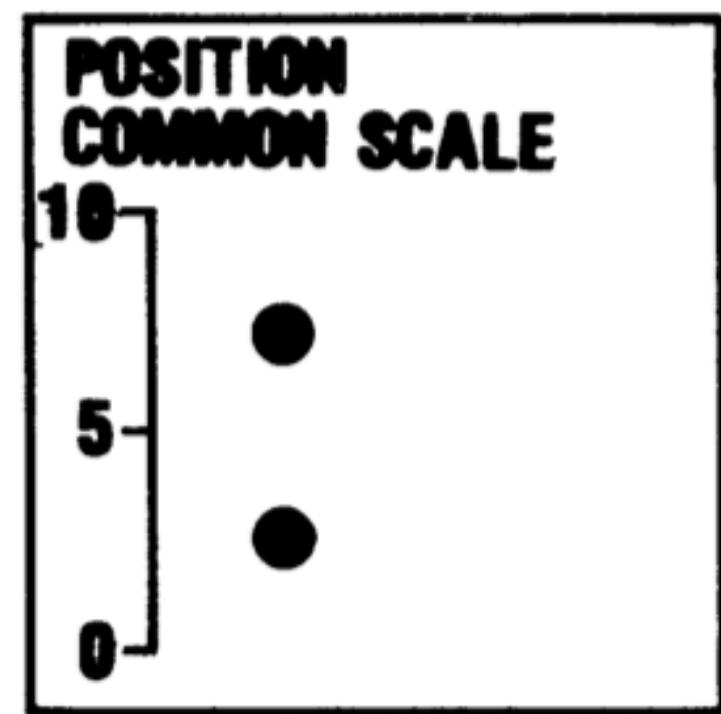
An Example Stacked Column Chart











COLOR SATURATION

Figure 1. Elementary perceptual tasks.

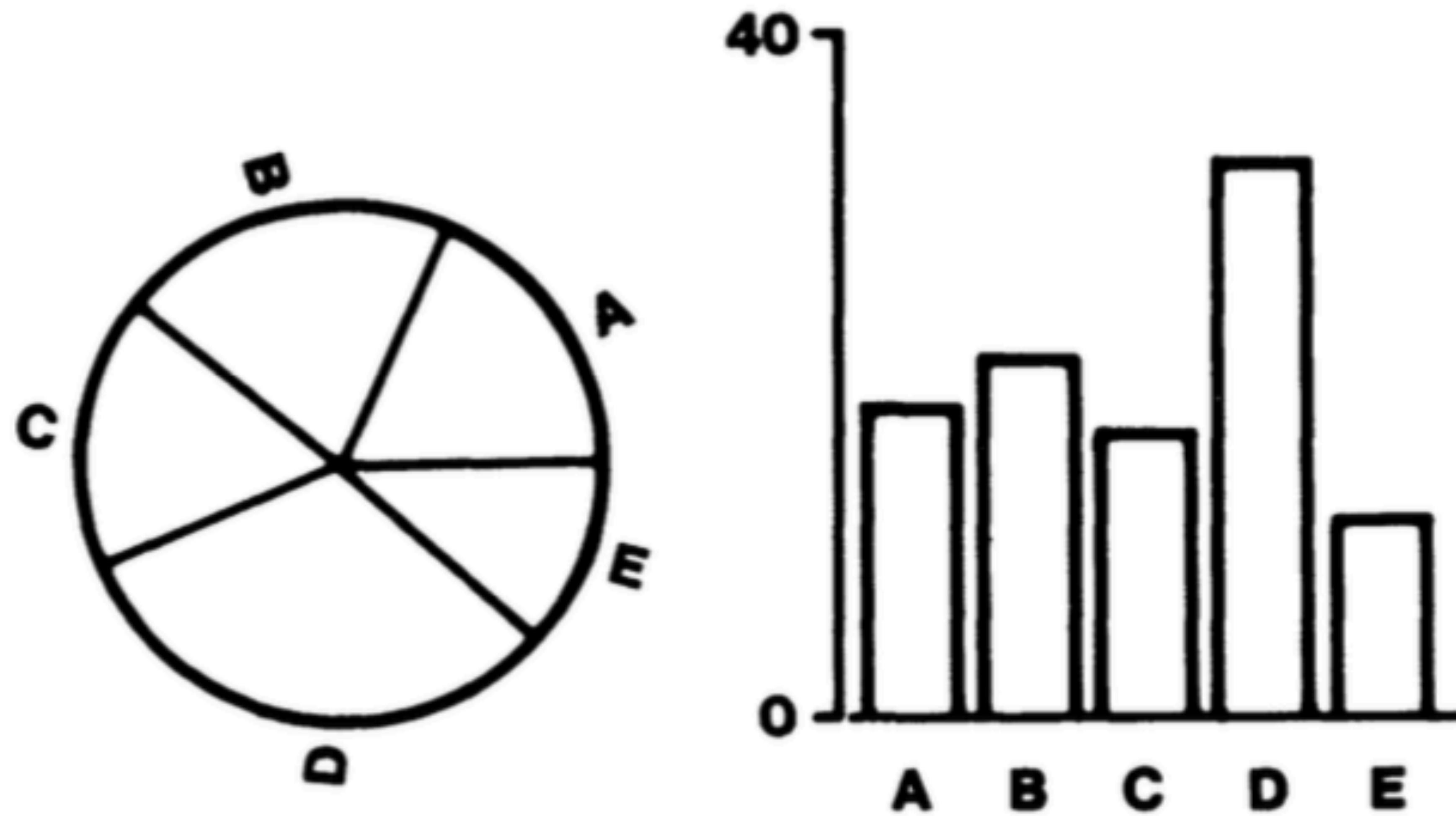


Figure 3. Graphs from position-angle experiment.

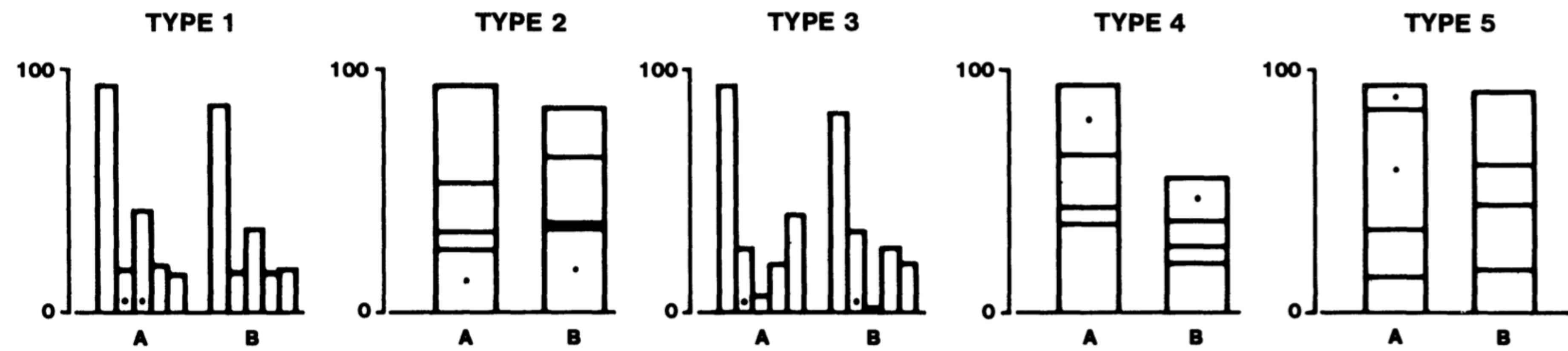
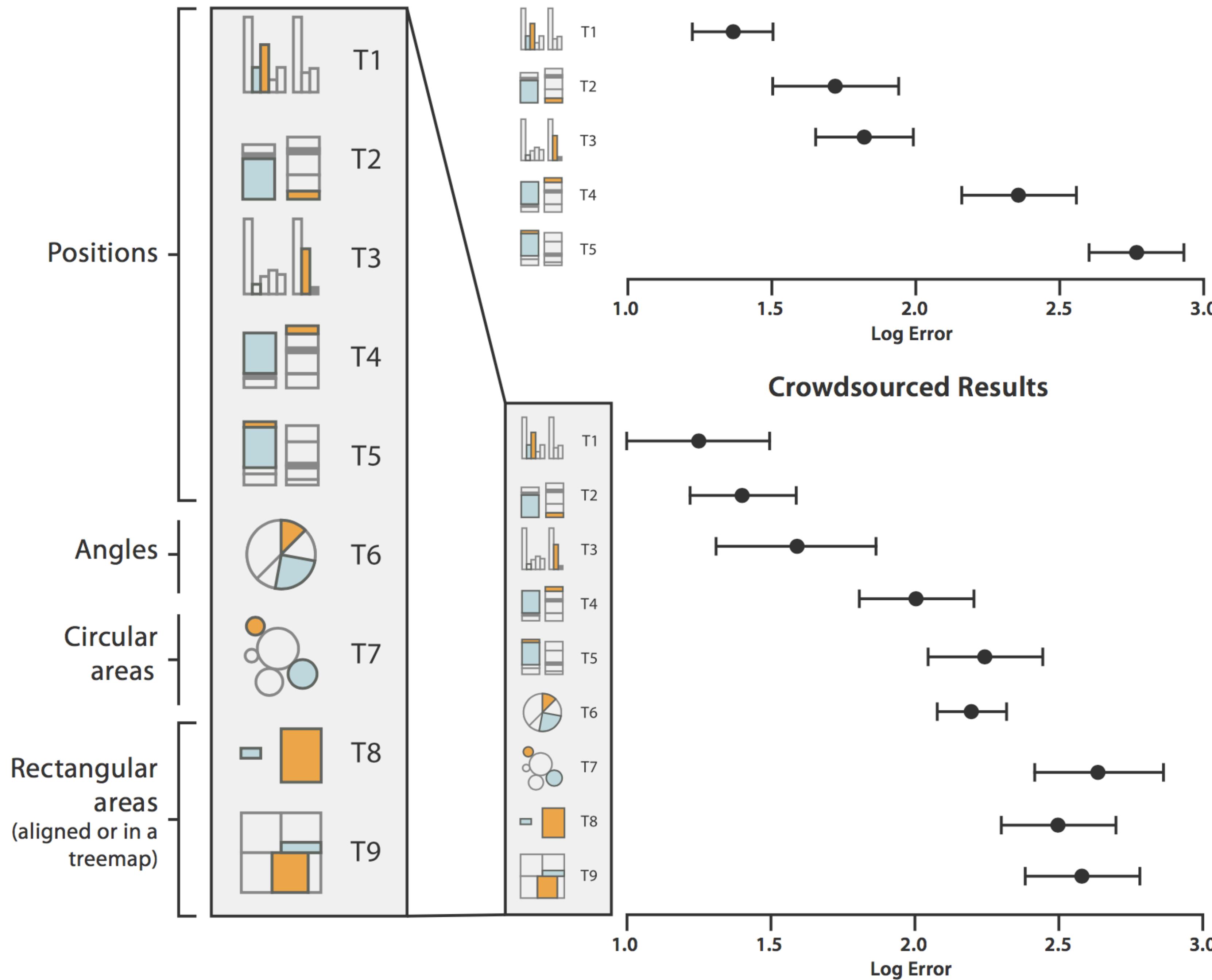


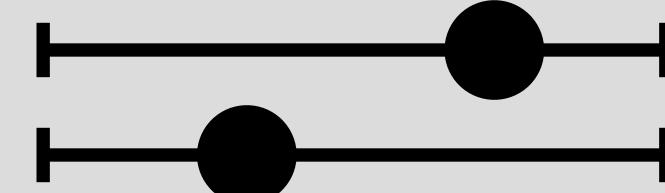
Figure 4. Graphs from position-length experiment.

Cleveland & McGill's Results



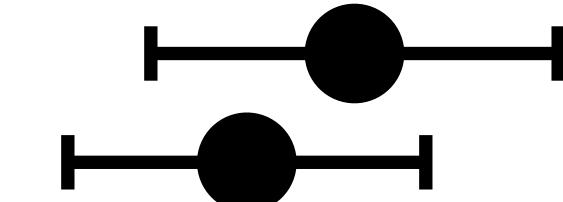
A Rough Hierarchy of Mappings

Position on a common scale

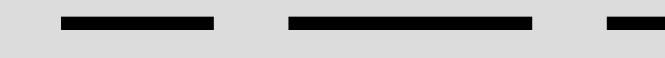


Better

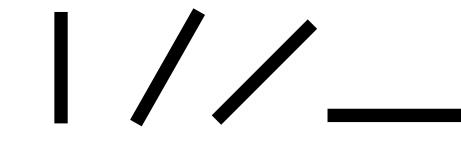
Position on unaligned scale



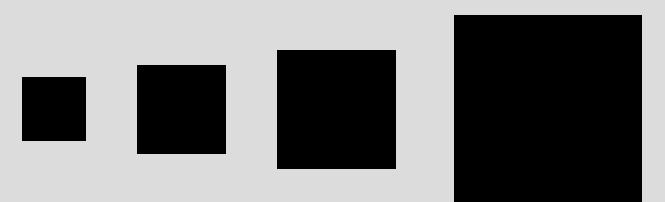
Length (1D as size)



Tilt or Angle



Area (2D as size)

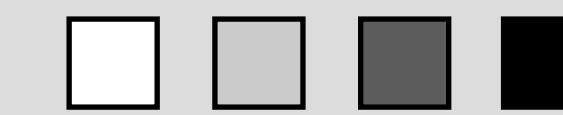


Effectiveness

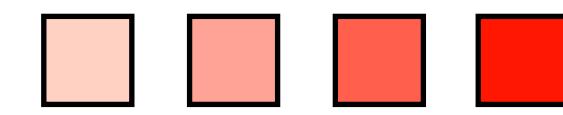
Depth (3D as Position)



Color luminance [brightness]



Color saturation [intensity]

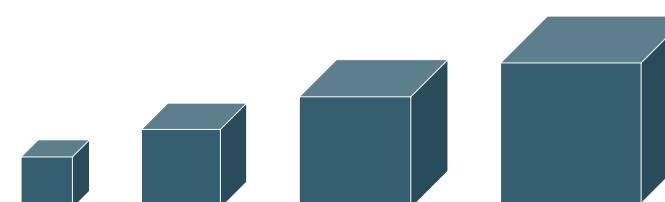


Curvature

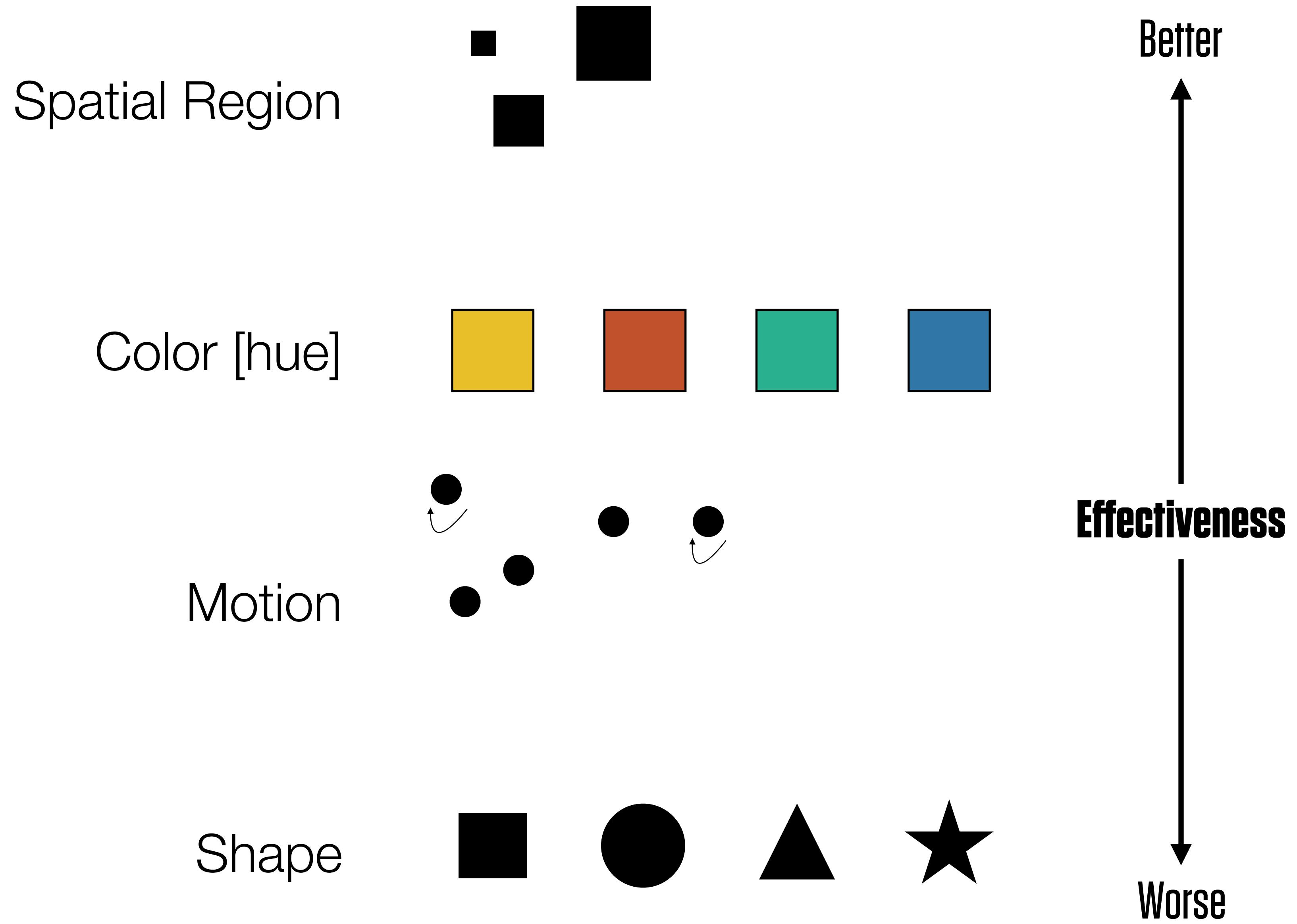


Worse

Volume (3D as size)



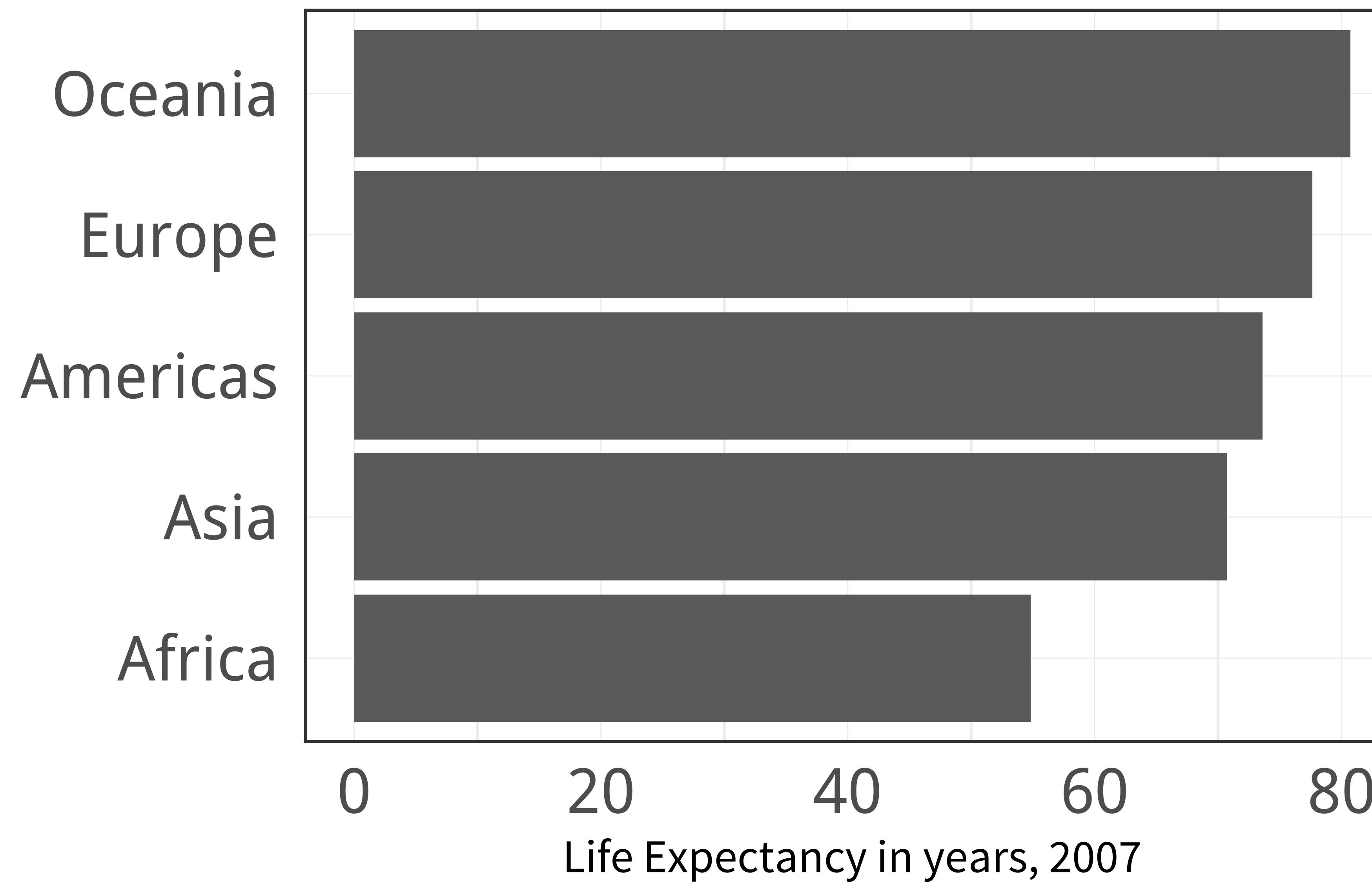
See e.g. Munzer (2014)

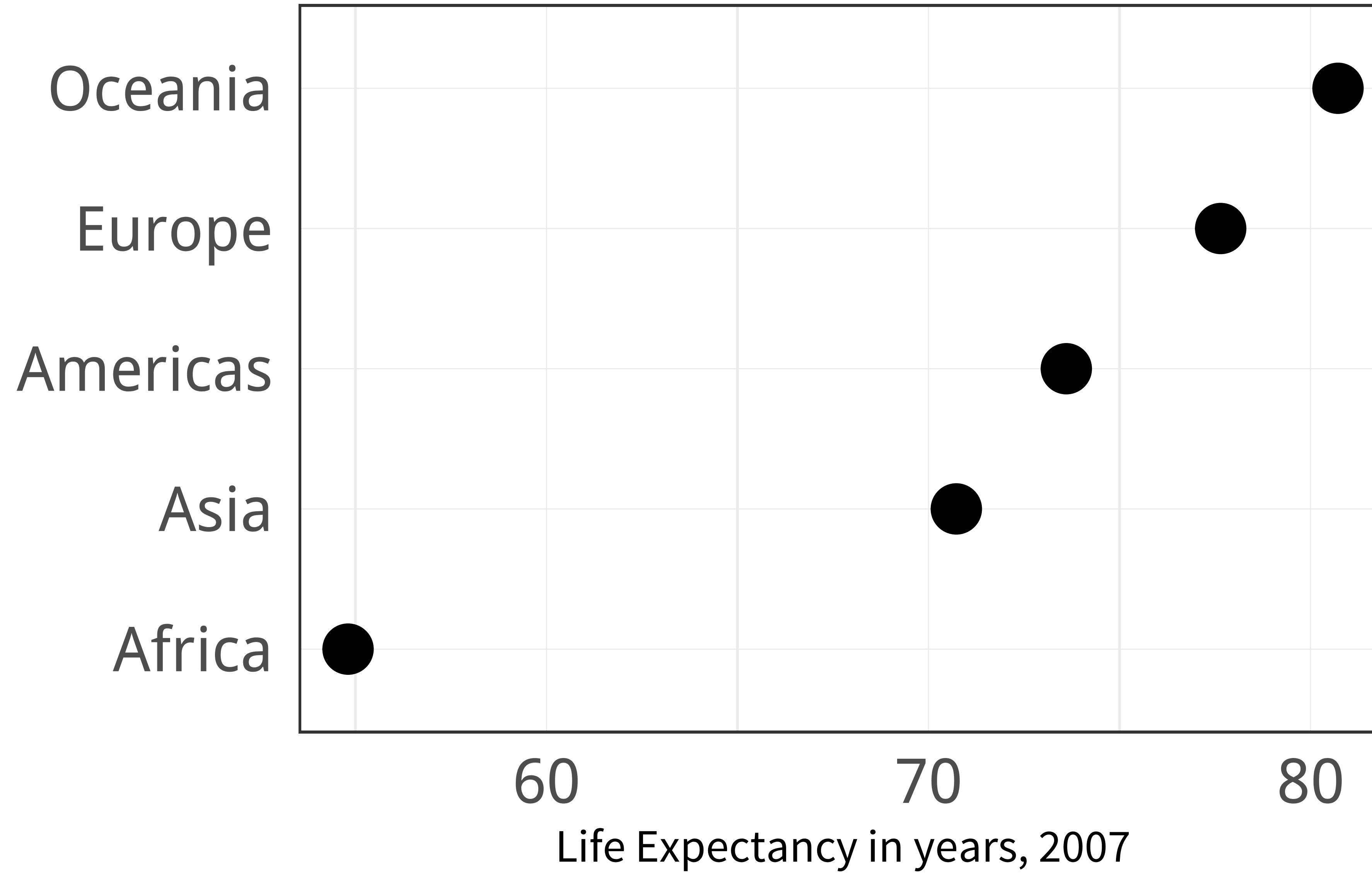


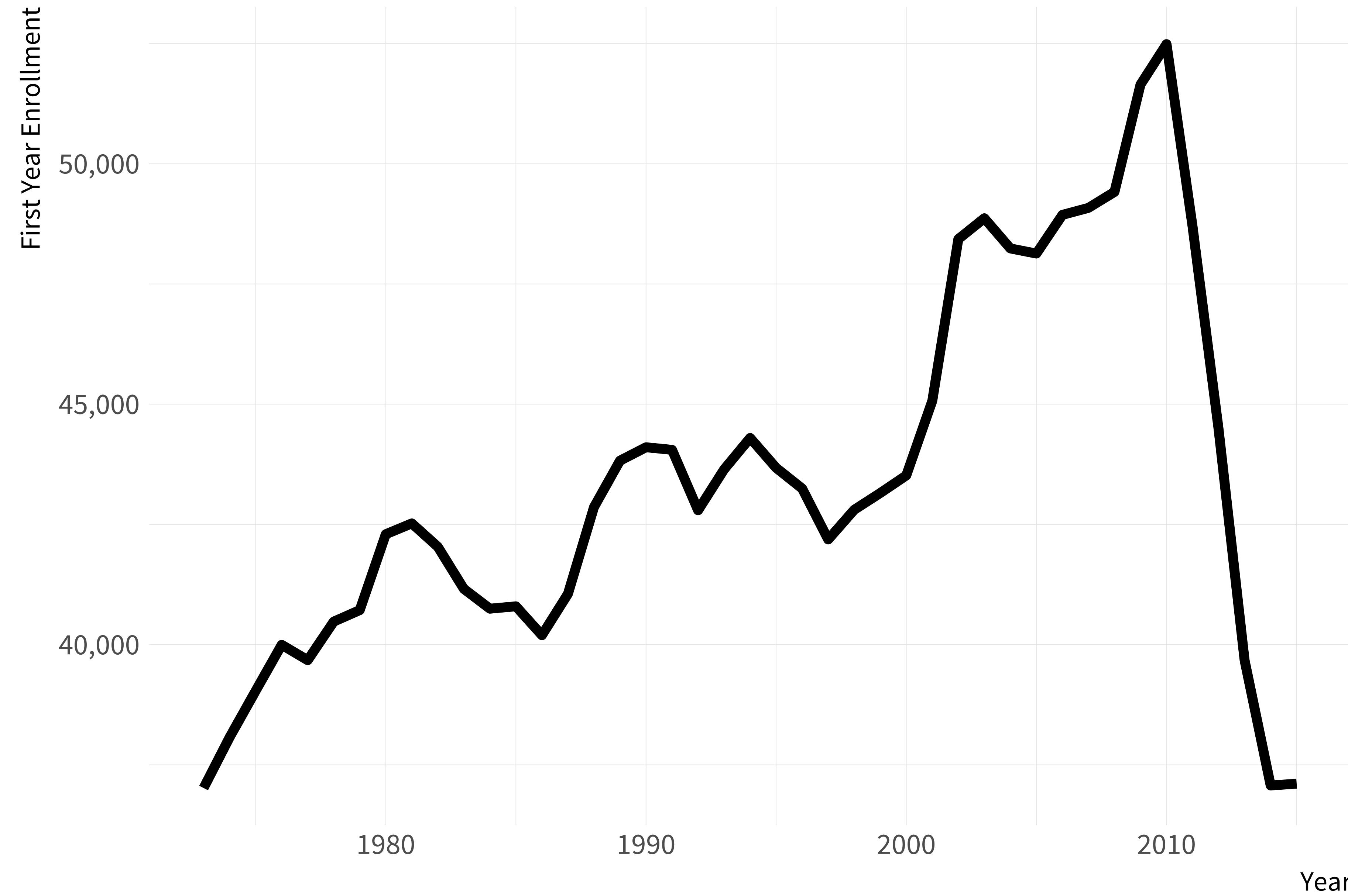
See e.g. Munzer (2014)

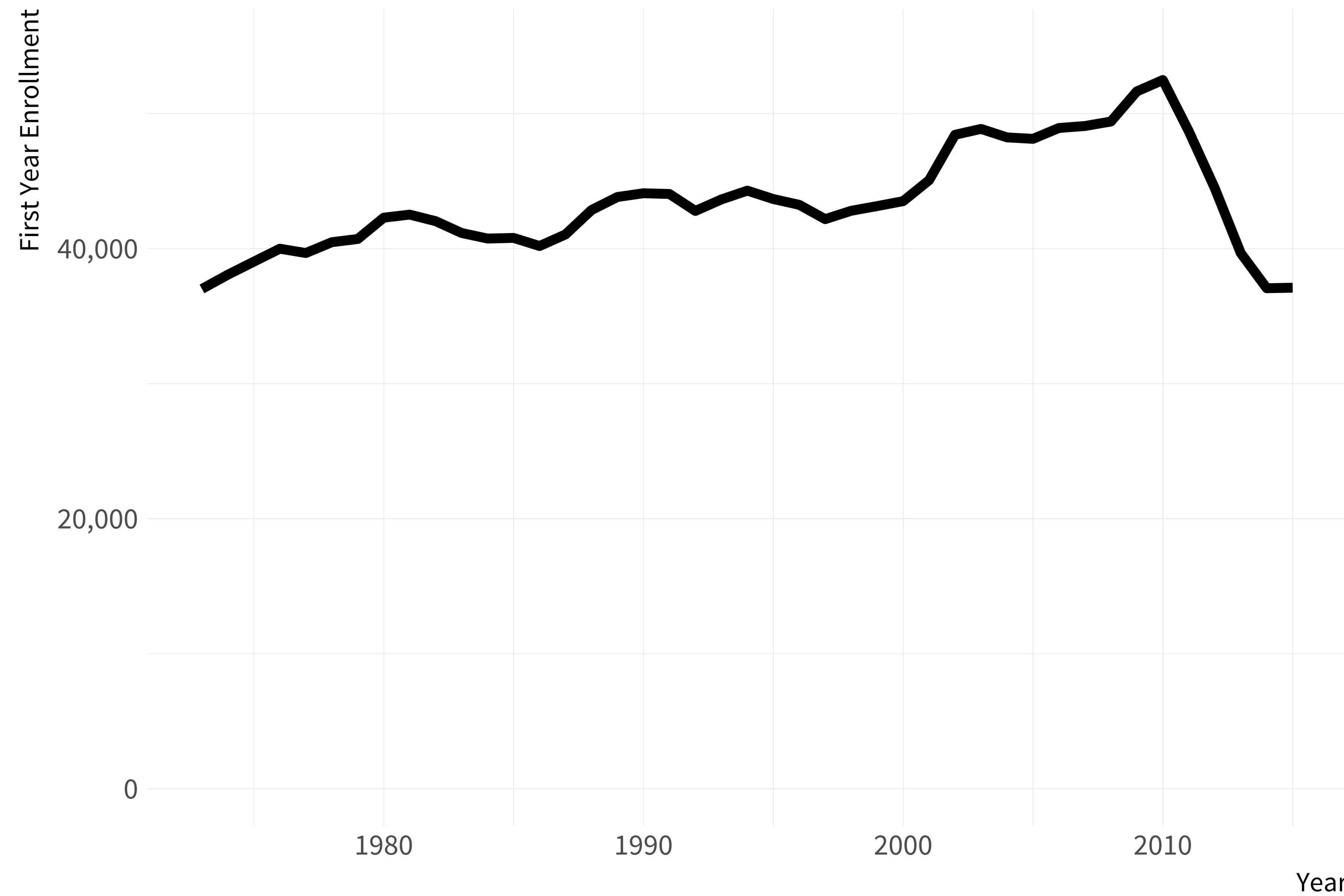
**HONESTY AND
GOOD JUDGMENT**

**Example:
Baselines**



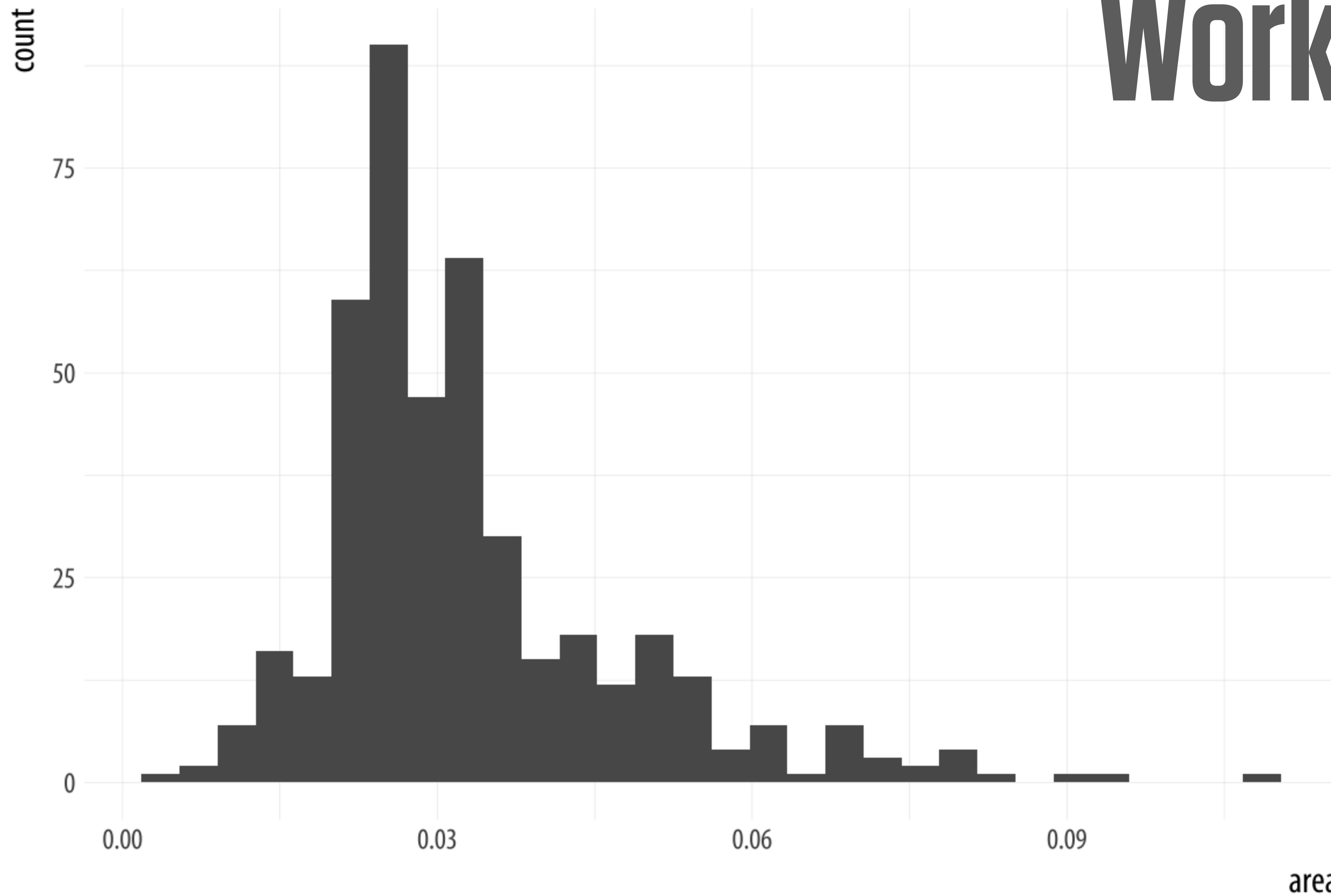


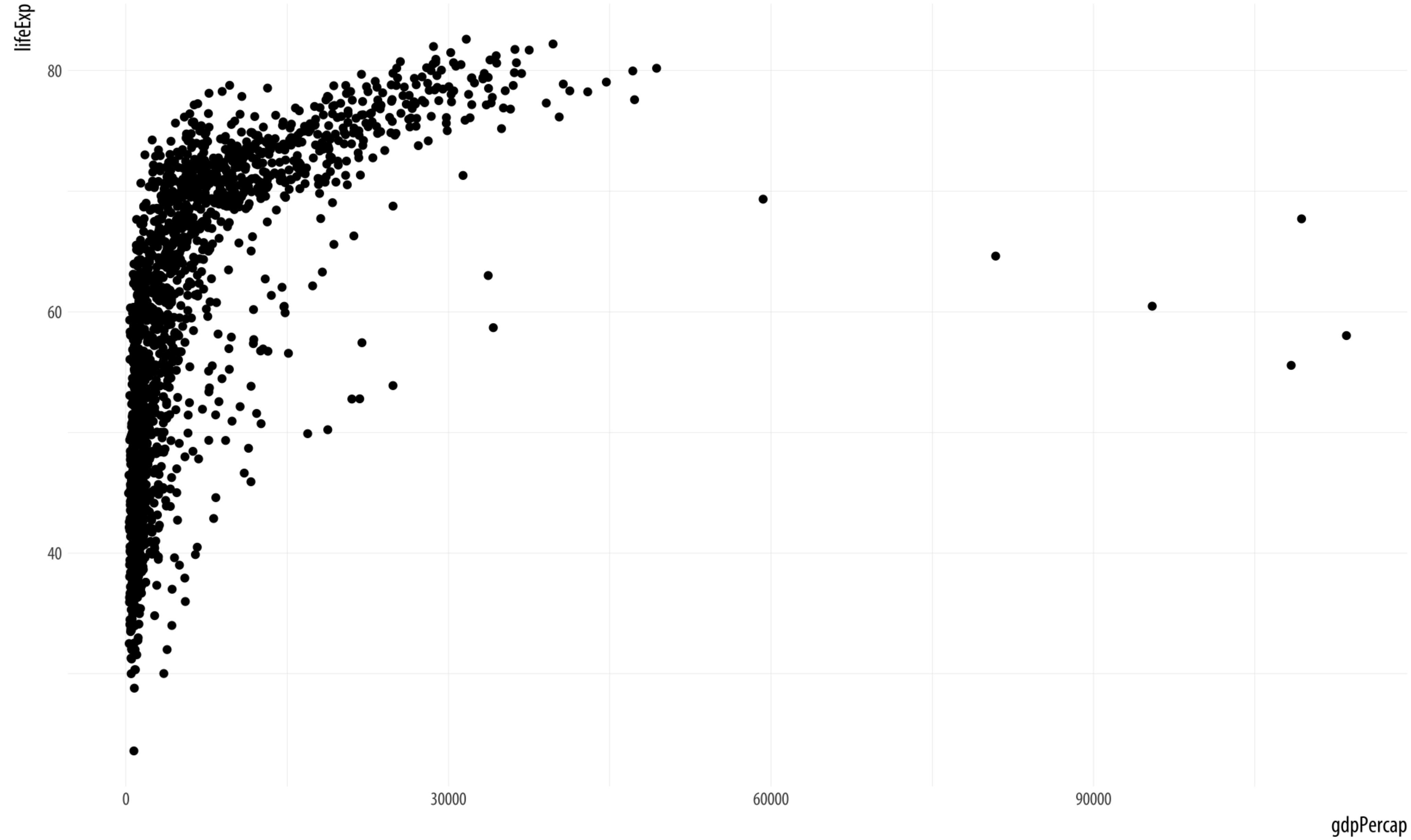


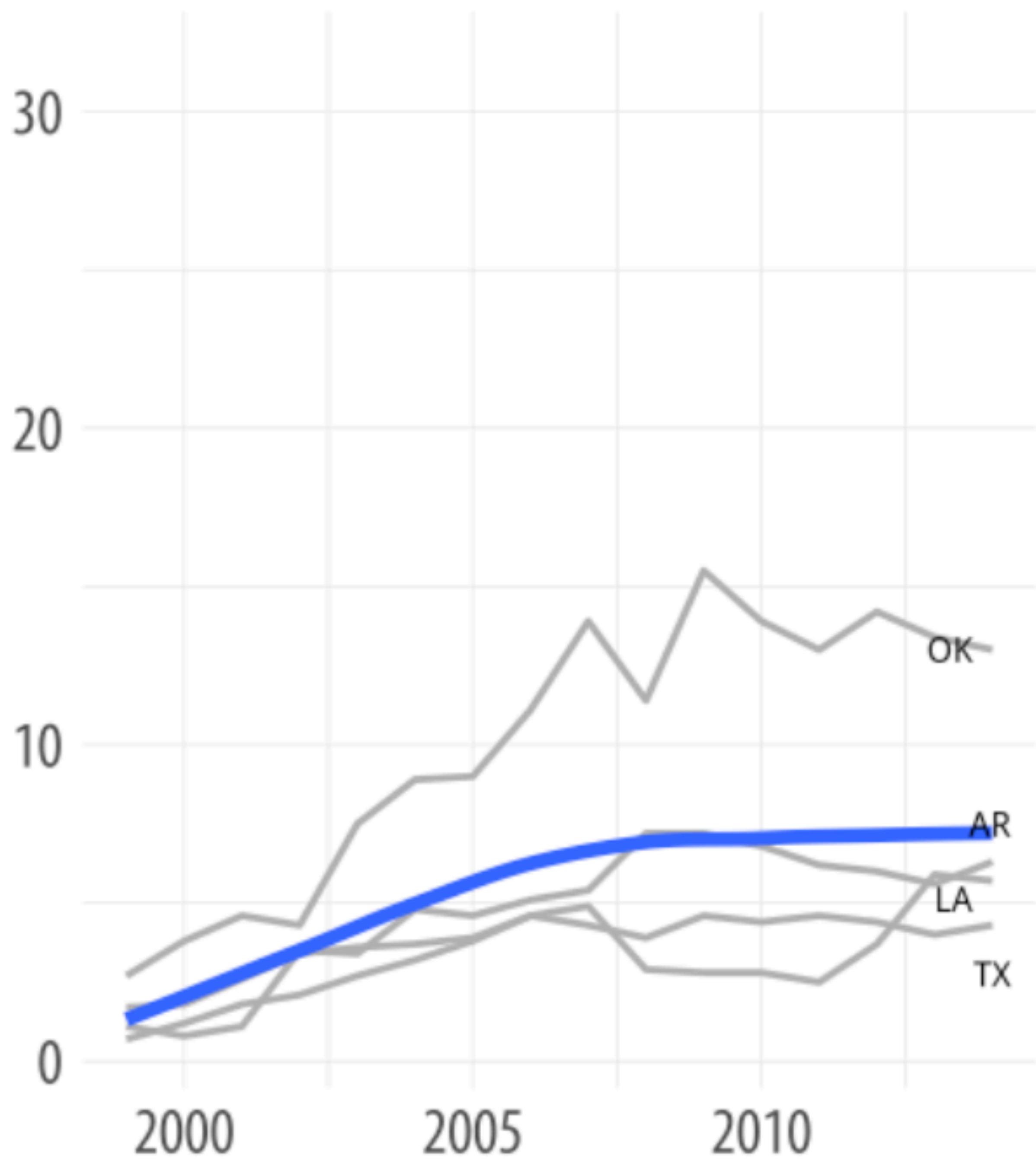


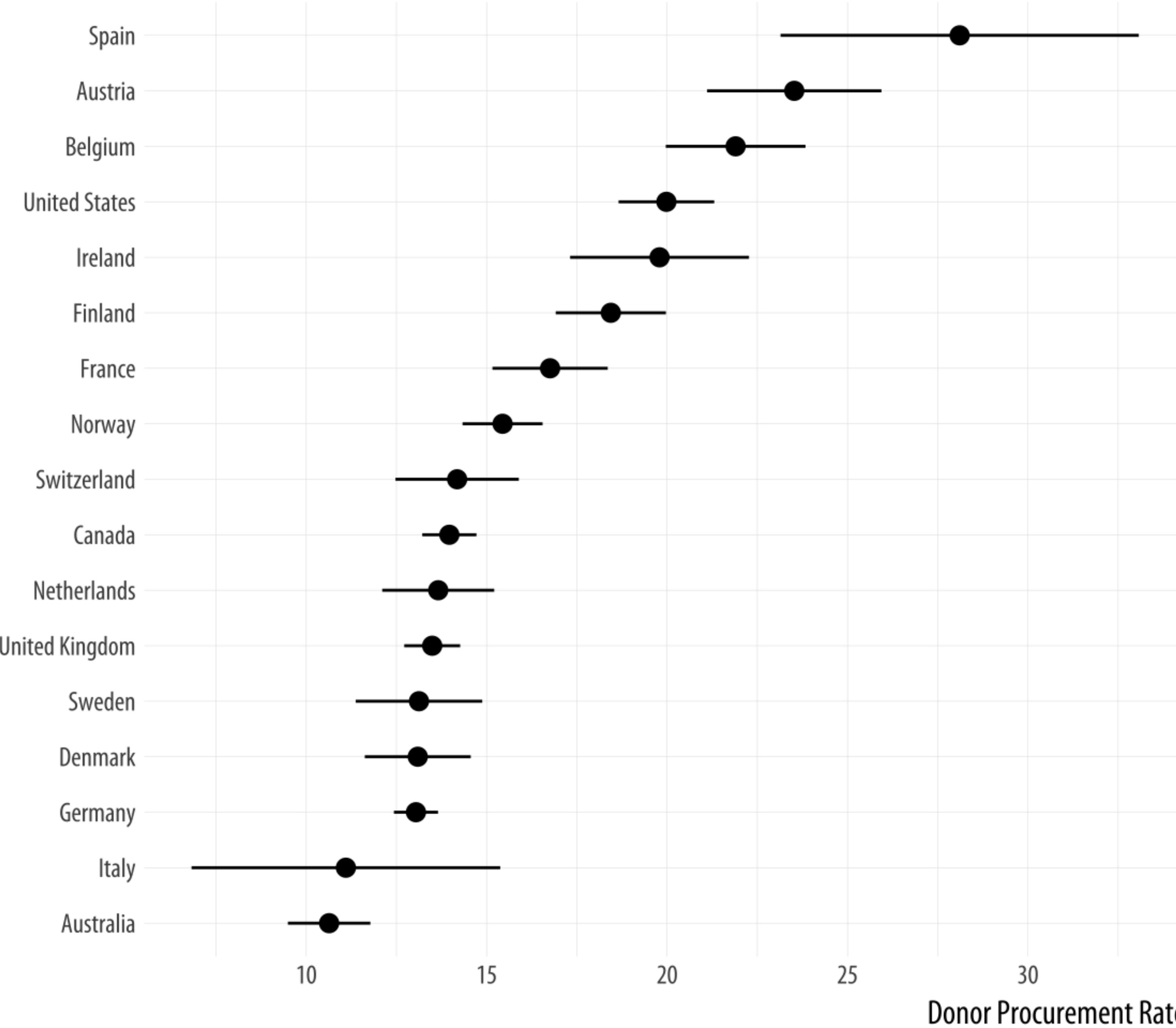
In Practice

Workhorses



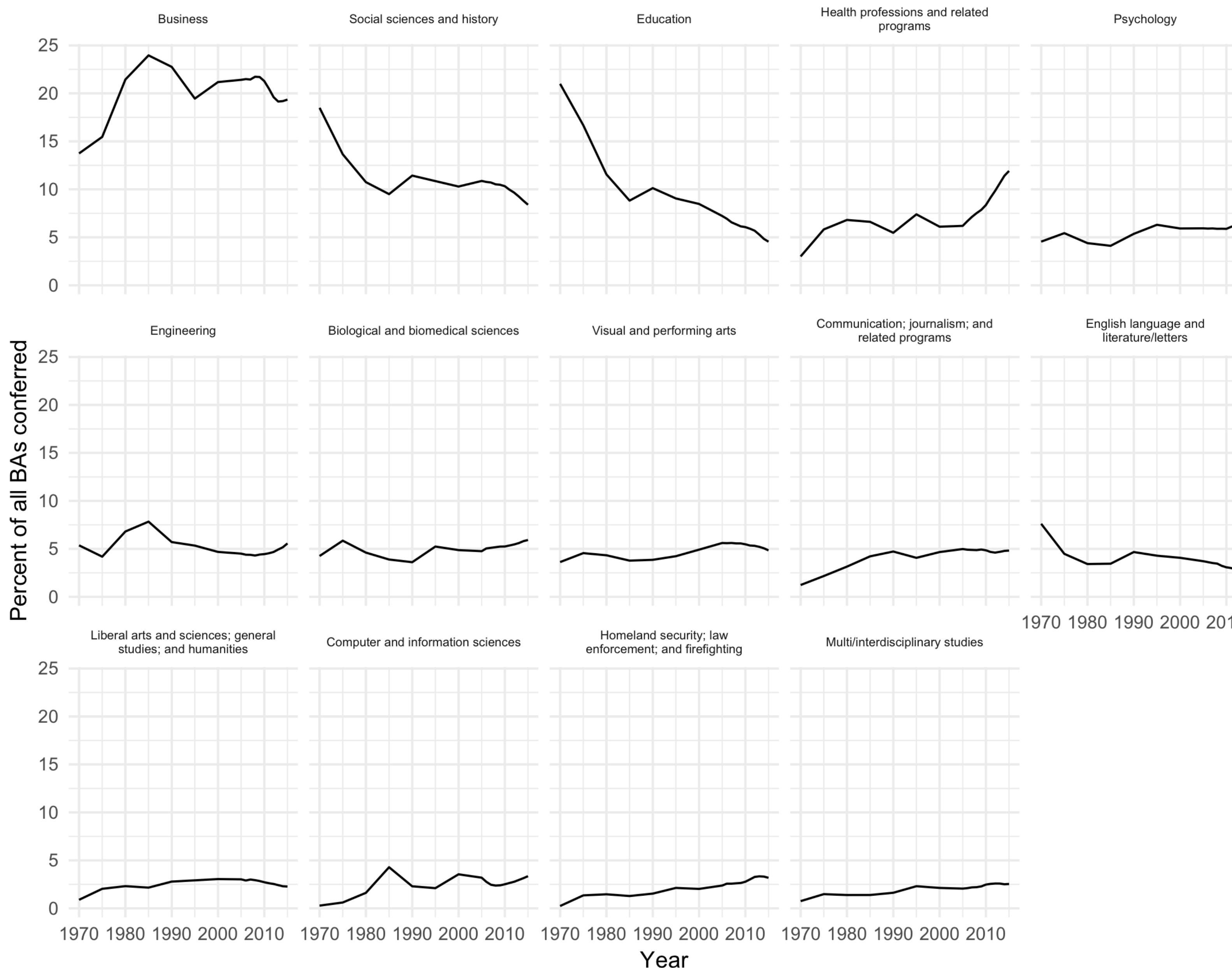






US Trends in Bachelor's Degrees Conferred, 1970-2015, for Areas averaging more than 2% of all degrees

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



Data from NCES Digest 2017, Table 322.10.

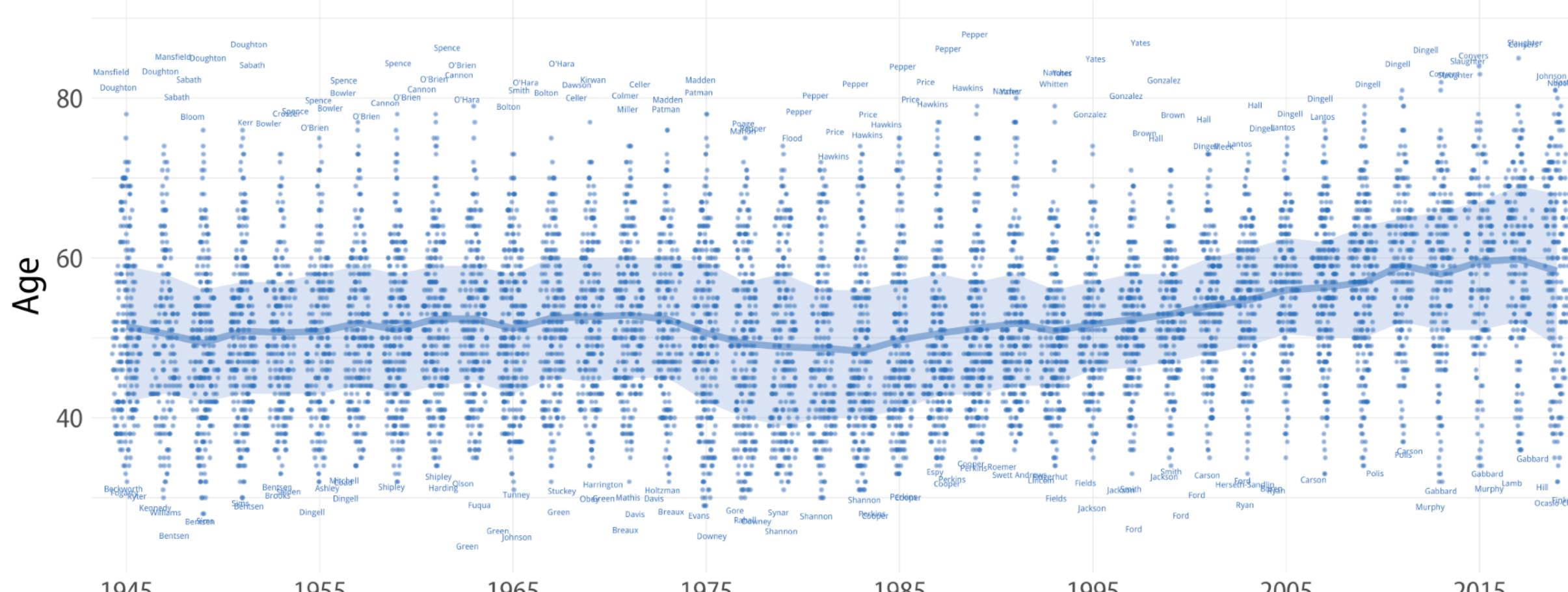
Show Ponies

Age Distribution of Congressional Representatives, 1945-2019

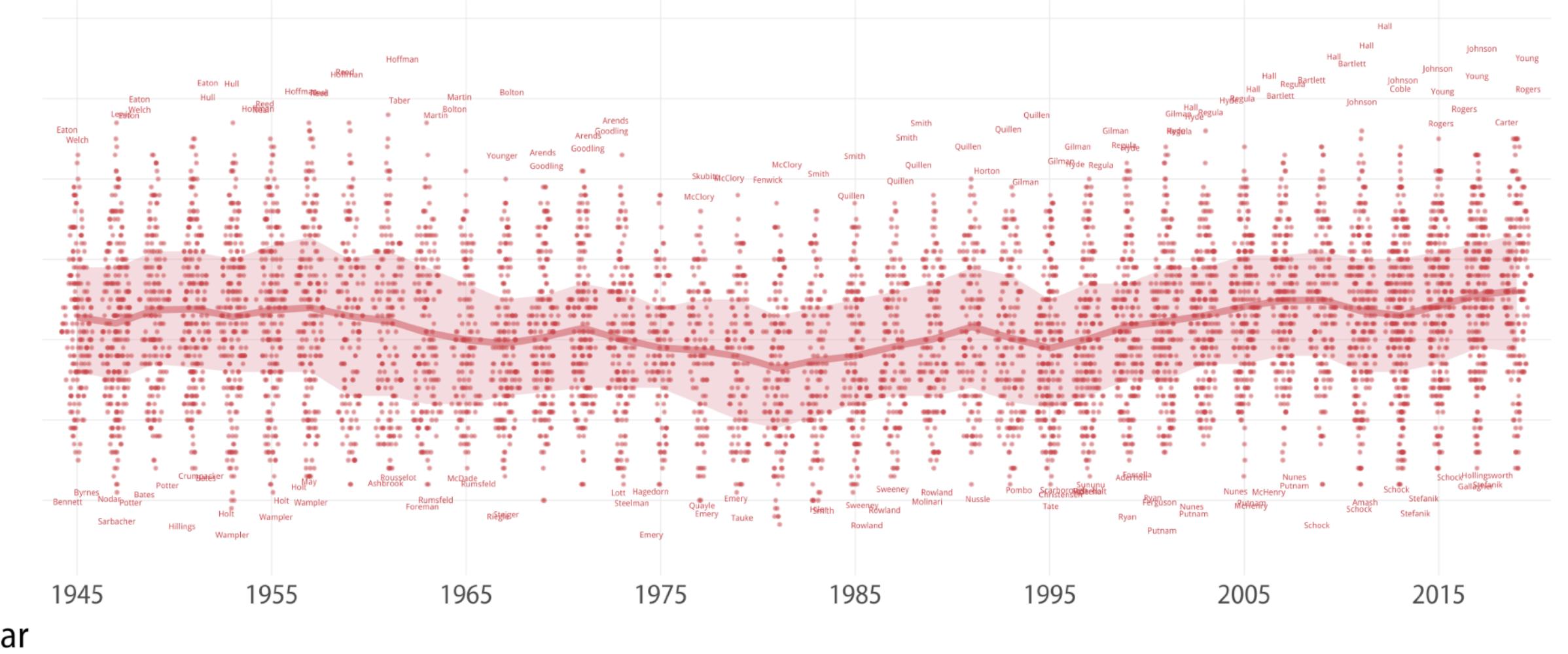
Trend line is mean age; bands are 25th and 75th percentiles of the range.

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

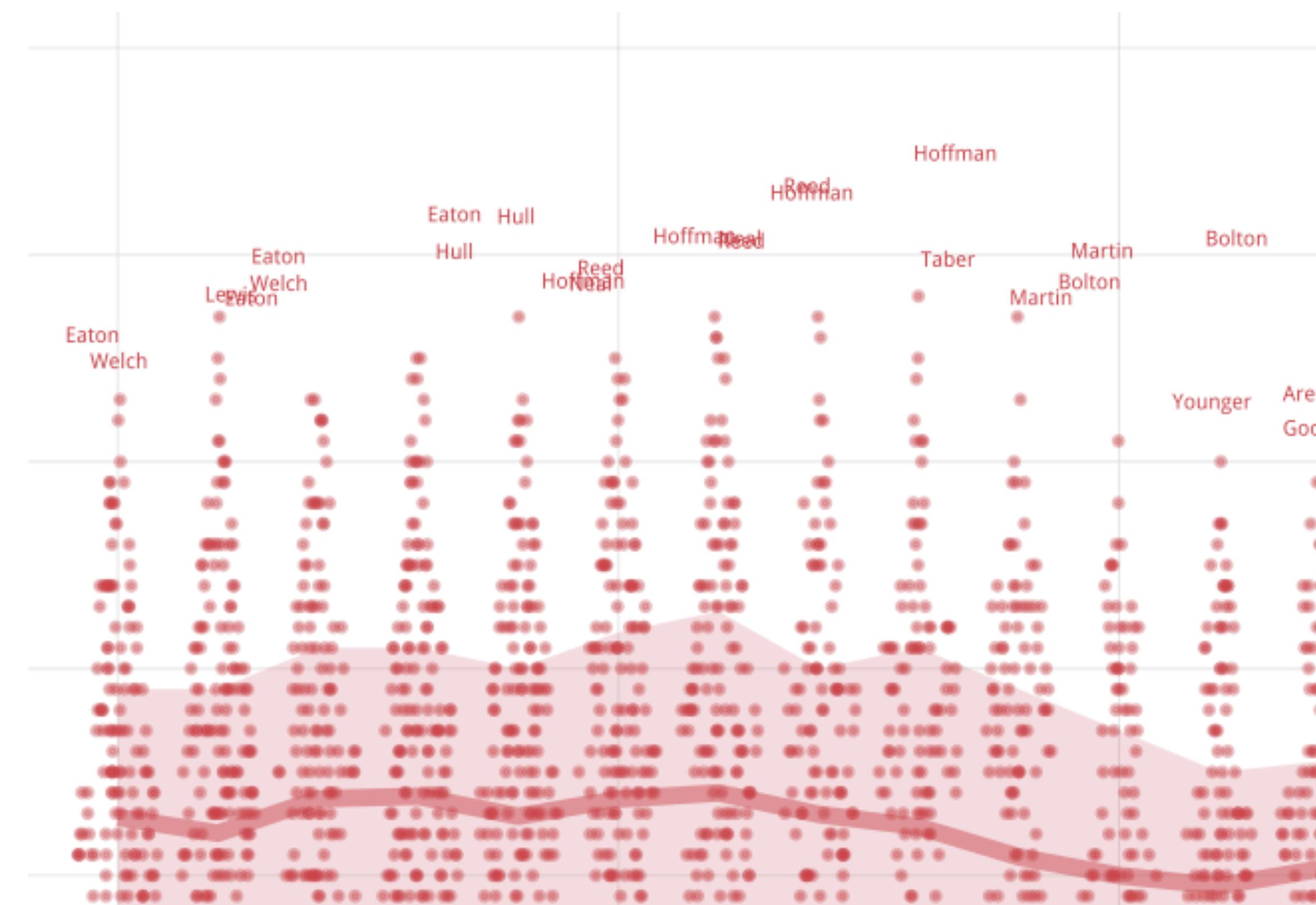
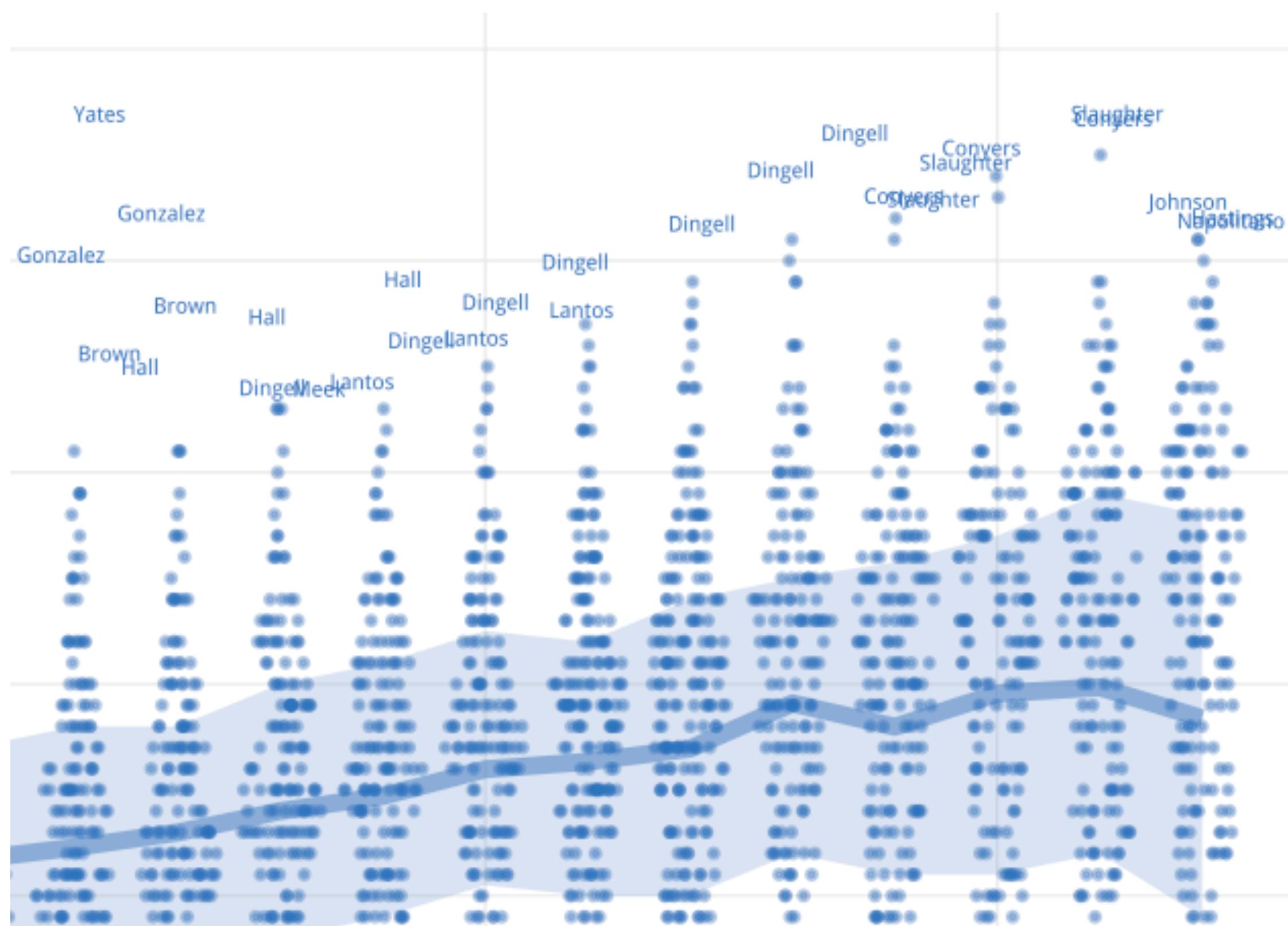
Democrats



Republicans



Show Ponies



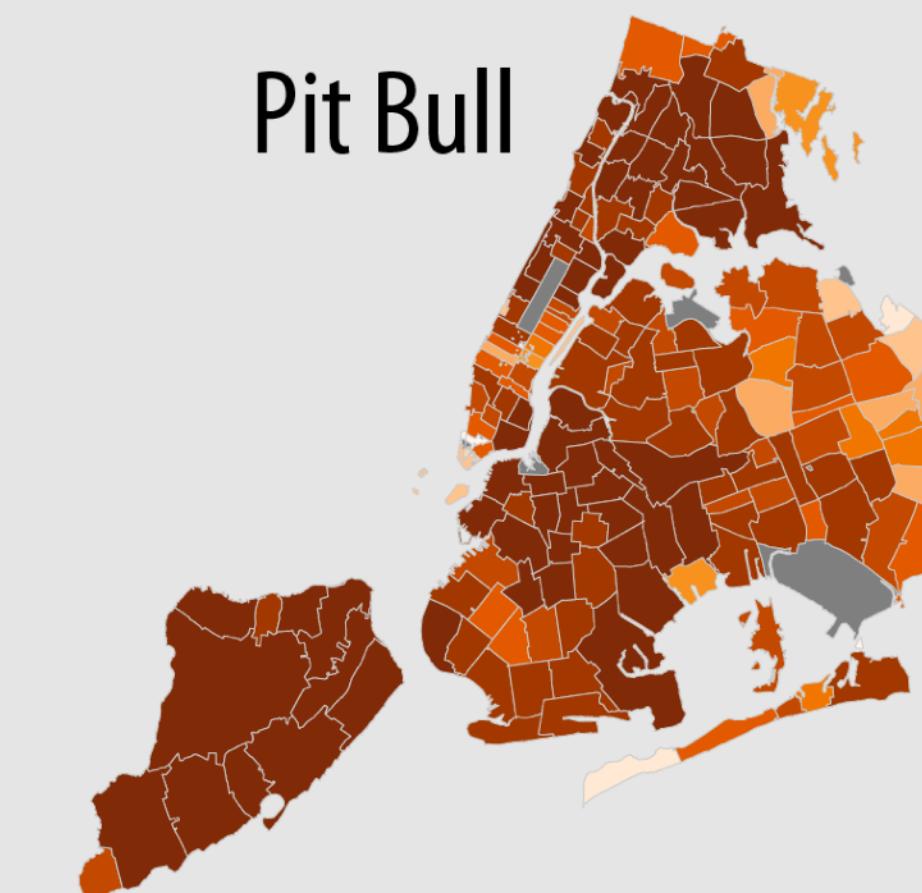
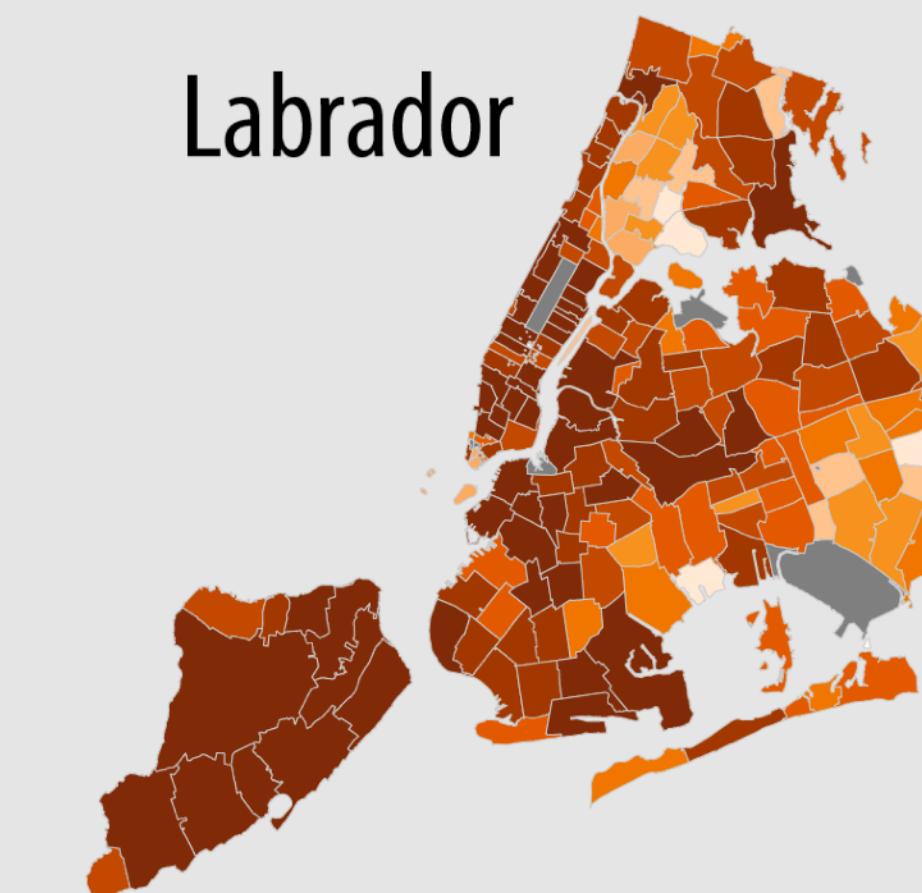
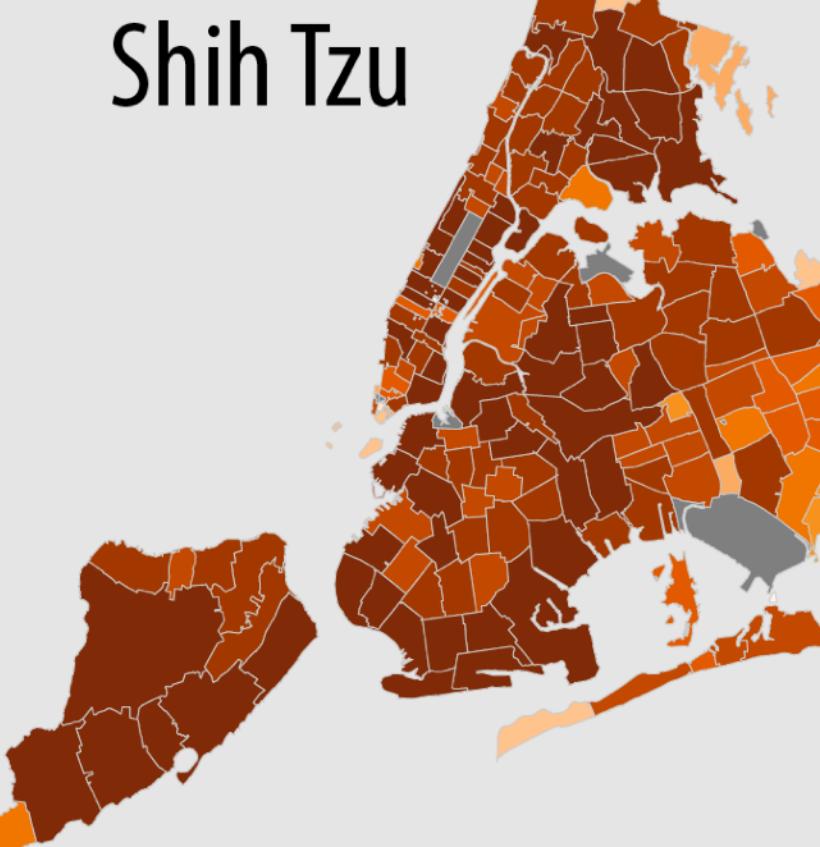
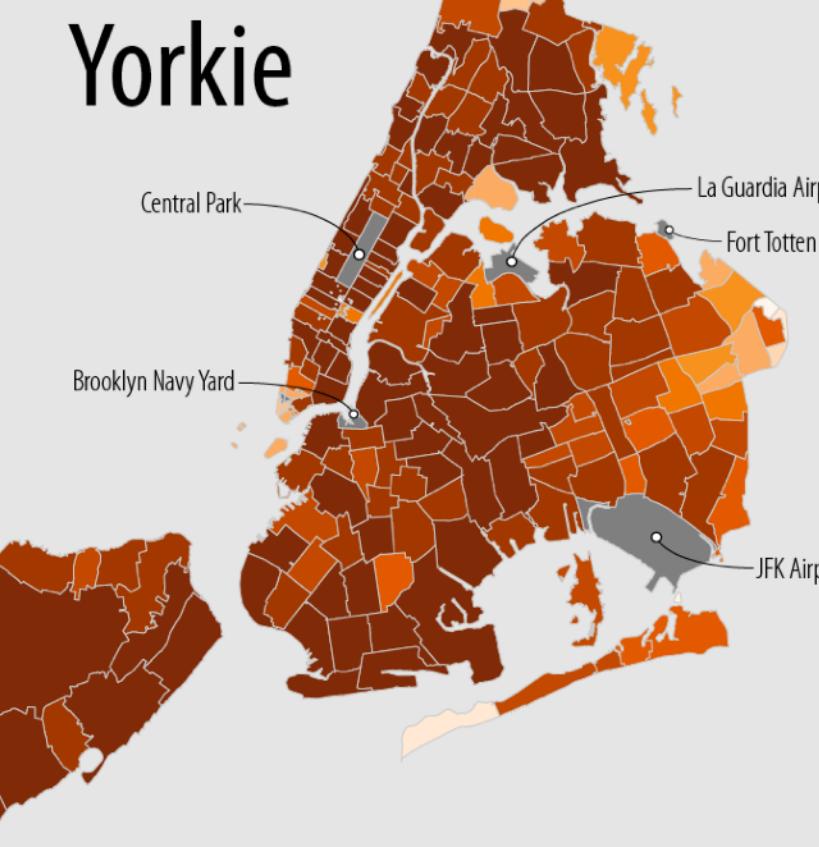
Relative prevalence of breed, from Most to Least



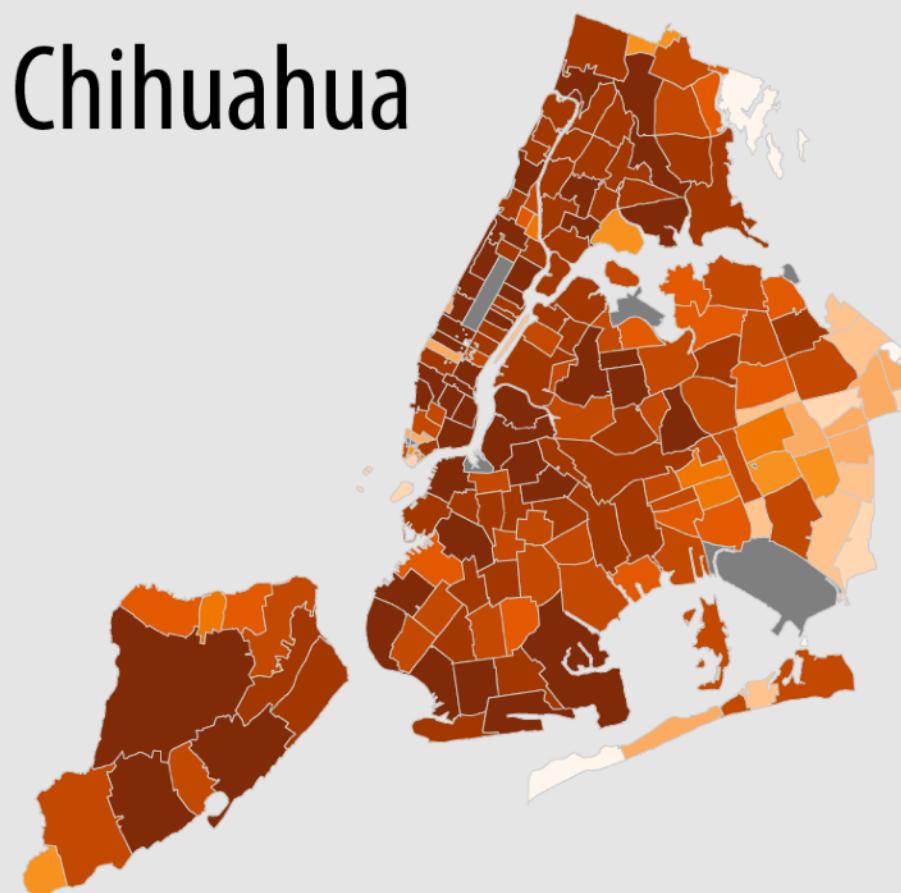
More Common

Less Common

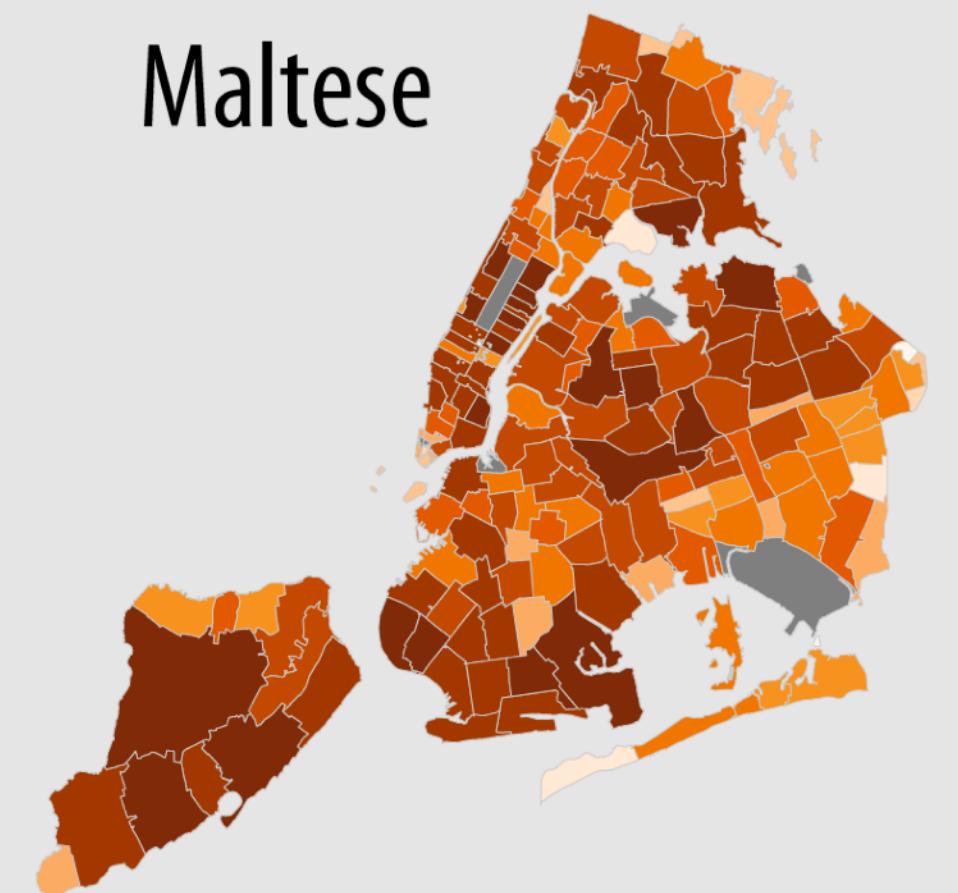
The most common
NYC breed is
“Mixed” or
“Unknown”



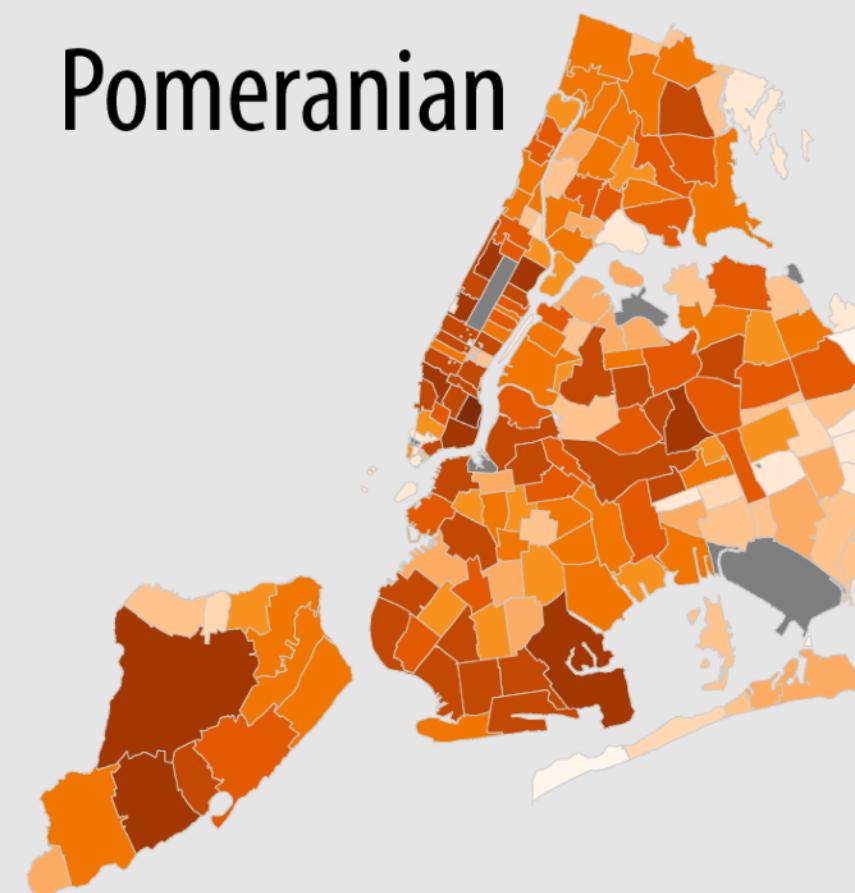
Chihuahua



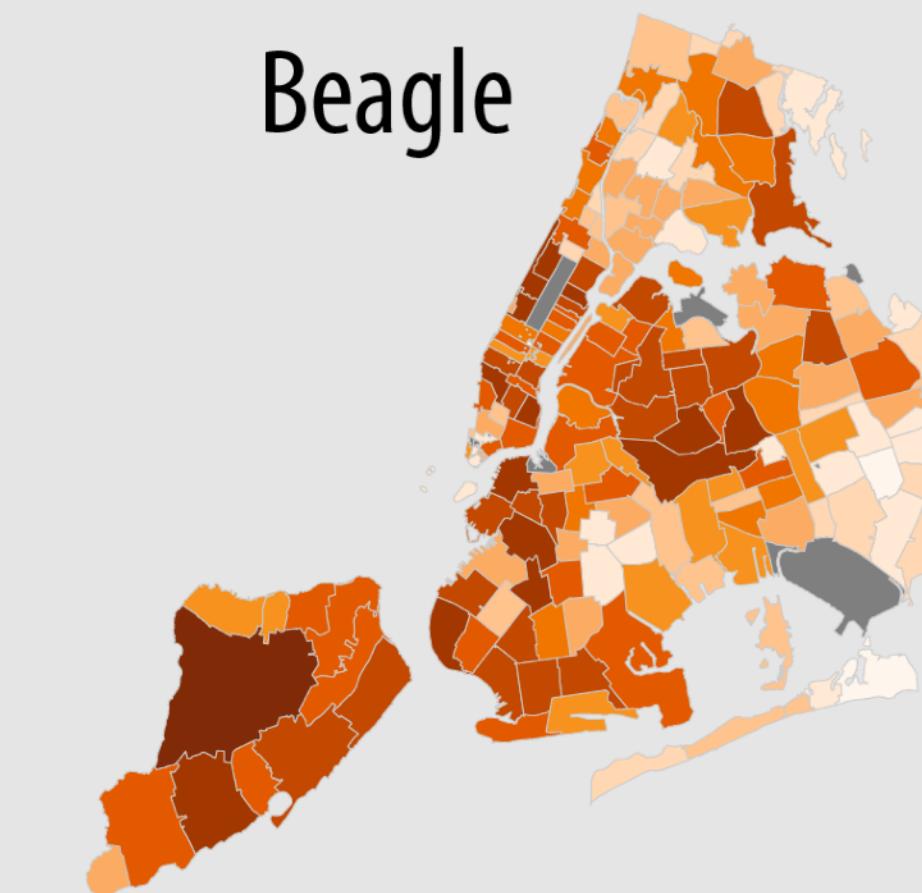
Maltese



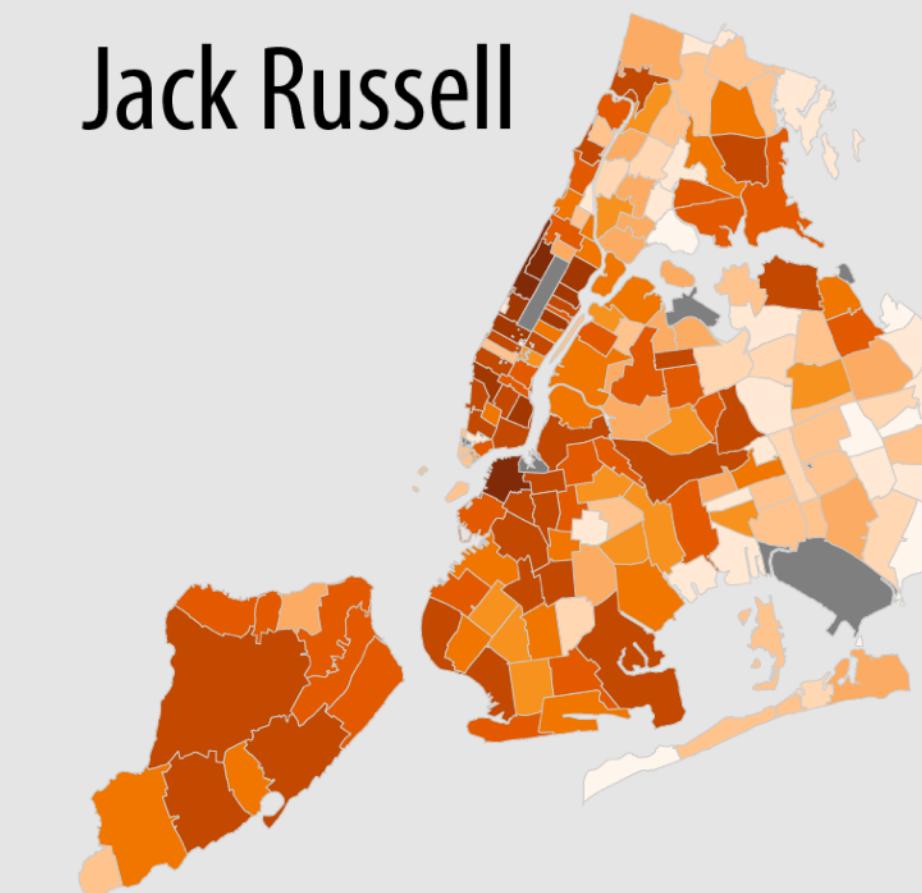
Pomeranian



Beagle

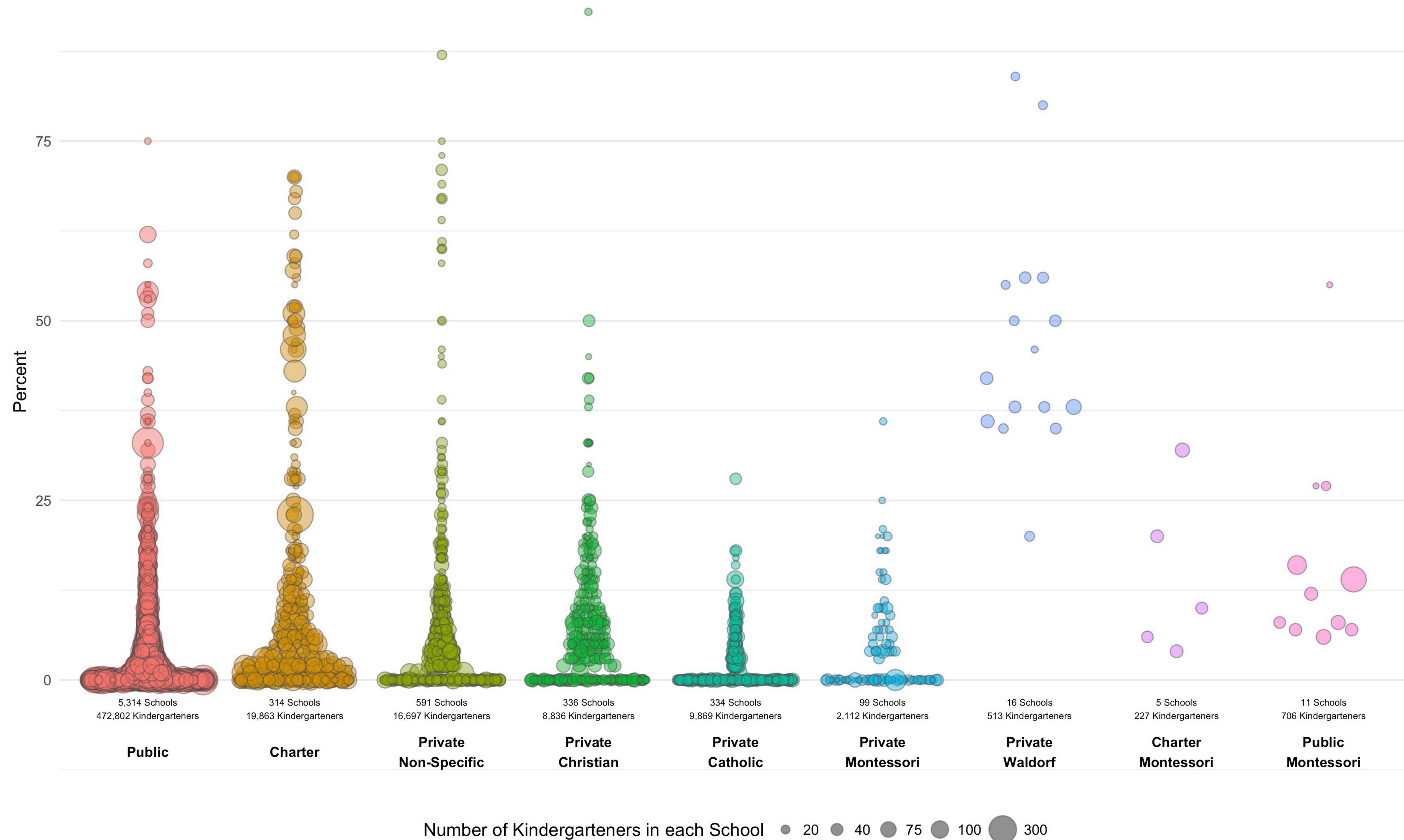


Jack Russell

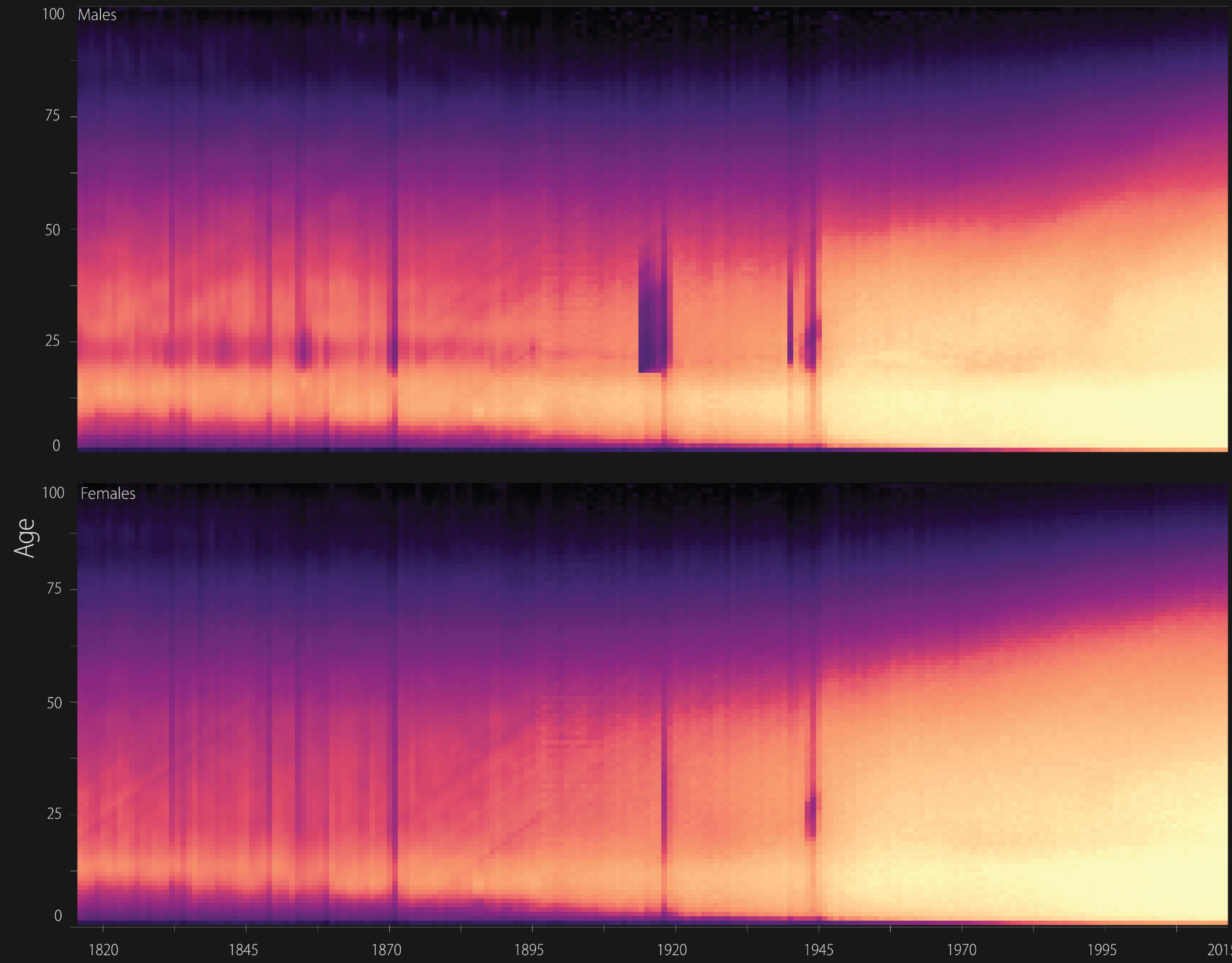


Vaccination Exemption Rates in California Kindergartens

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



MORTALITY IN FRANCE 1816 – 2016



Unicorns

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Drawn by M. Minard, Inspecteur Général des Ponts et Chaussées en retraite

Paris, le 20 Novembre 1869

Les nombres d'hommes perdus sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres les zones. Le rouge désigne les hommes qui ont péri en Russie, le noir ceux qui en sont sortis. — Les renseignements qui ont servi à dresser la carte sont tirés dans les ouvrages de M. M. Chiers, de Léglur, de Fezoridac, de Chambray et le journal intime de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'aïl la diminution de l'armée, j'ai supposé que les corps de l'Armée Dernière et du Maréchal Davout qui avaient été détachés sur Moscou au début et qui rejoignirent Orléans et Wladiwostok, avaient toujours marché avec l'armée.

