

Data Visualization with R and `ggplot2`

Data Visualization: Session 1

Kieran Healy
Code Horizons, May 2022

Housekeeping

Housekeeping

10:30am to 12:30pm US EST

Housekeeping

10:30am to 12:30pm US EST

Break from 12:30pm to 1:30pm

Housekeeping

10:30am to 12:30pm US EST

Break from 12:30pm to 1:30pm

1:30pm to 3:00pm

Use the Zoom chat to ask questions, or raise a hand with 

In between class sessions



My Setup and Yours

My Setup and Yours

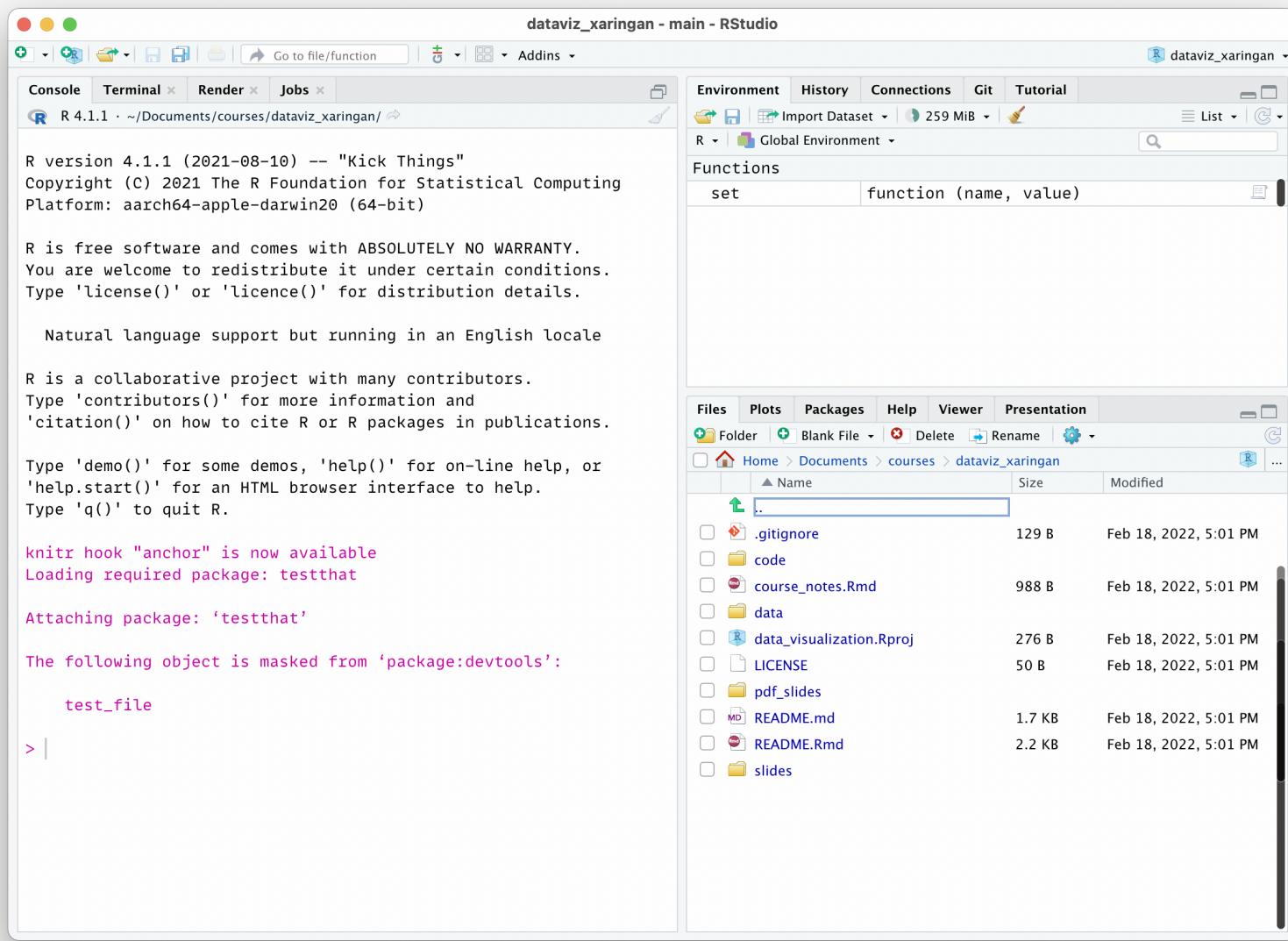
Talking, Slides, and Live-Coding in RStudio

Follow along and take notes in RStudio yourself

The course packet is also an RStudio project

Get up and Running





dataviz_xaringan - main - RStudio

course_notes.Rmd x 01-looking-at-data.Rmd x Knit on Save ABC Knit Run Addins Environment History Connections Git Tutorial Import Dataset 271 MiB List Global Environment Functions set function (name, value)

```
1 ---  
2 title: "Data Visualization Notes"  
3 author: "<YOUR NAME HERE>"  
4 date: "April 2022"  
5 output: html_document  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10 ```  
11  
12 ## Welcome  
13  
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.
```

11:1 (Top Level) R Markdown

Console Terminal Render Jobs R 4.1.1 · ~/Documents/courses/dataviz_xaringan/ ls.

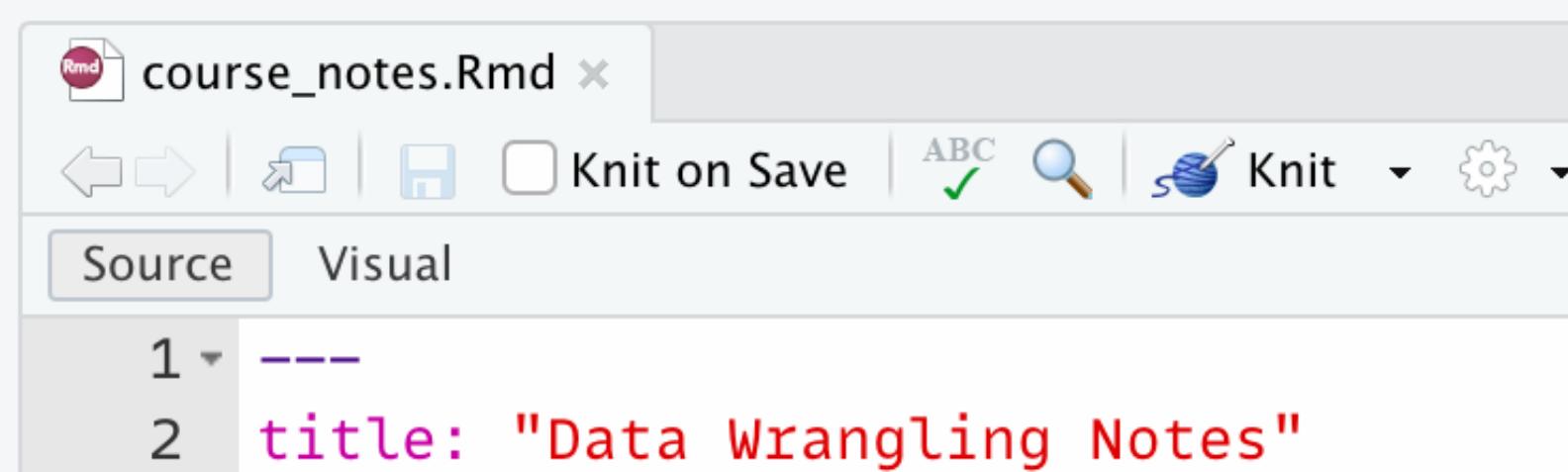
Natural language support but running in an English locale

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.

Files Plots Packages Help Viewer Presentation New Folder New Blank File Delete Rename More Home > Documents > courses > dataviz_xaringan Name Size Modified .. .gitignore 129 B Feb 18, 2022, 5:01 PM code course_notes.Rmd 990 B Feb 19, 2022, 8:07 AM data data_visualization.Rproj 276 B Feb 18, 2022, 5:01 PM LICENSE 50 B Feb 18, 2022, 5:01 PM pdf_slides README.md 1.7 KB Feb 18, 2022, 5:01 PM README.Rmd 2.2 KB Feb 18, 2022, 5:01 PM slides

Try knitting your notes



The screenshot shows the RStudio interface with a file named "course_notes.Rmd" open. The toolbar includes icons for back, forward, save, and knit, along with "Knit on Save" and search functions. Below the toolbar, there are two tabs: "Source" (which is selected) and "Visual". The main pane displays the following R Markdown code:

```
1 ---  
2 title: "Data Wrangling Notes"
```

Don't worry if it's not clear what's happening at this point.

Now write the following code

Write this out inside the "code chunk" in your notes.

```
library(tidyverse)
library(gapminder)

p <- ggplot(data = gapminder,
             mapping = aes(x = gdpPercap,
                            y = lifeExp))

p + geom_point()
```

... And knit your document again.

Now write the following code

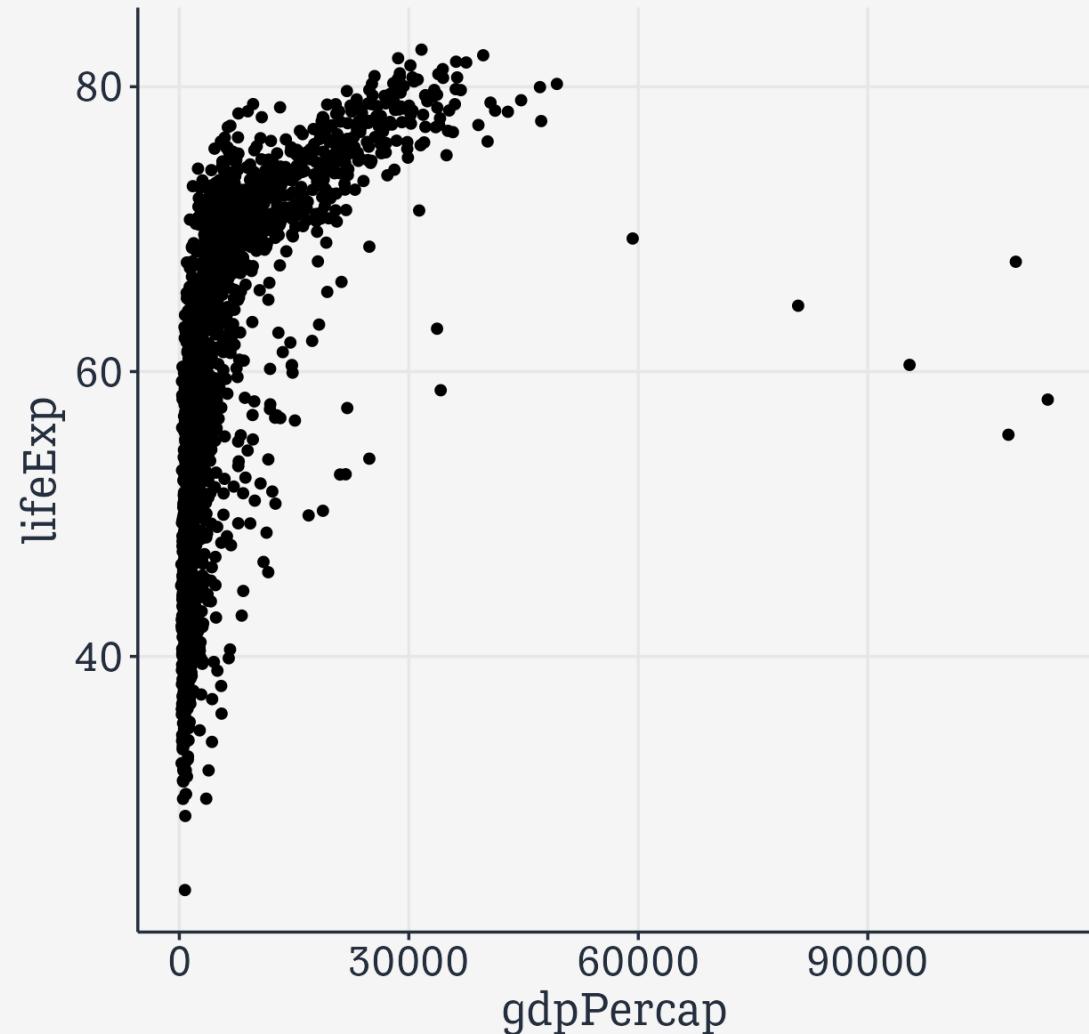
Write this out inside the "code chunk" in your notes.

```
library(tidyverse)
library(gapminder)

p <- ggplot(data = gapminder,
             mapping = aes(x = gdpPercap,
                            y = lifeExp))

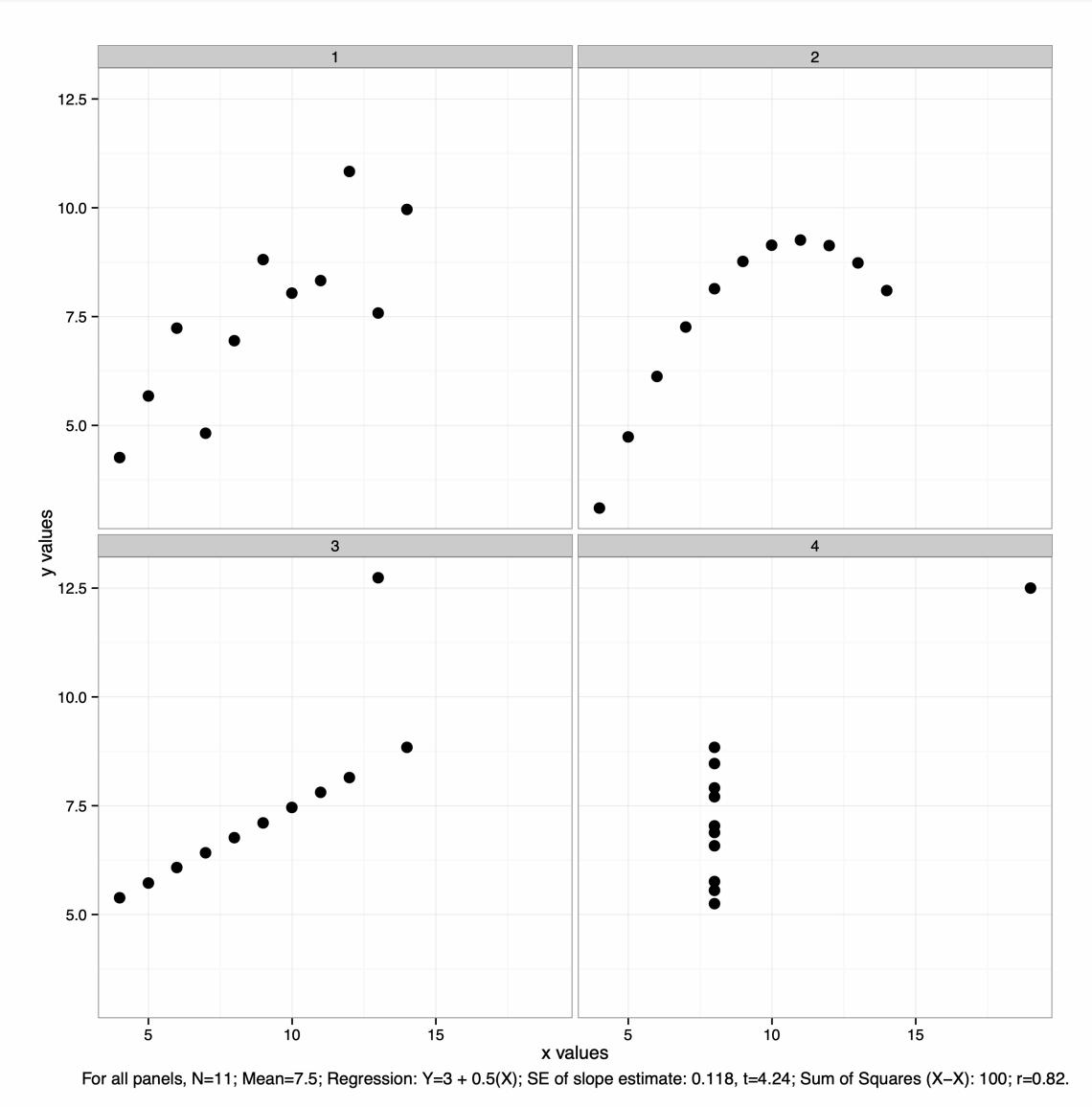
p + geom_point()
```

... And knit your document again.

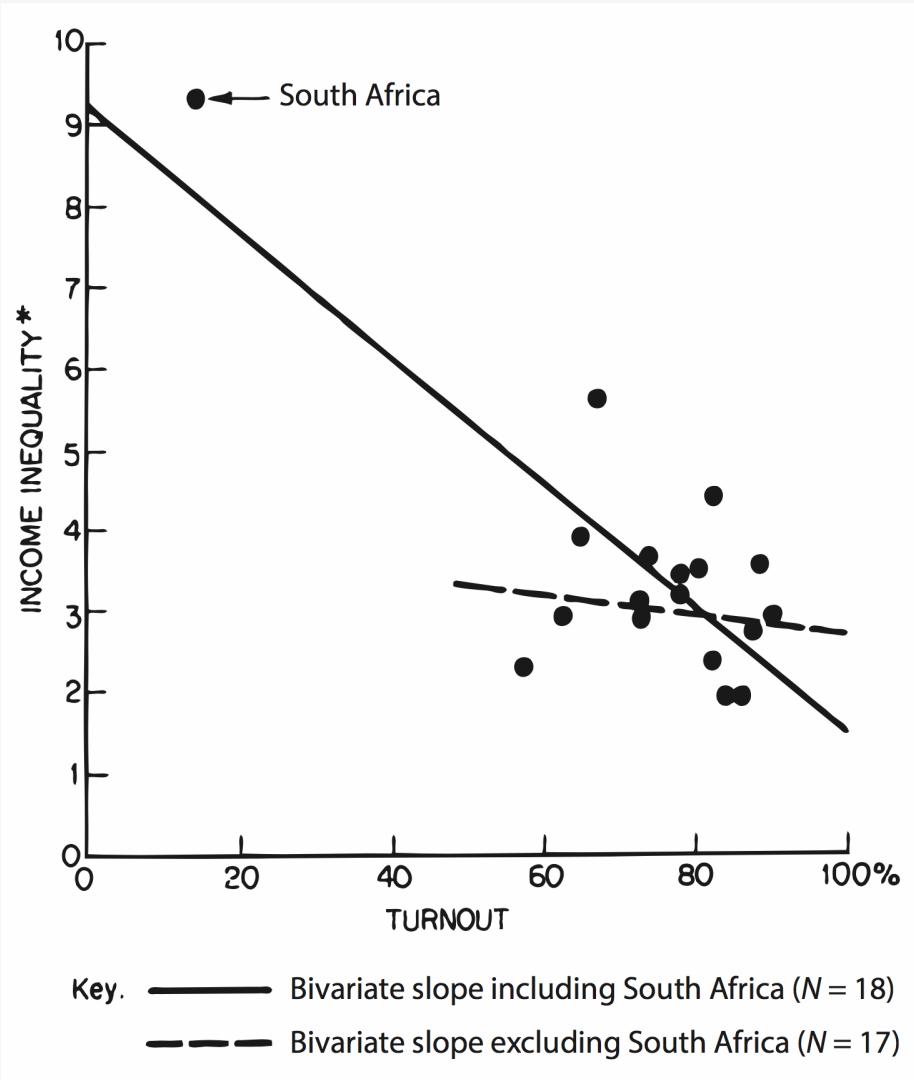


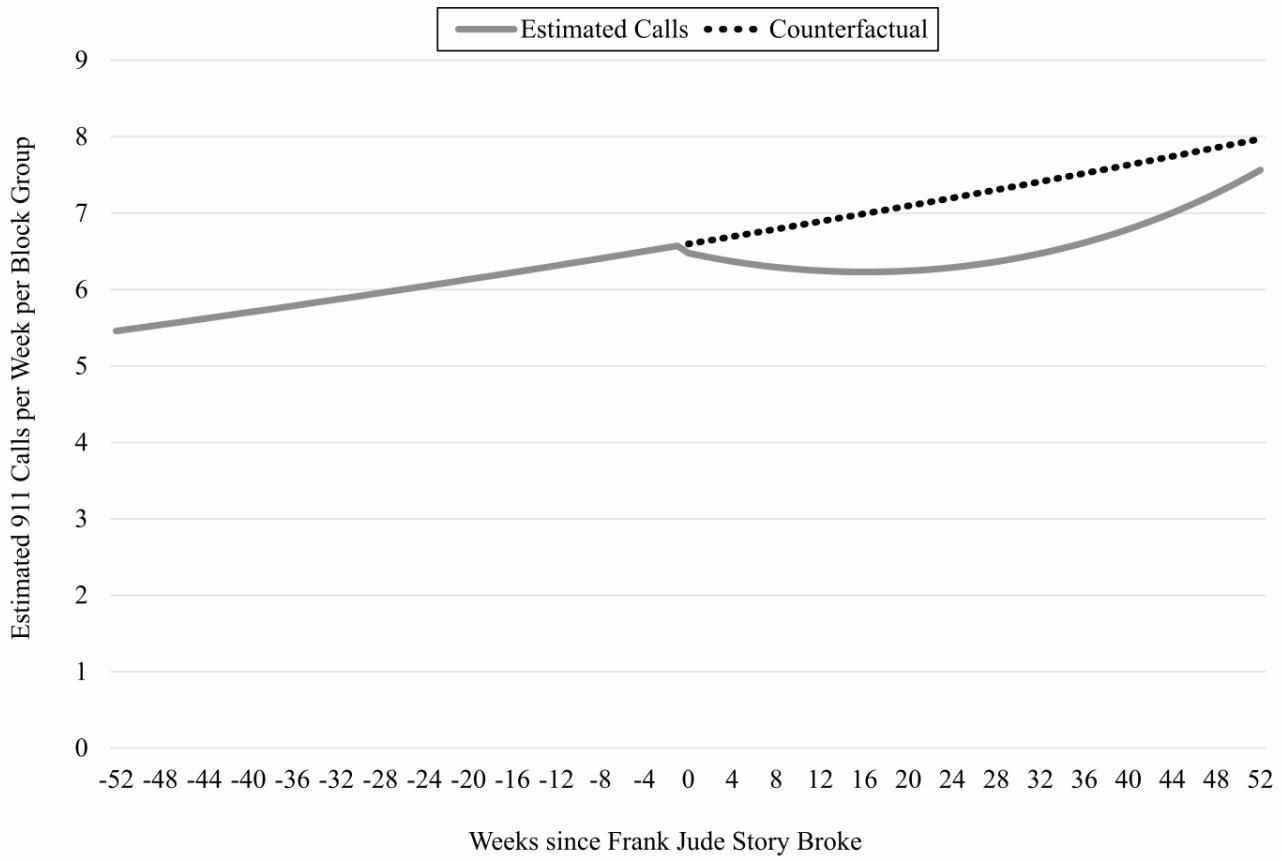
You should
look at
your data

Seeing things

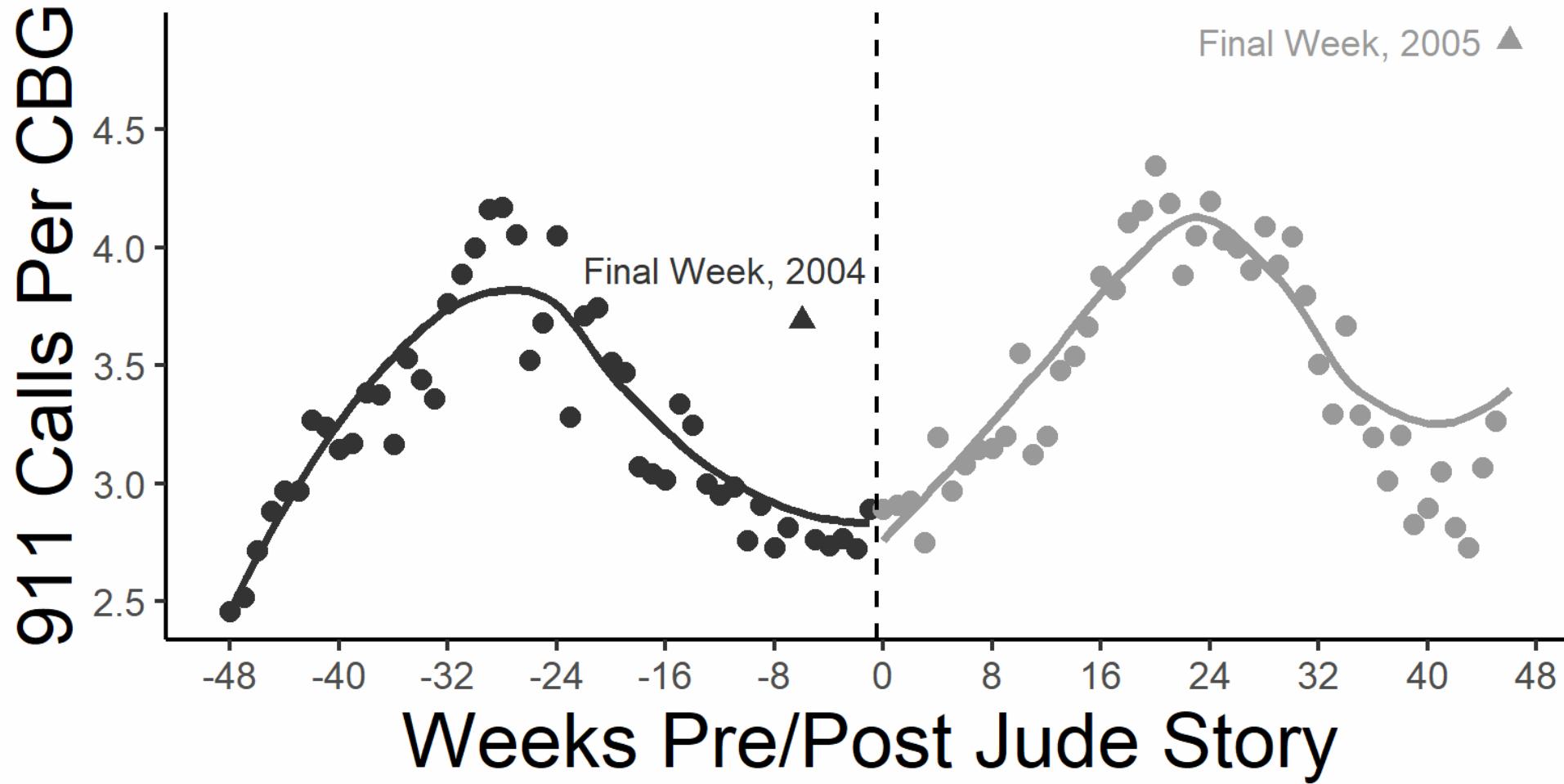


Anscombe's Quartet

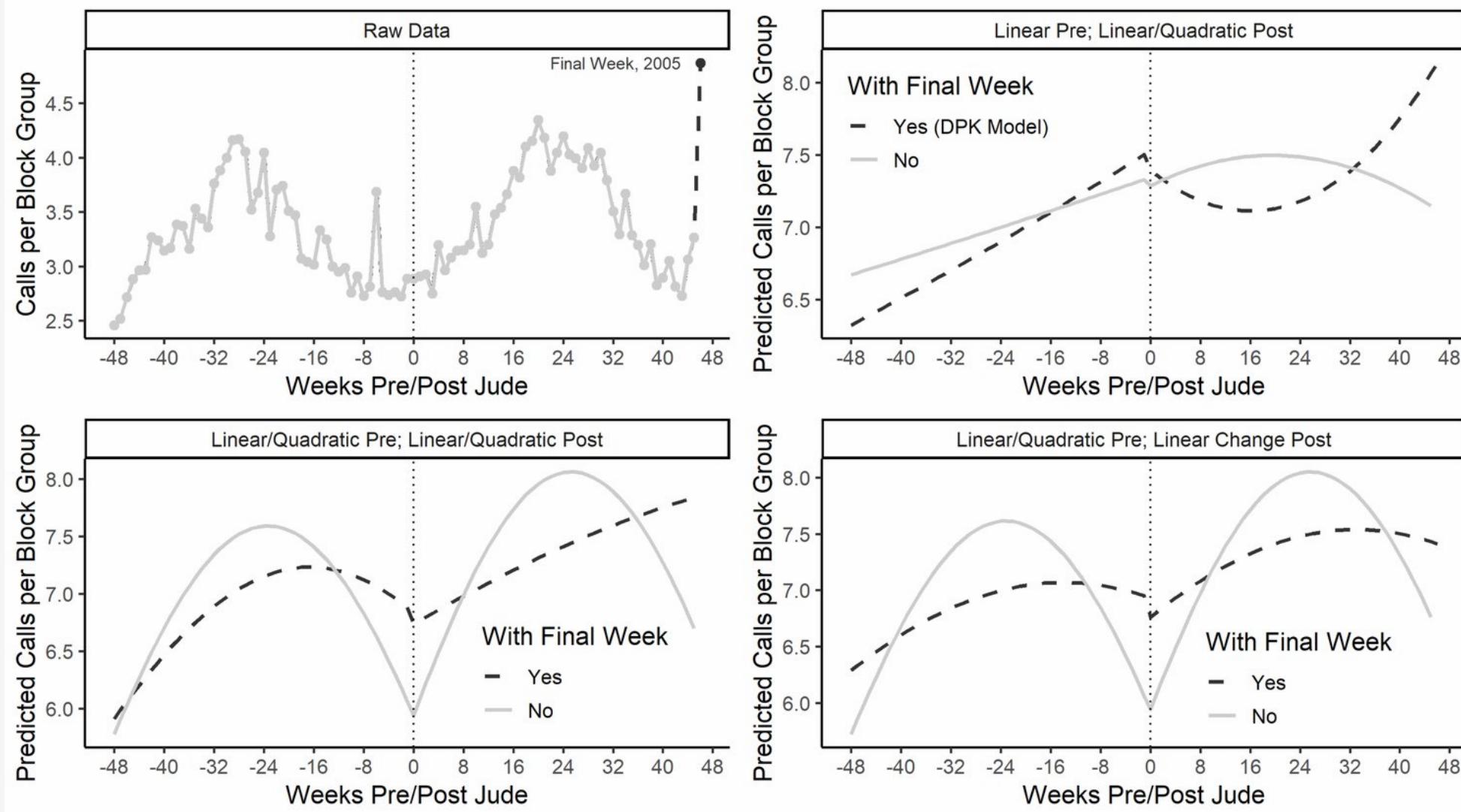


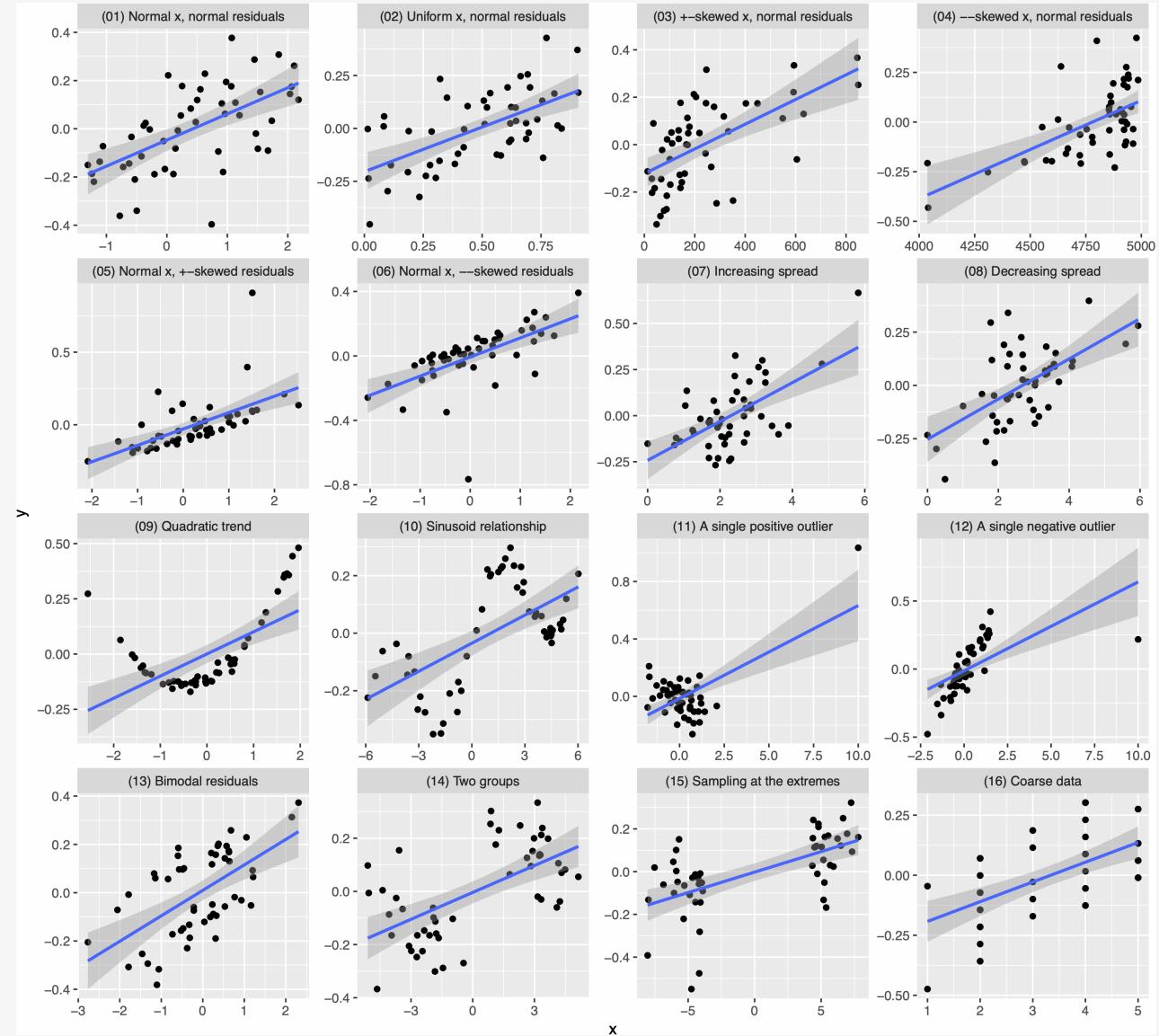


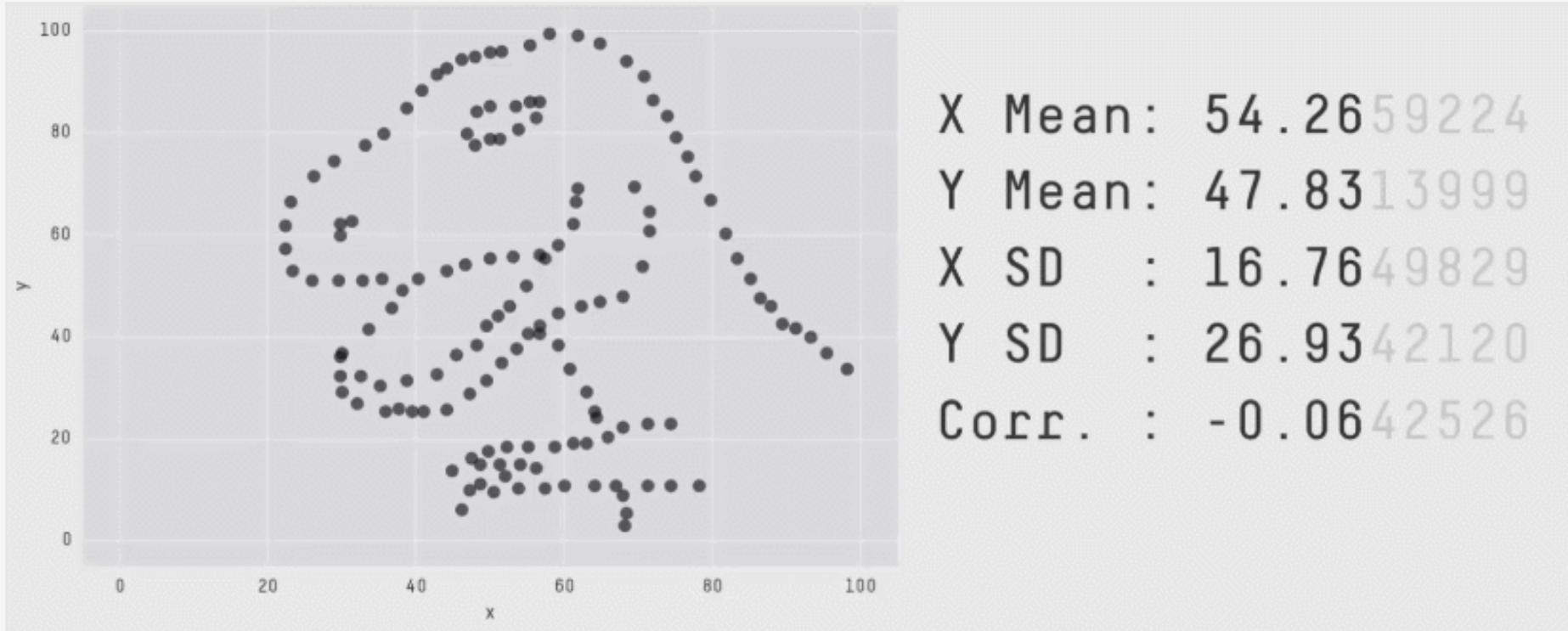
All Neighborhoods



Zoorob (2020)

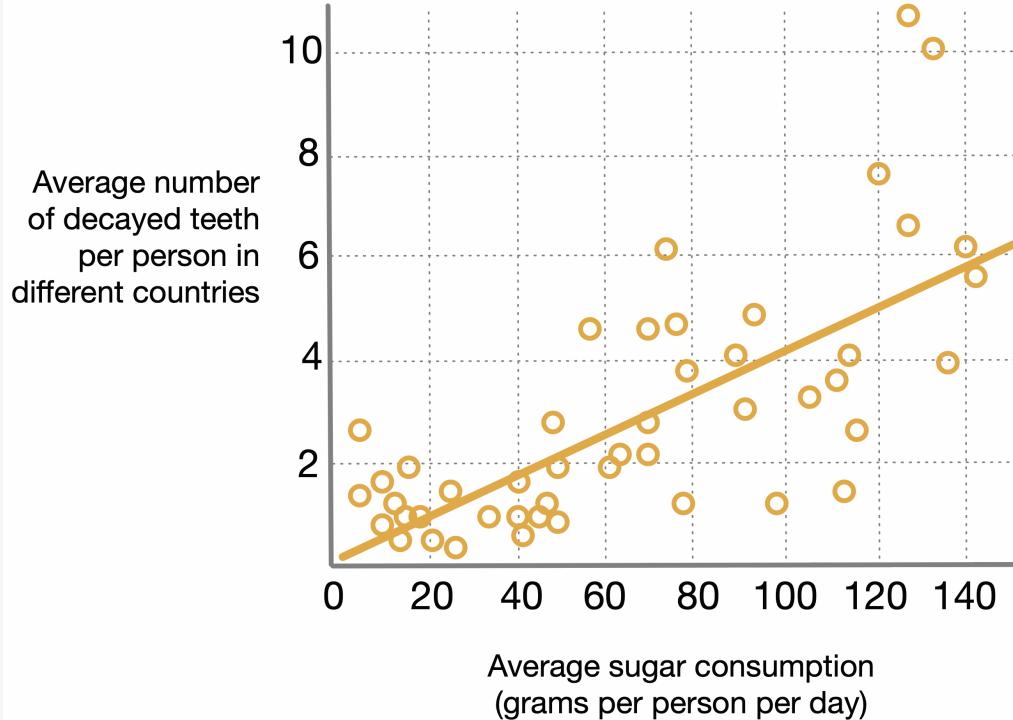






Cairo; Matejka & Fitzmaurice

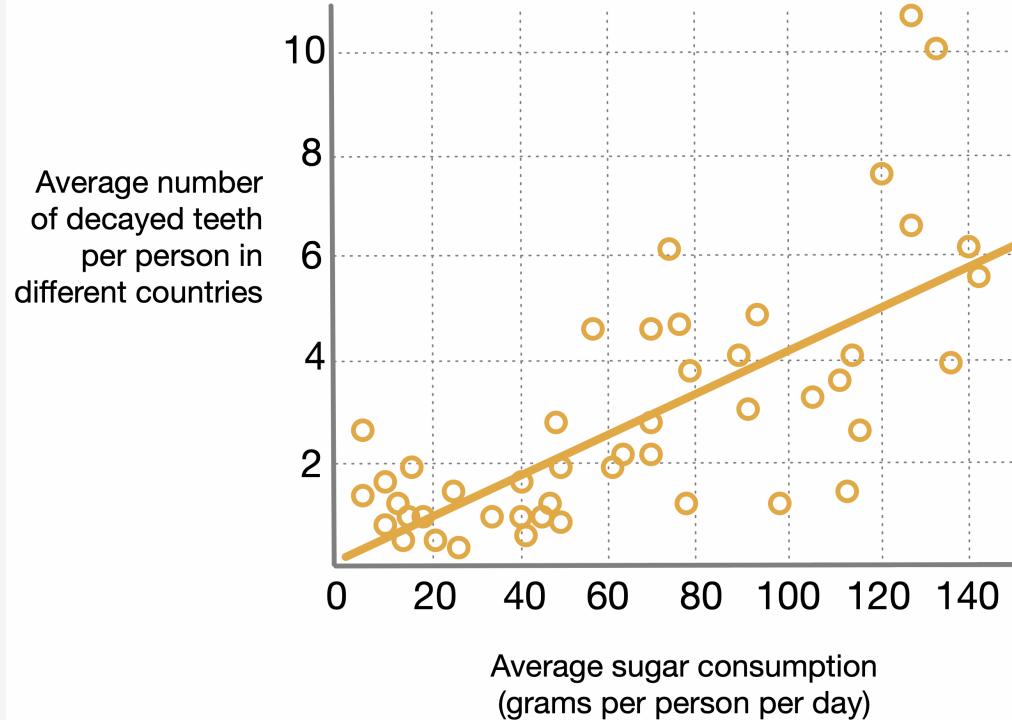
Which of the following statements best describes the data in the graph below?



Pew Research

- 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

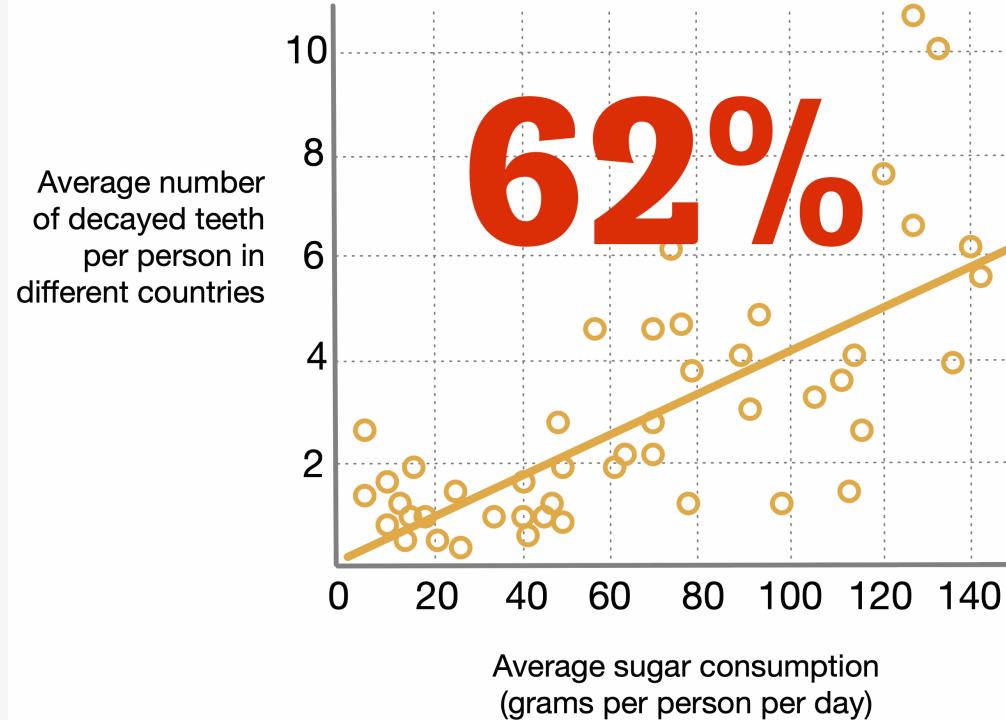
Which of the following statements best describes the data in the graph below?



Pew Research

- 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

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

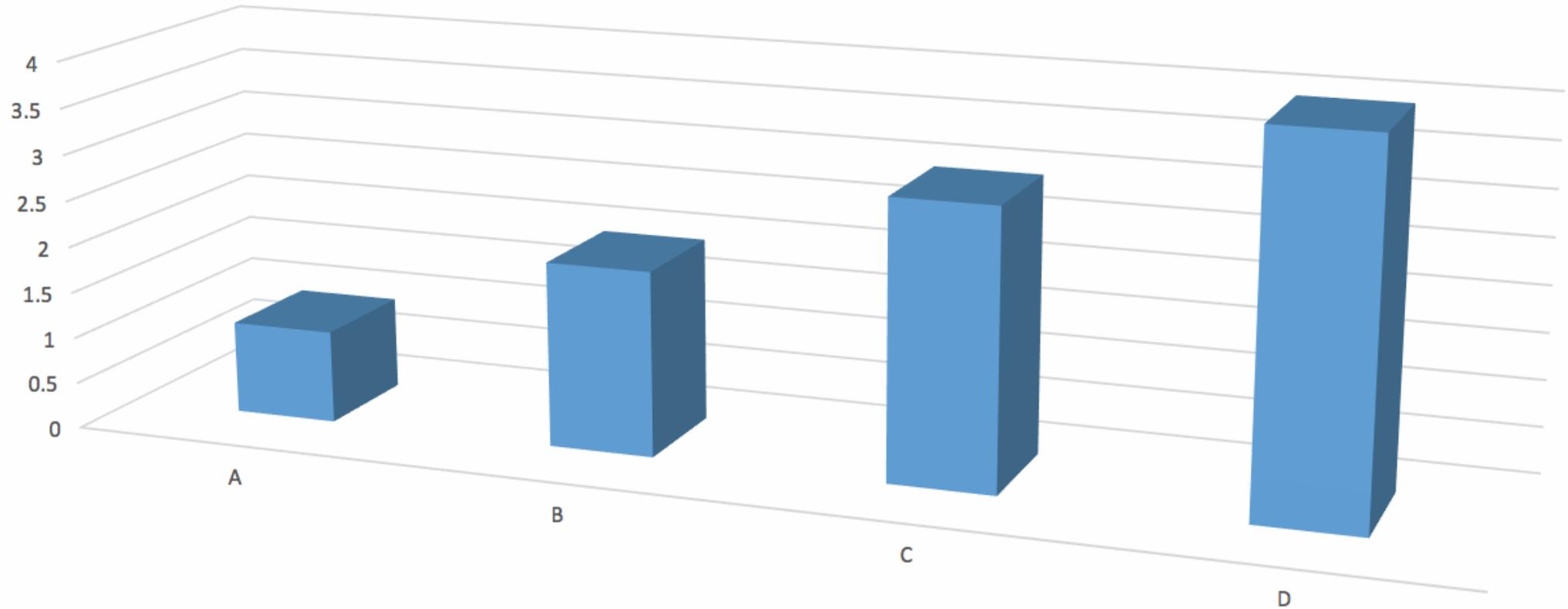


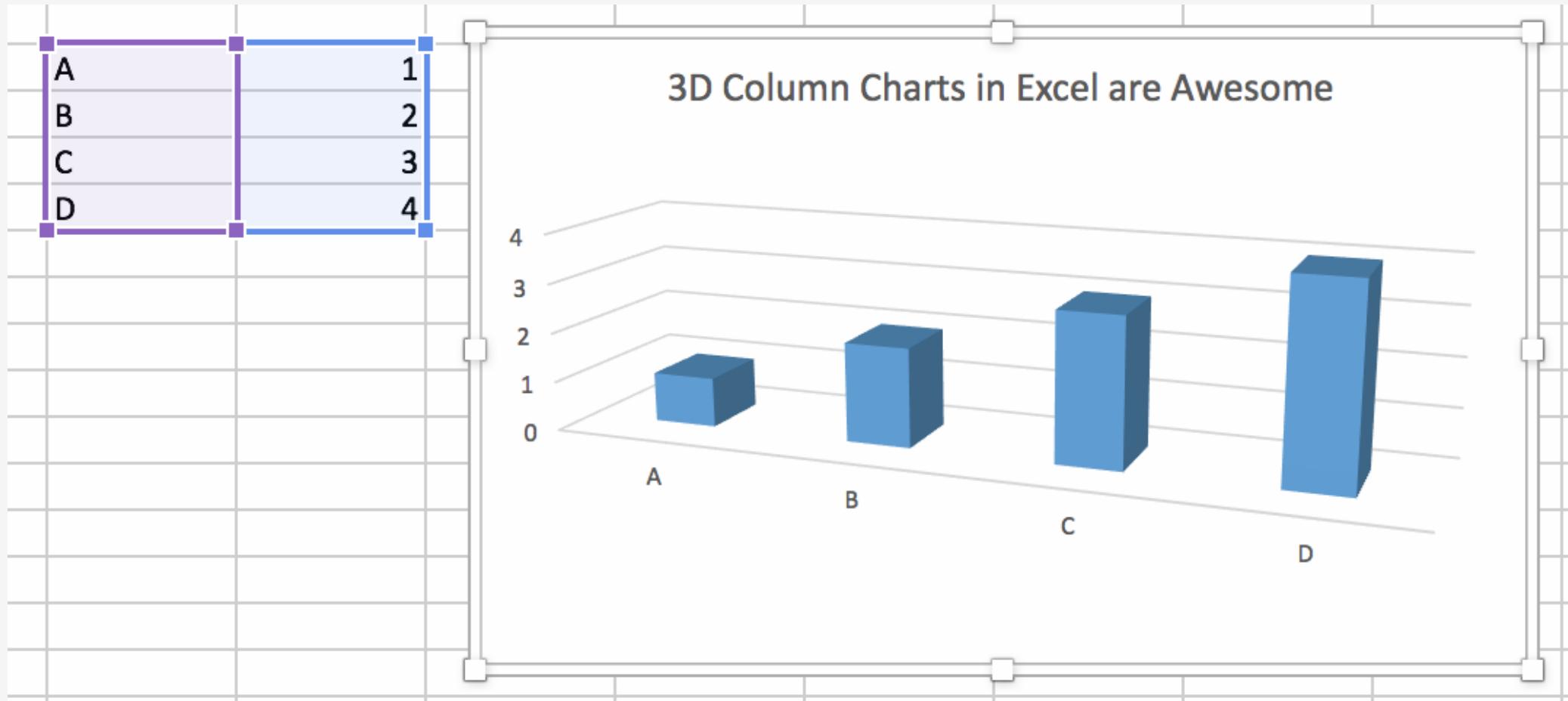
Pew Research

- 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





Bad Taste

Bad Taste

Bad Data

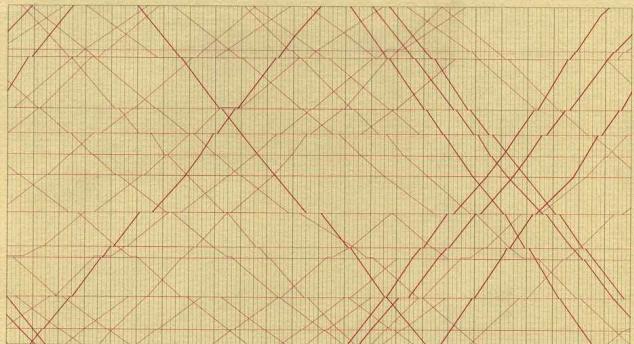
Bad Taste

Bad Data

Bad Perception

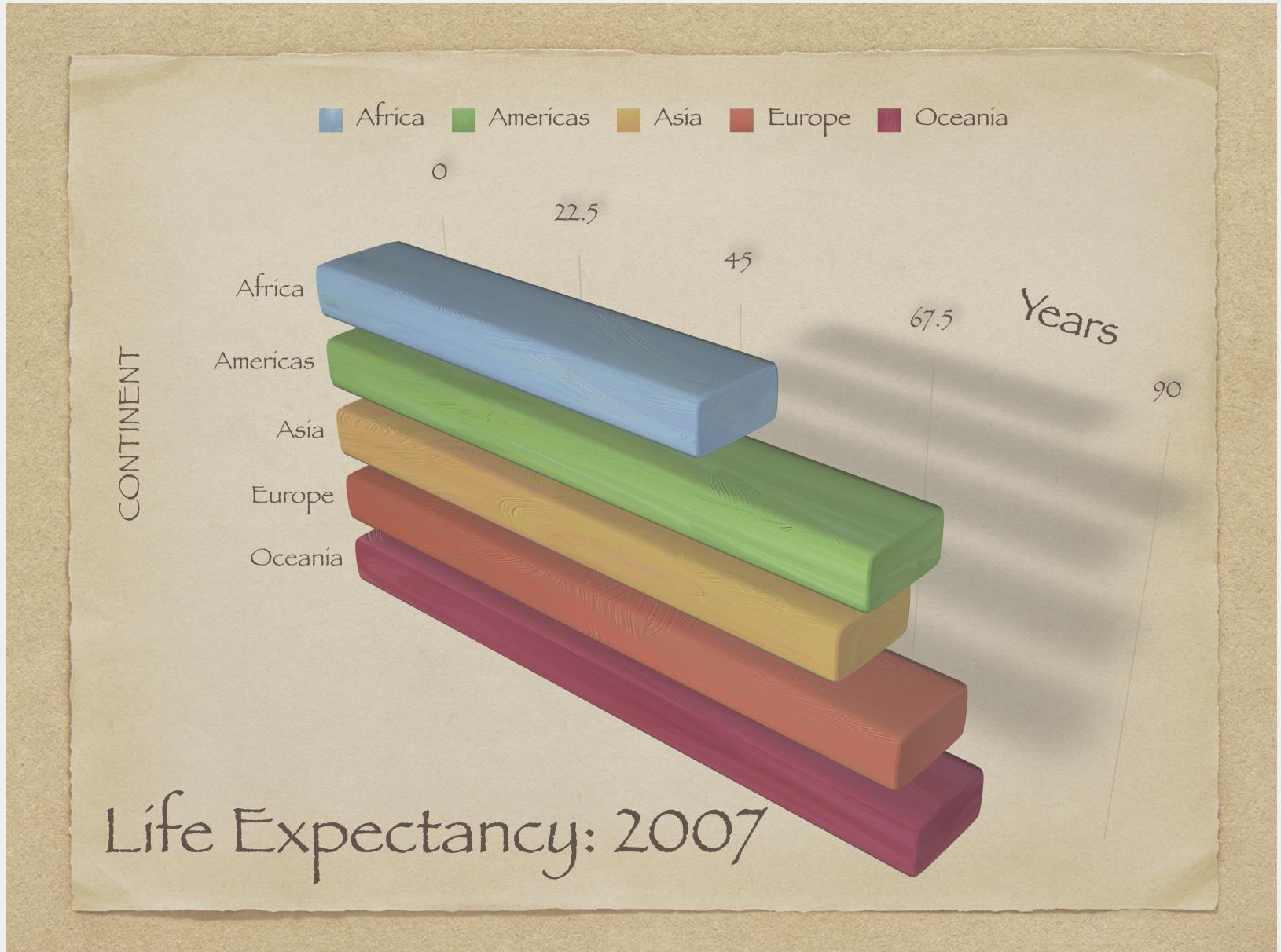
Taste: Simplify, Simplify?

Tufte's "Data to Ink Ratio"



The Visual Display
of Quantitative Information

EDWARD R. TUFTE

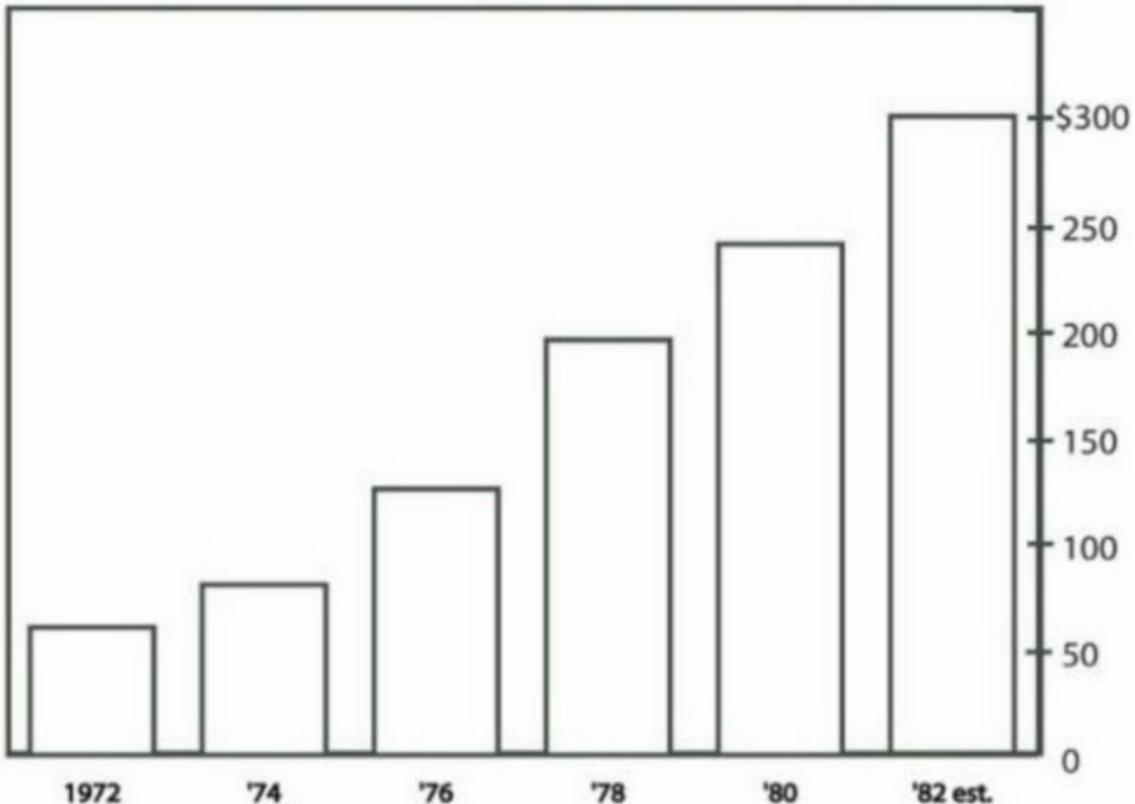


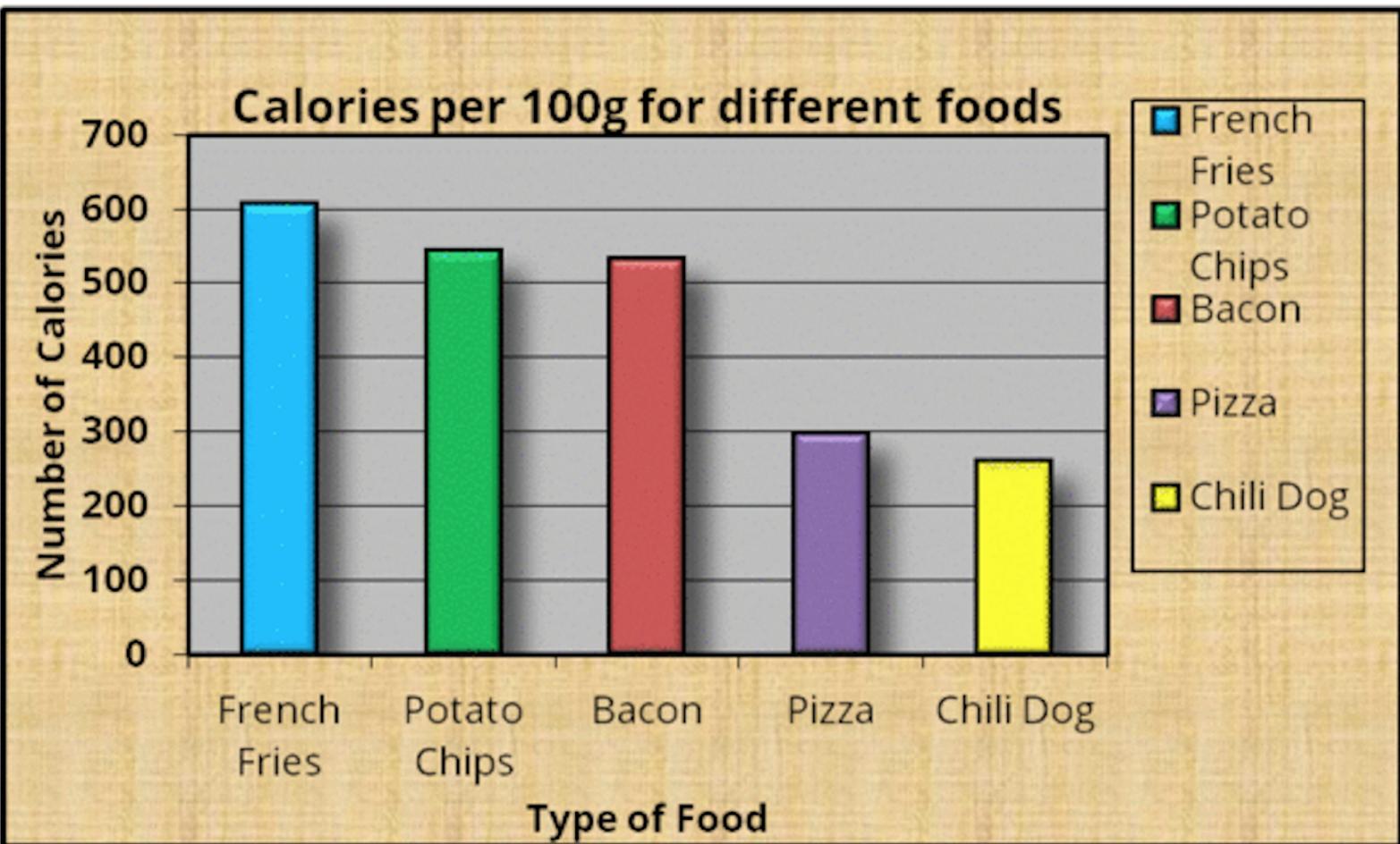
MONSTROUS COSTS

Total House and Senate campaign expenditures,
in millions



MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions



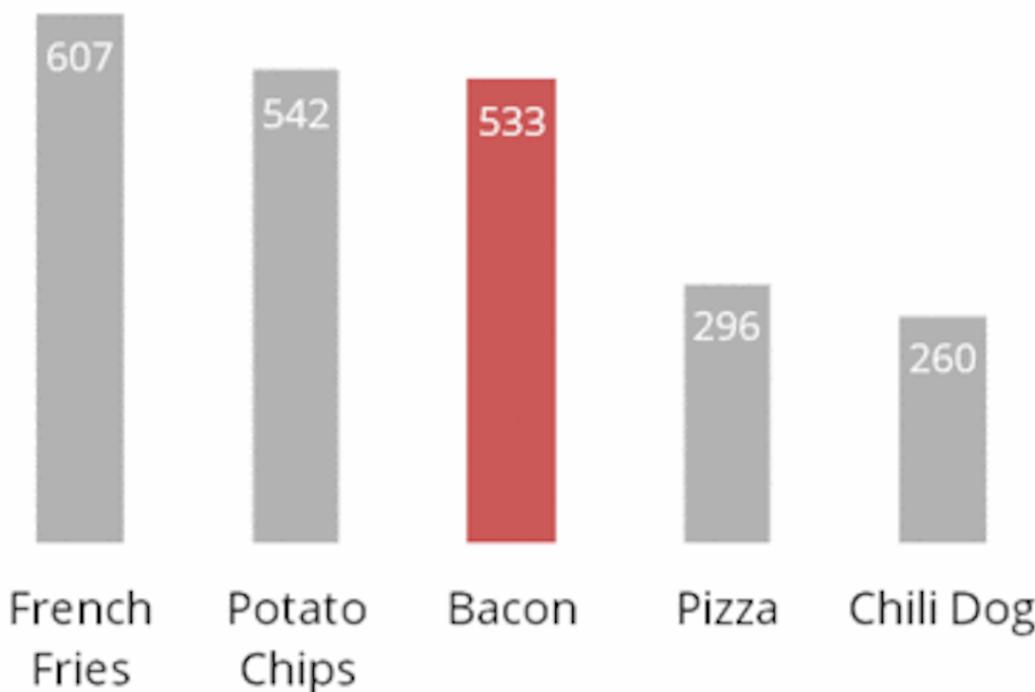


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Darkhorse Analytics

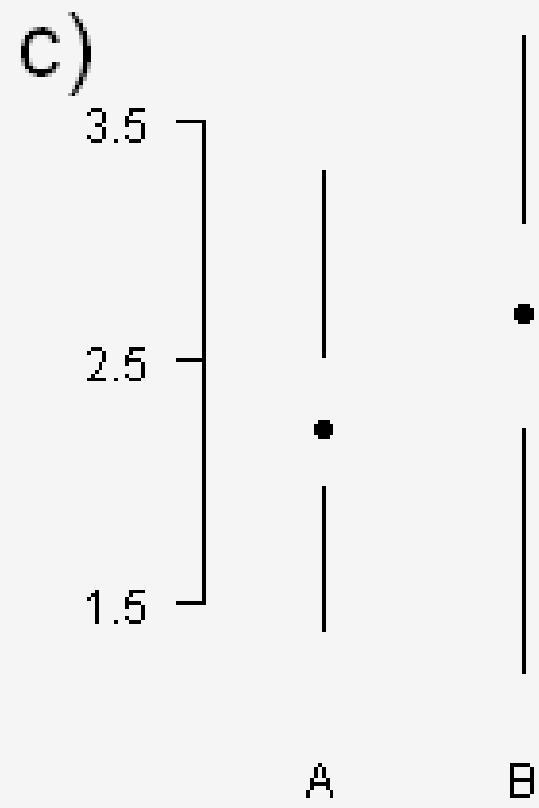
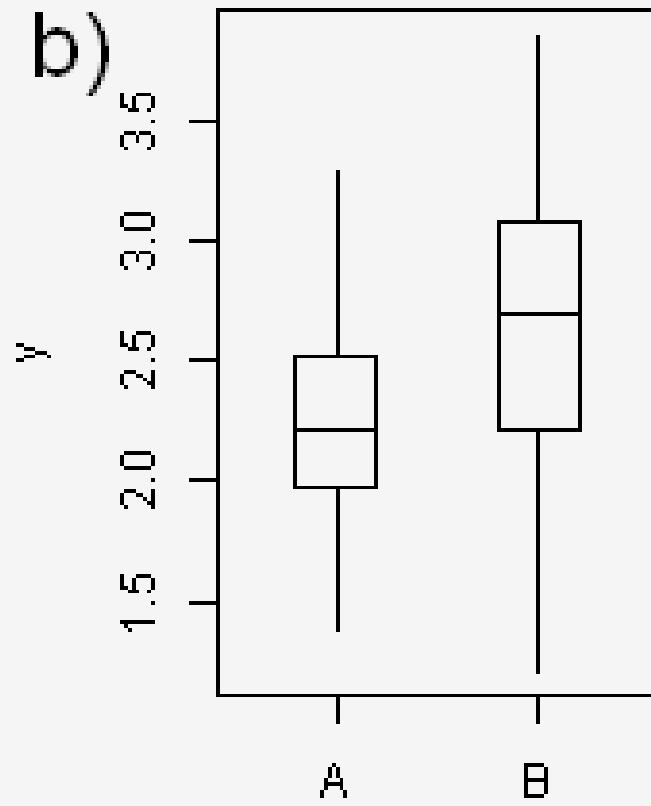
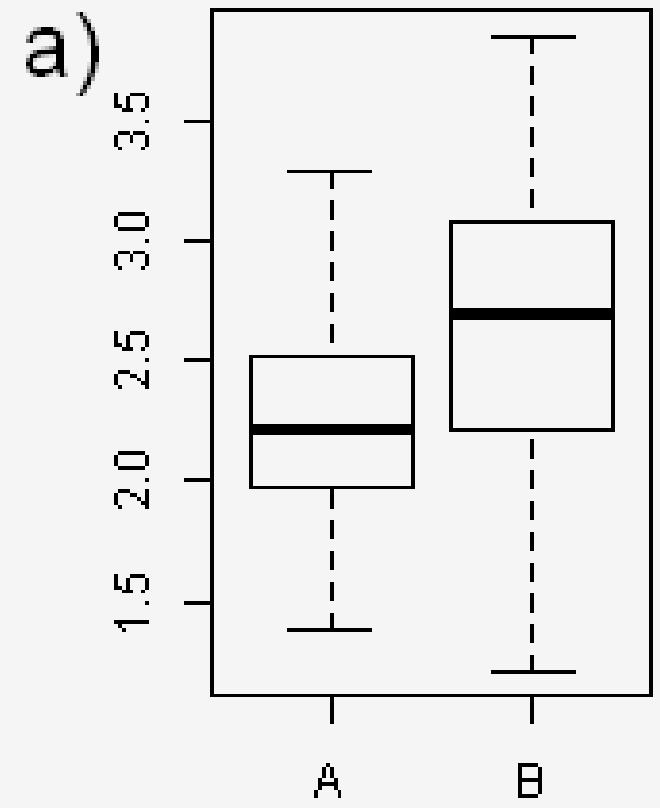
Calories per 100g



Created by Darkhorse Analytics

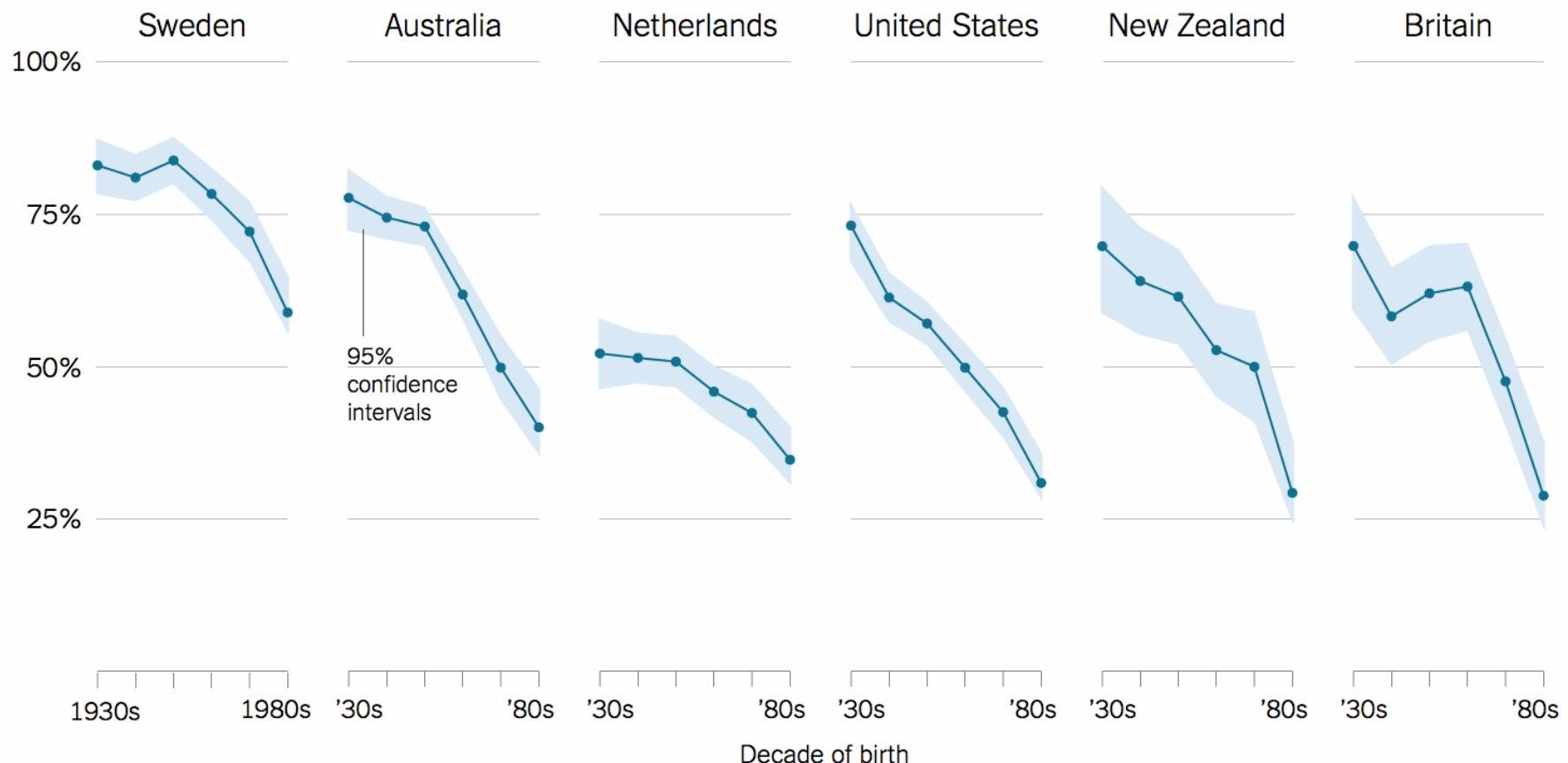
www.darkhorseanalytics.com

Darkhorse Analytics

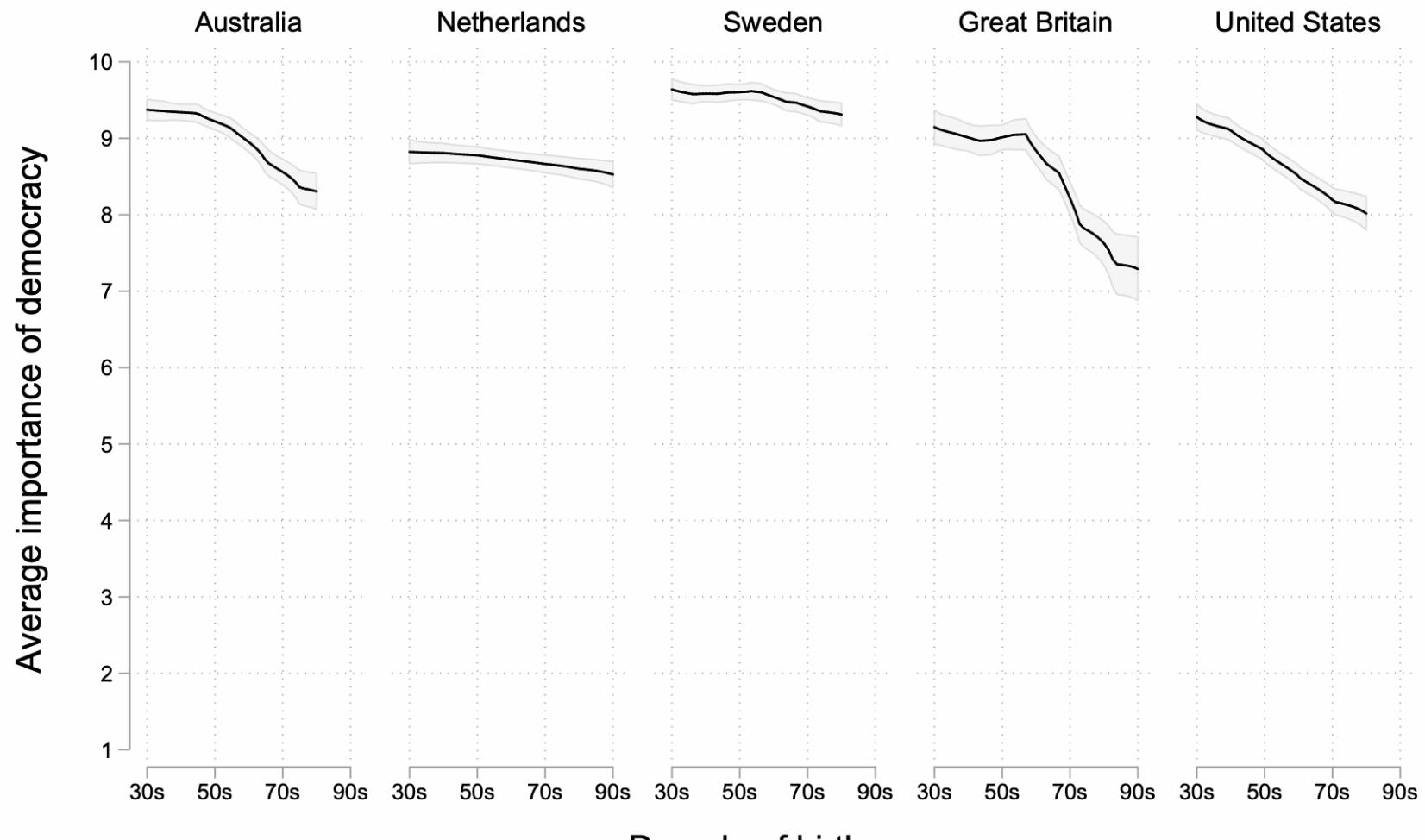


Data: **Junk-Free Junk Charts**

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



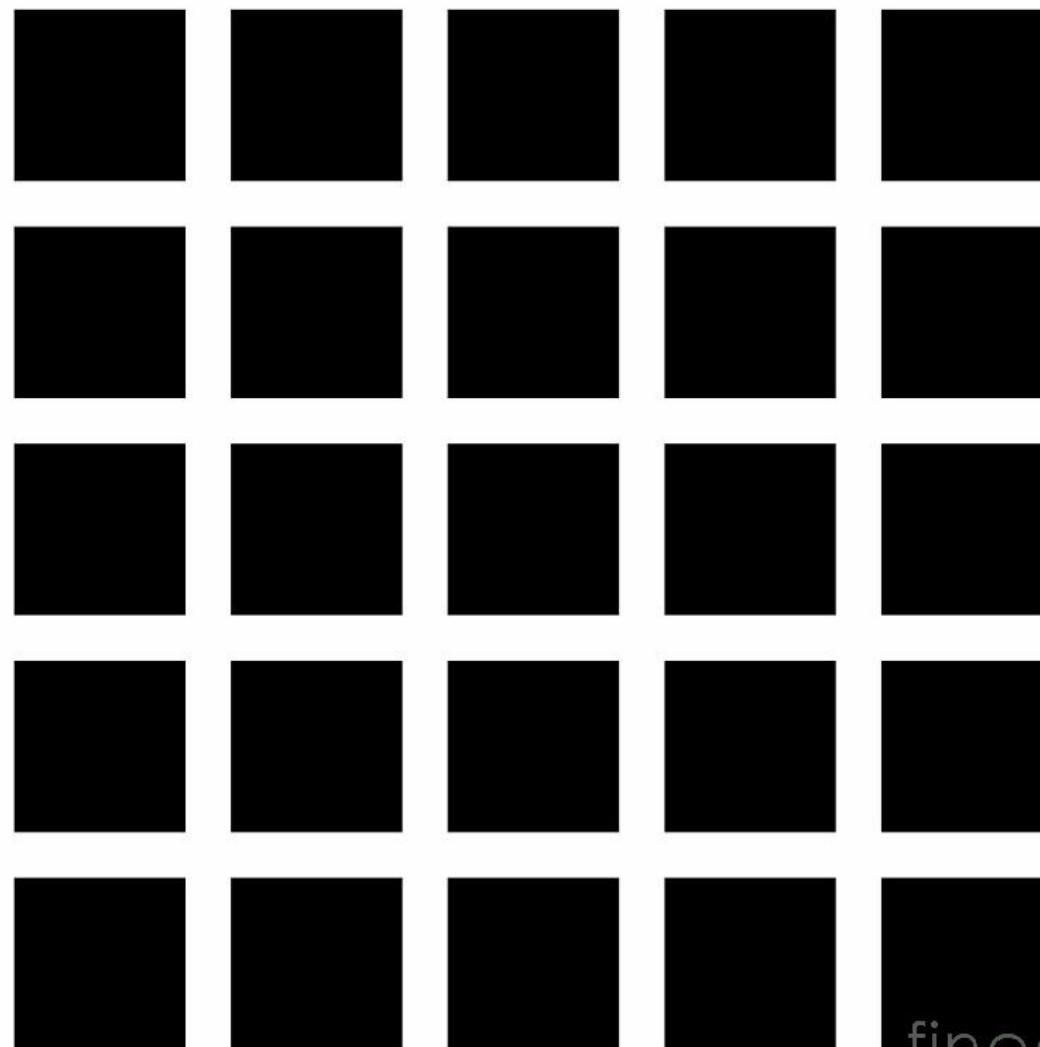
Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times



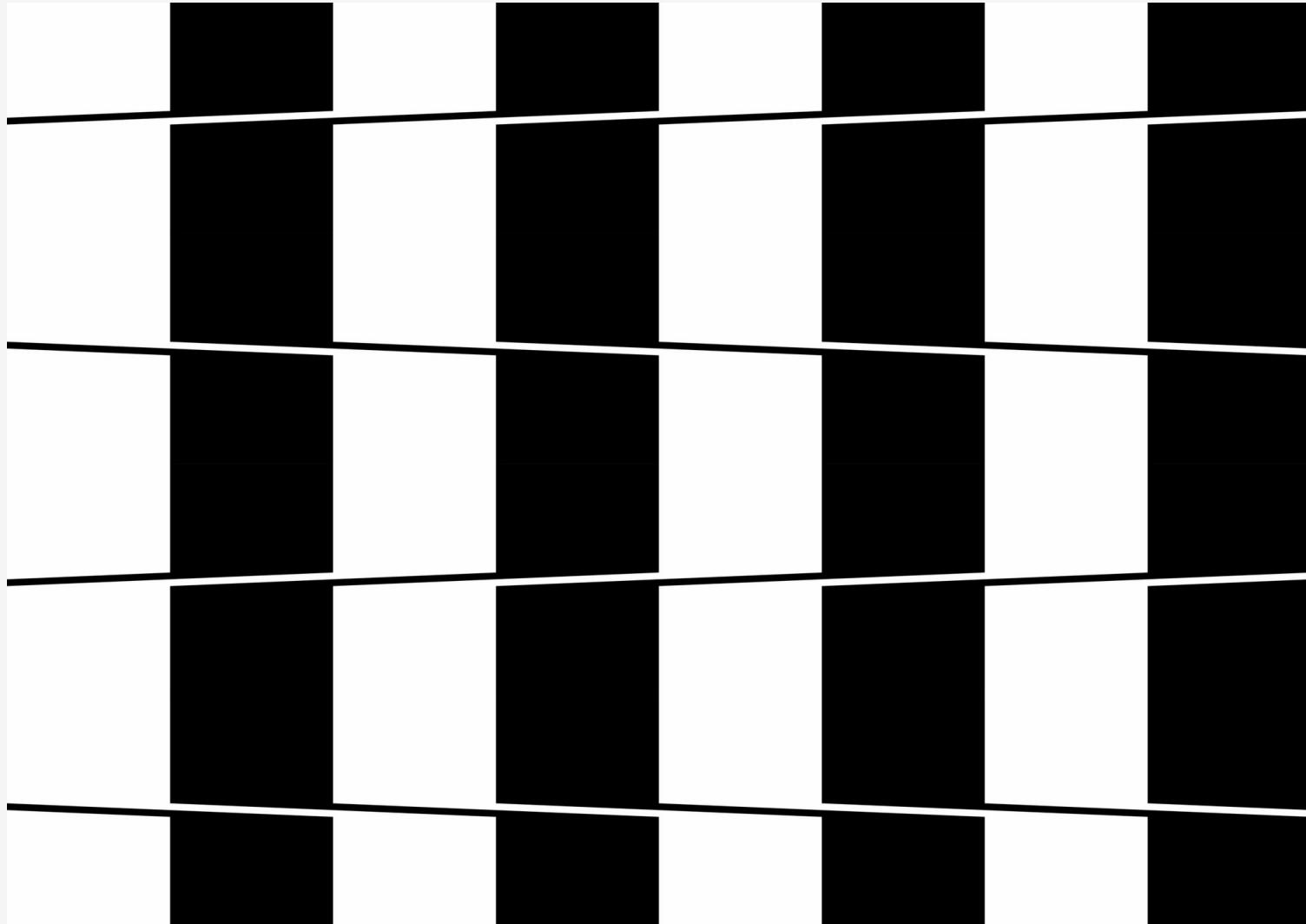
Graph by Erik Voeten, based on WVS 5

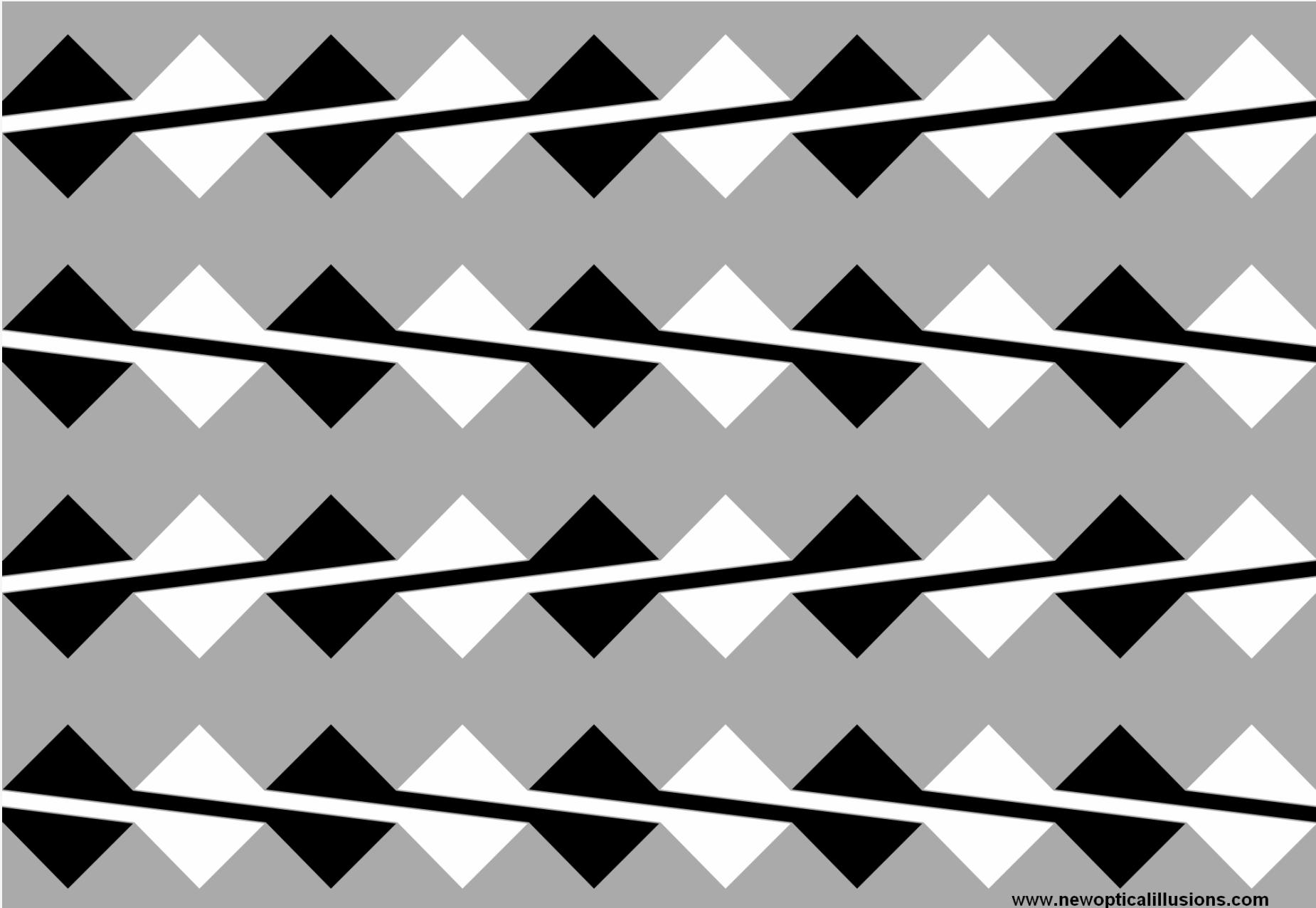
Perception: Seeing and Not Seeing

Edges & Contrasts



fine
© 2010

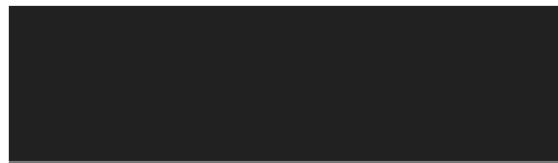




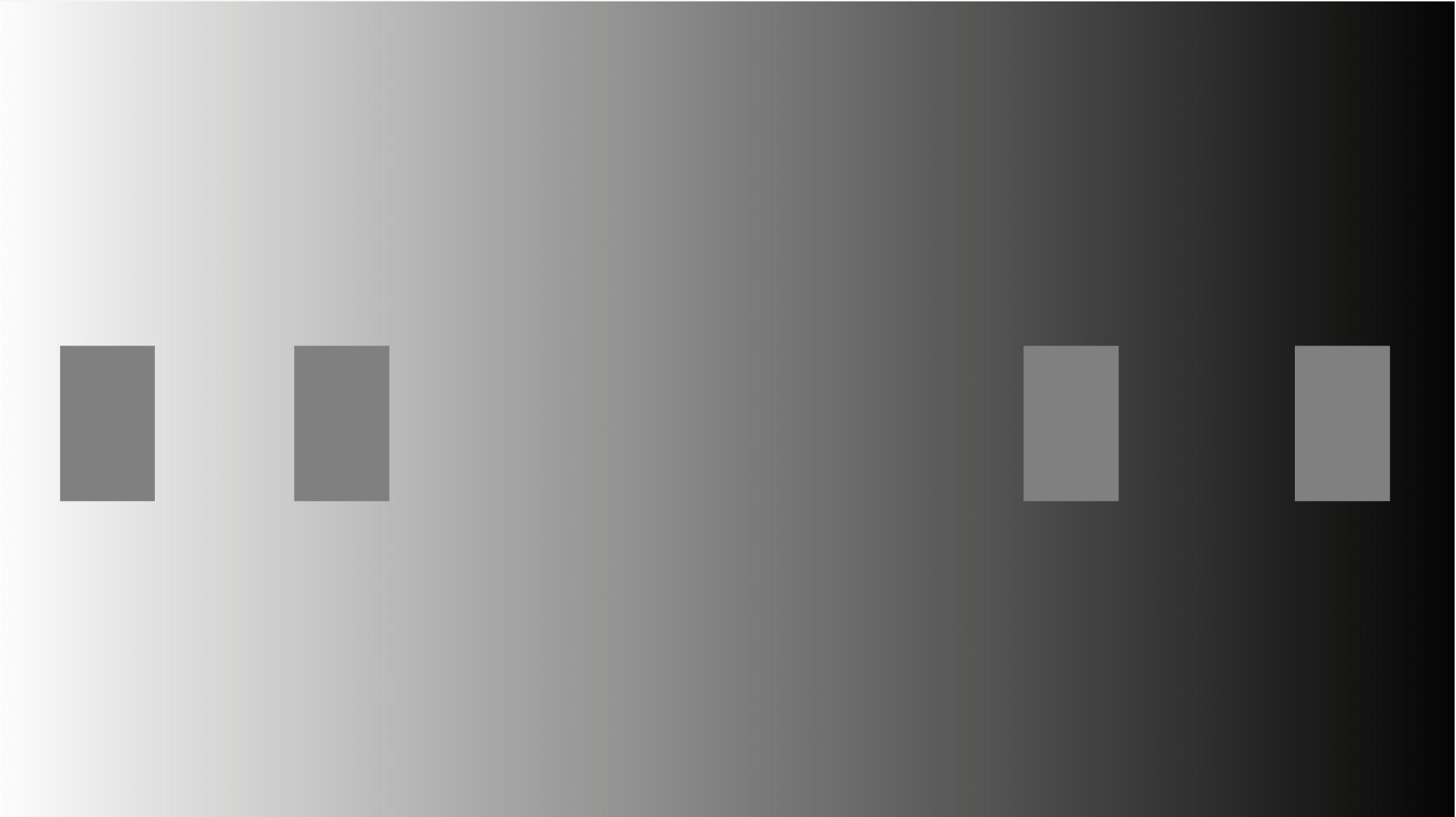


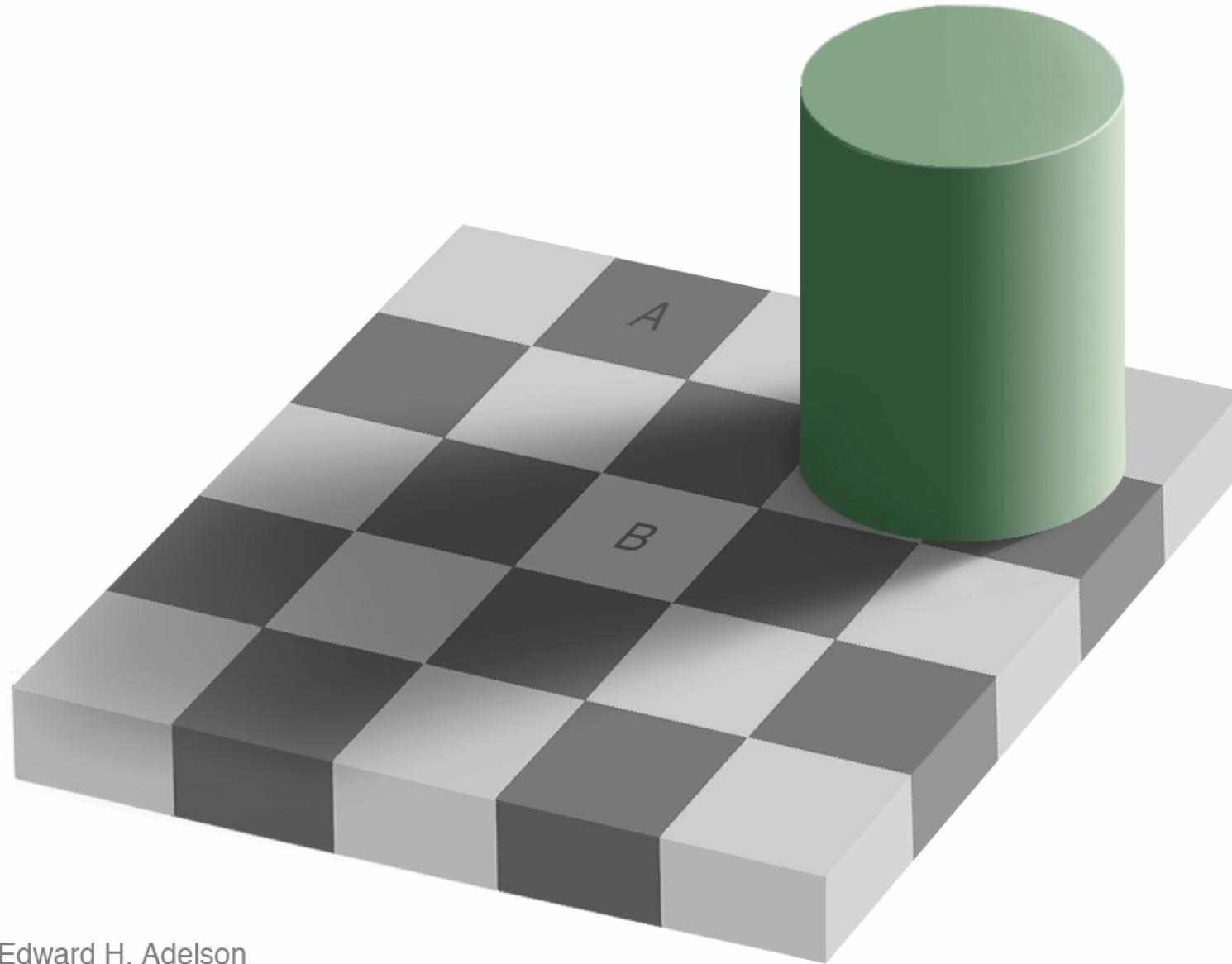
These are two perfectly
geometrical circles



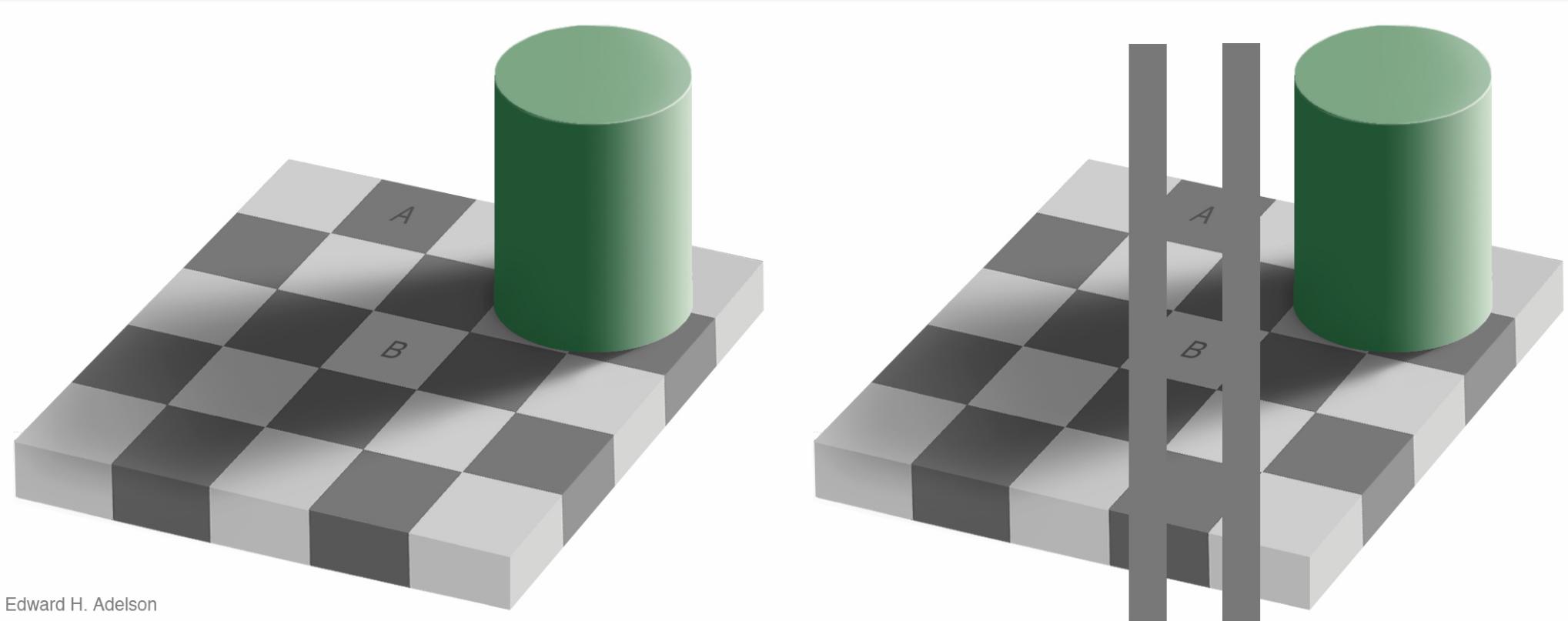








Edward H. Adelson

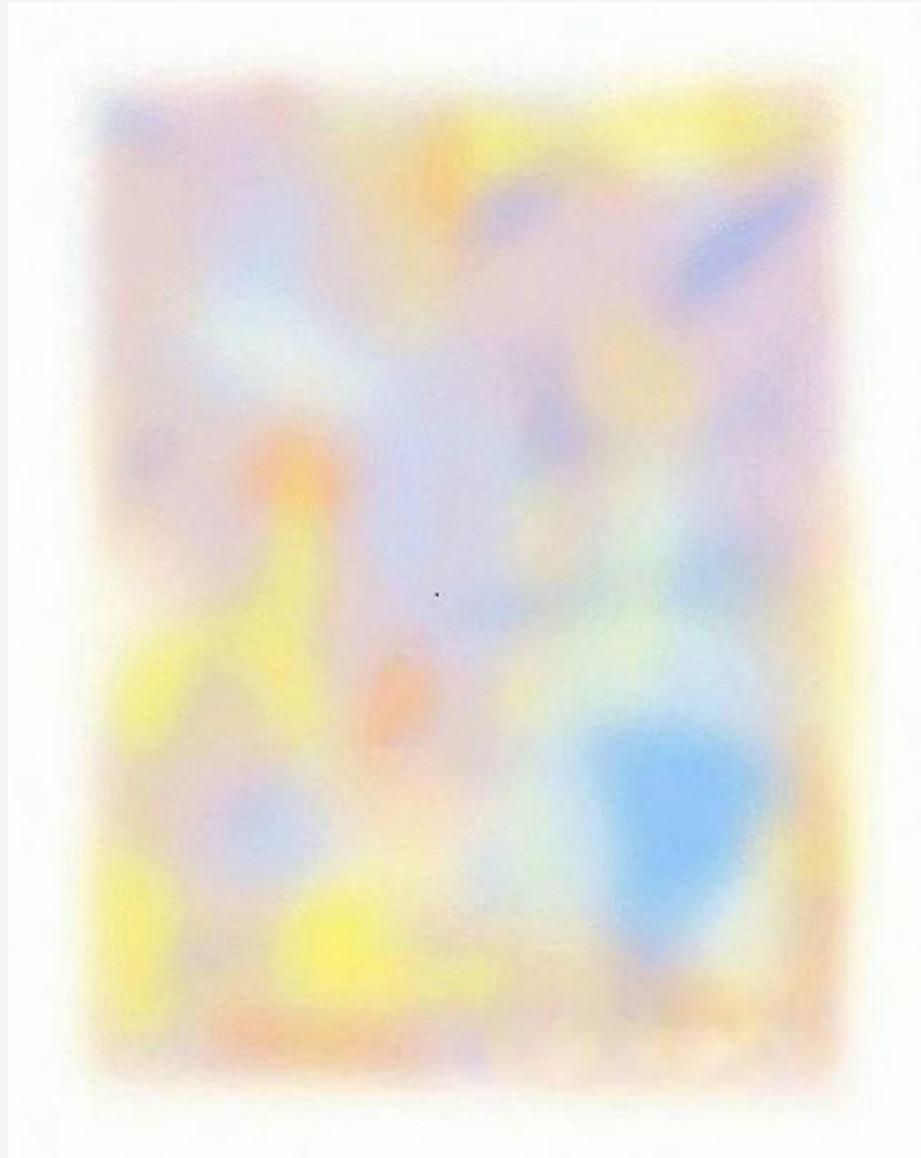


Edward Adelson

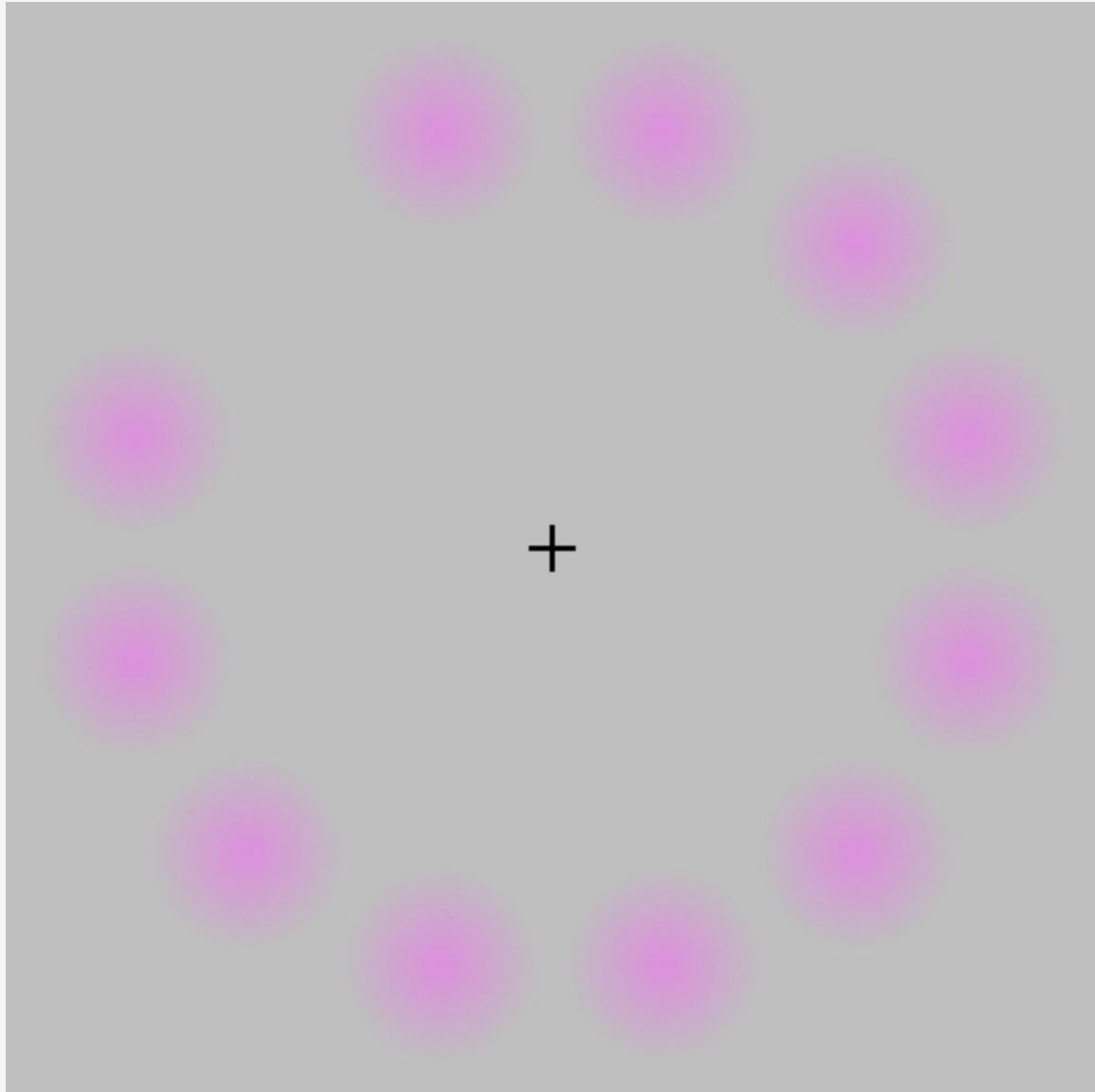


Edward Adelson

Luminance and Color

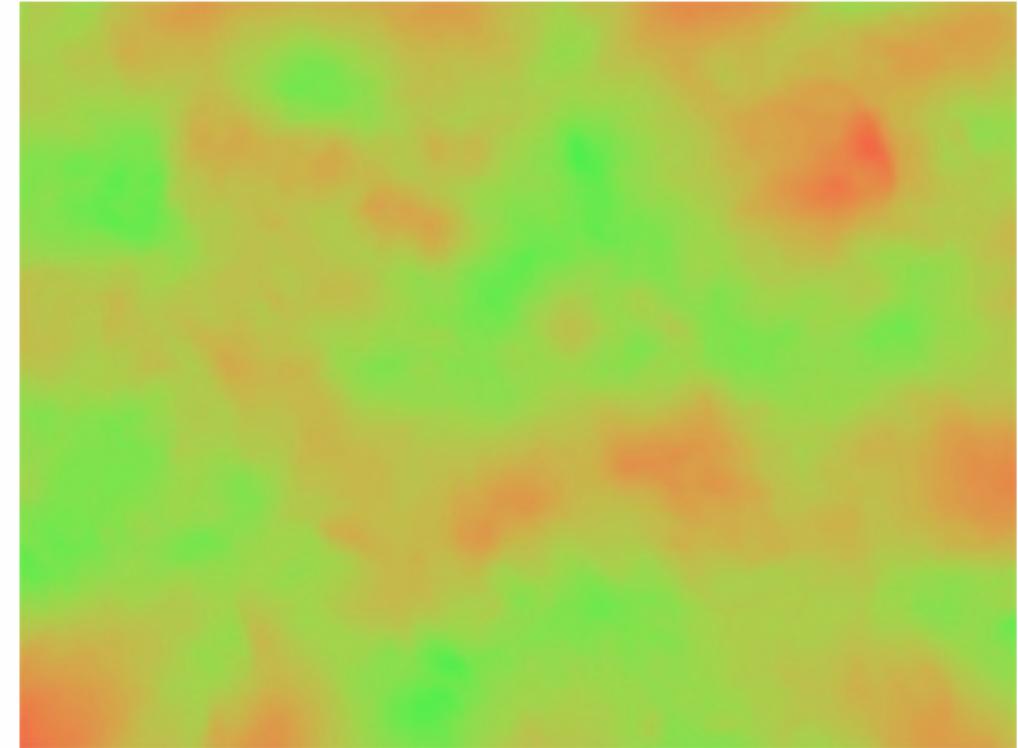


Troxler effect

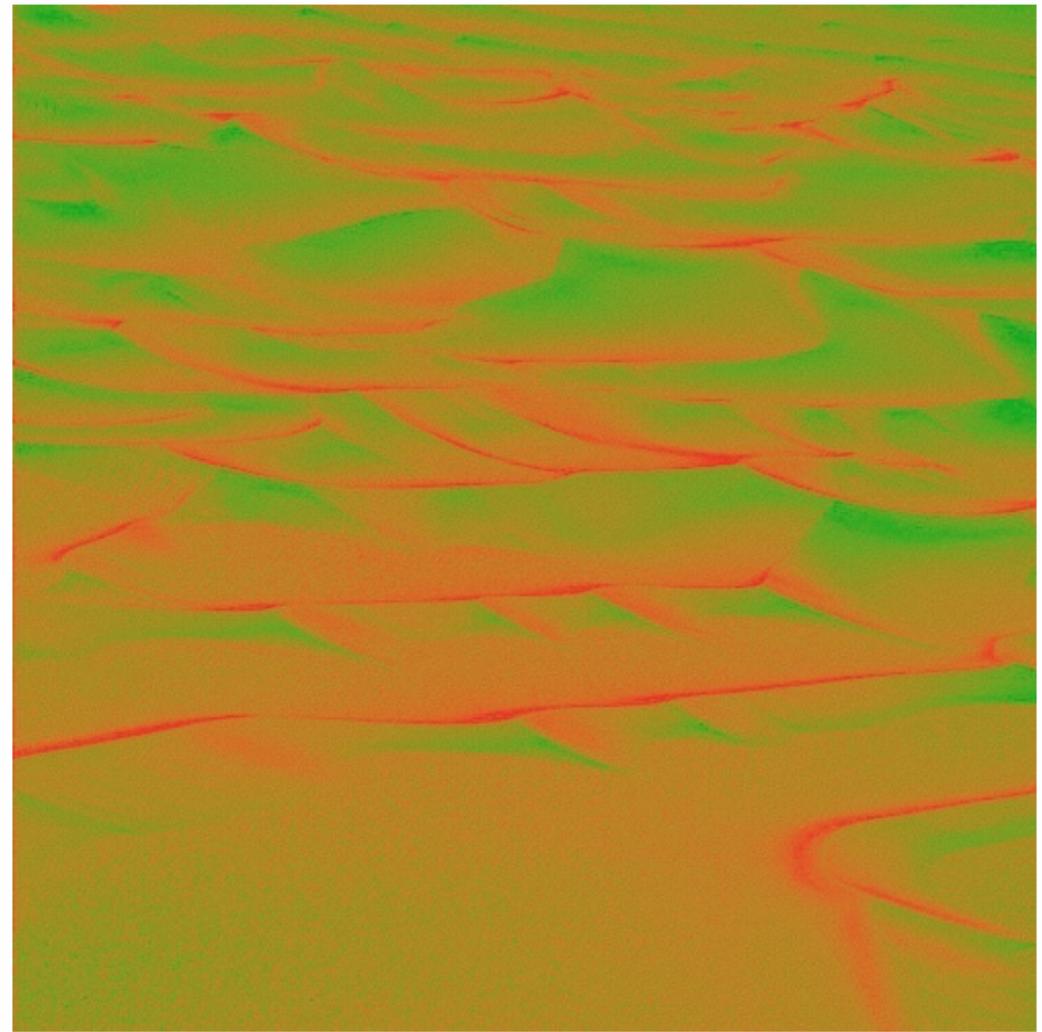
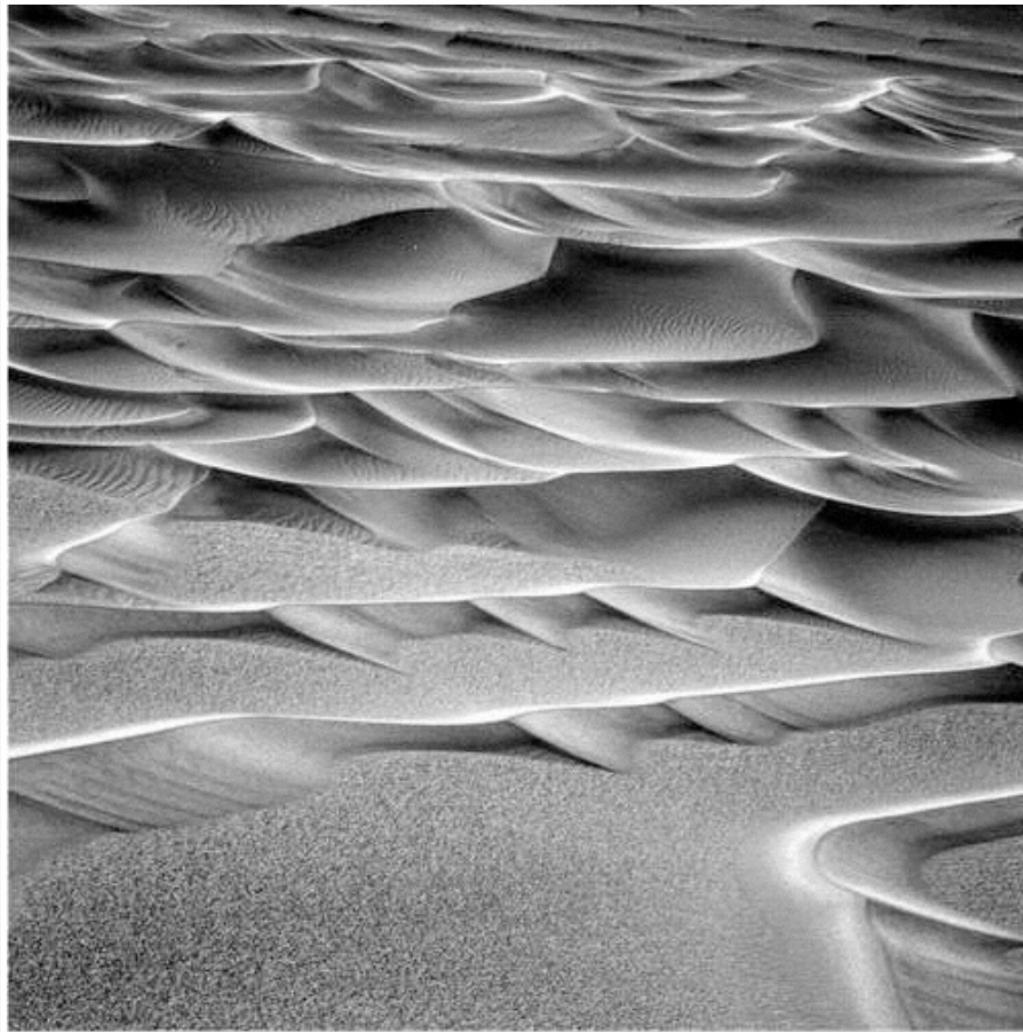


+

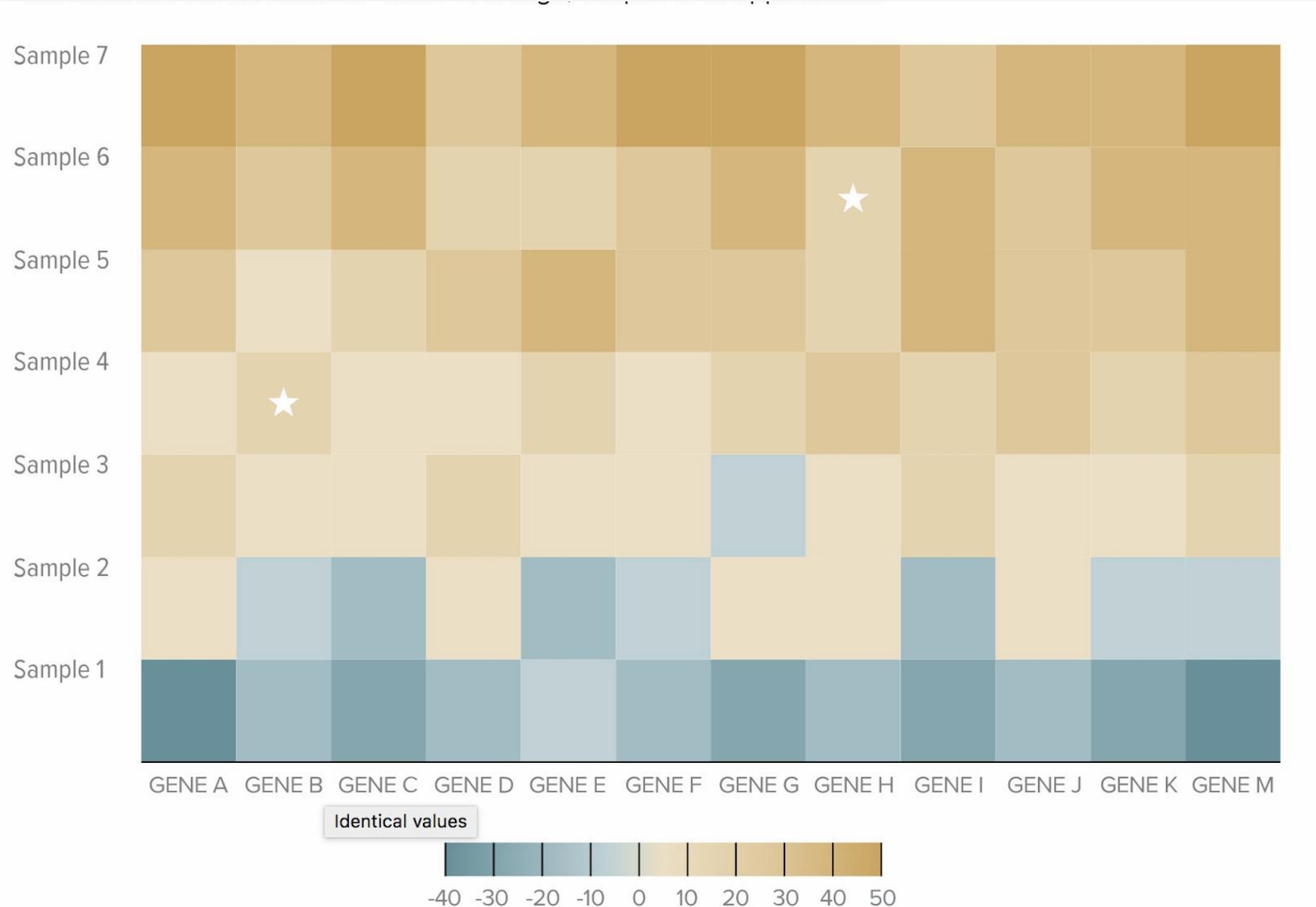
"Lilac Chaser"



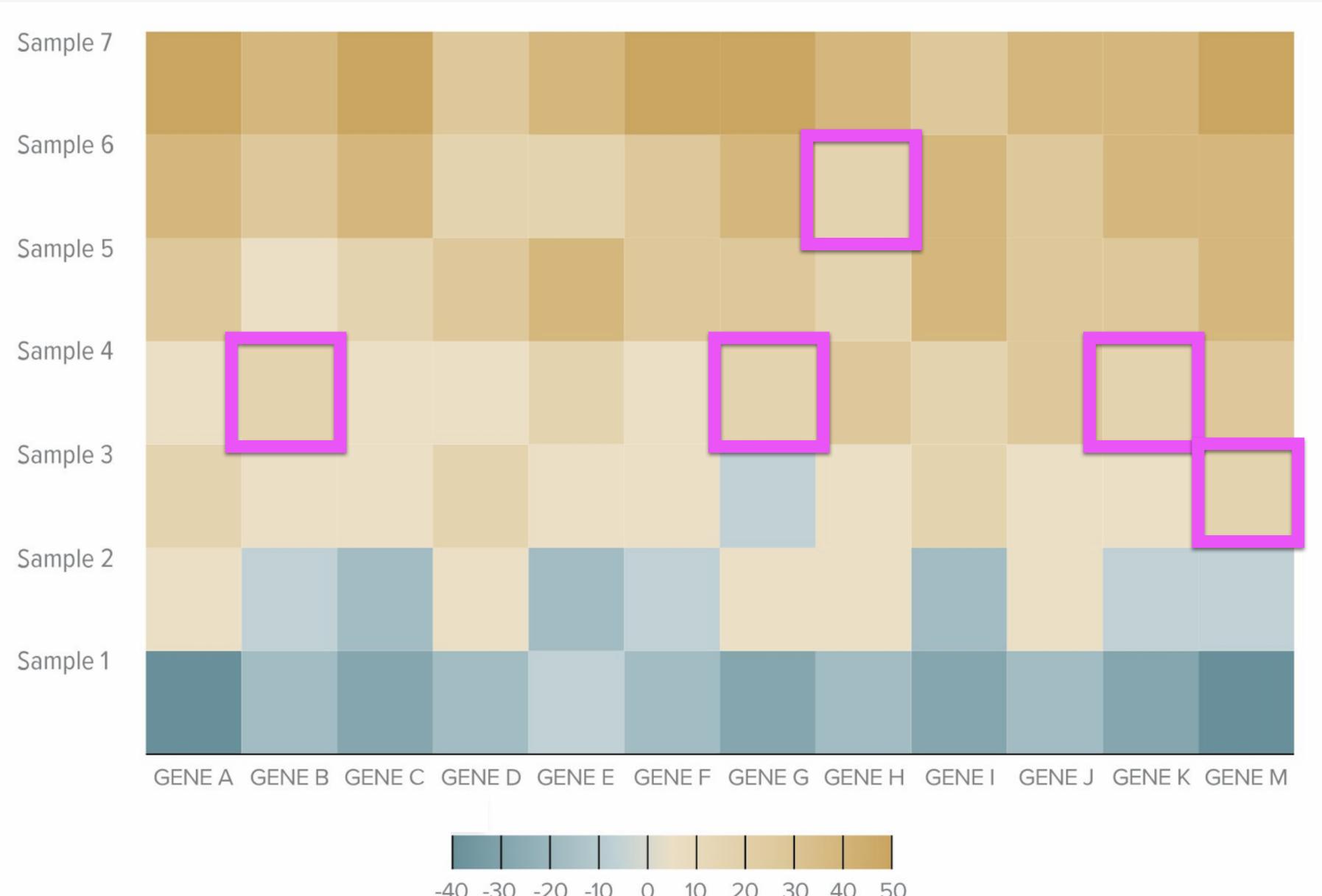
Colin Ware

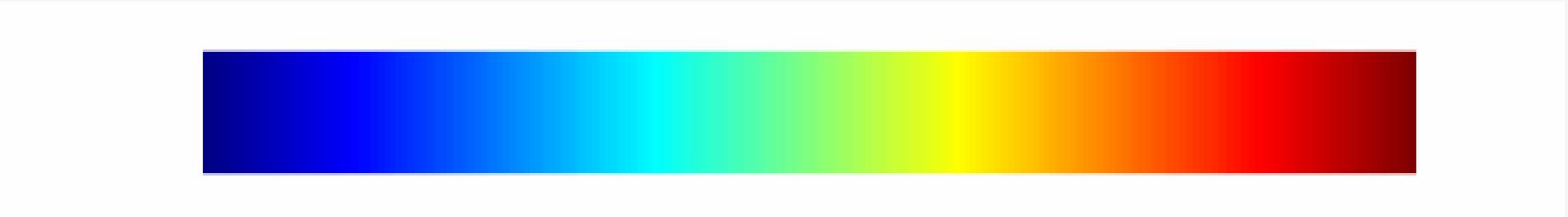


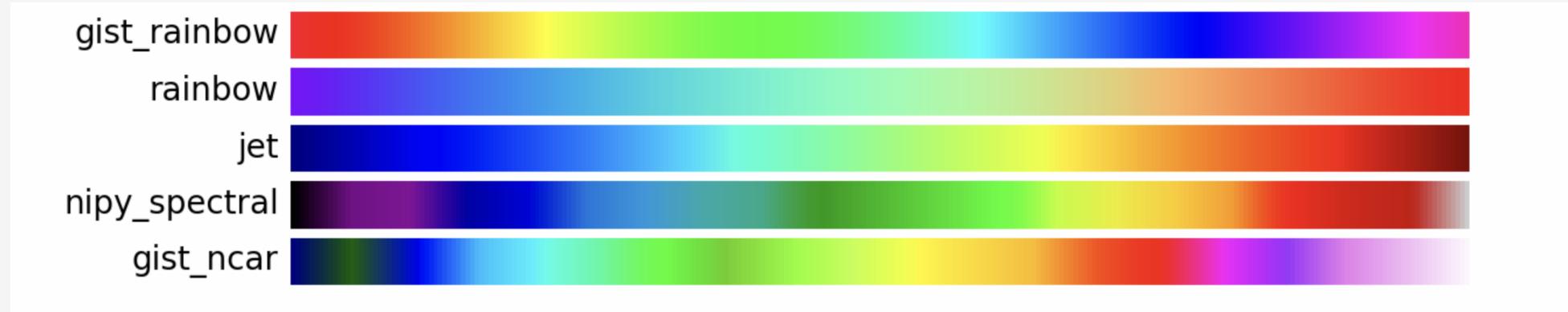
Colin Ware





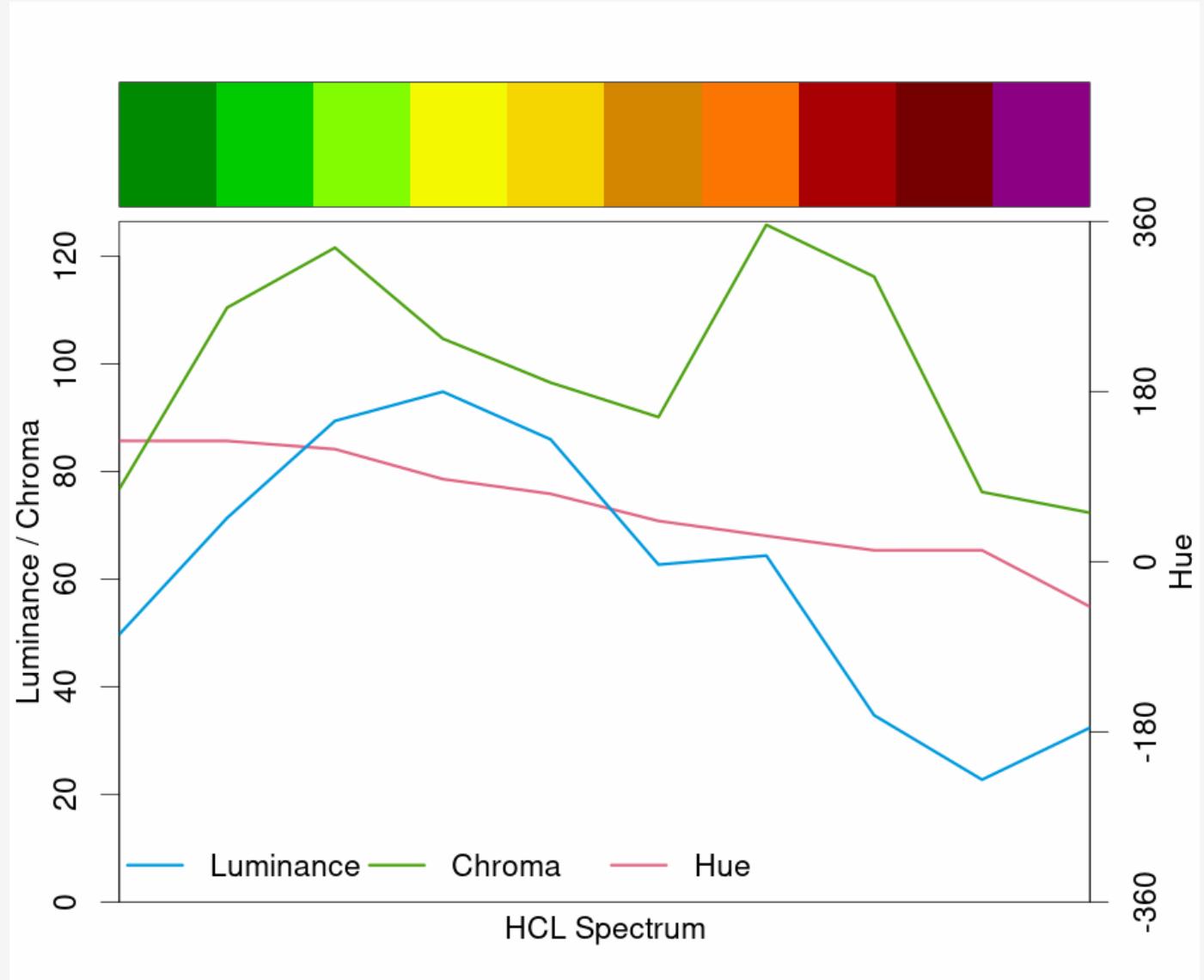


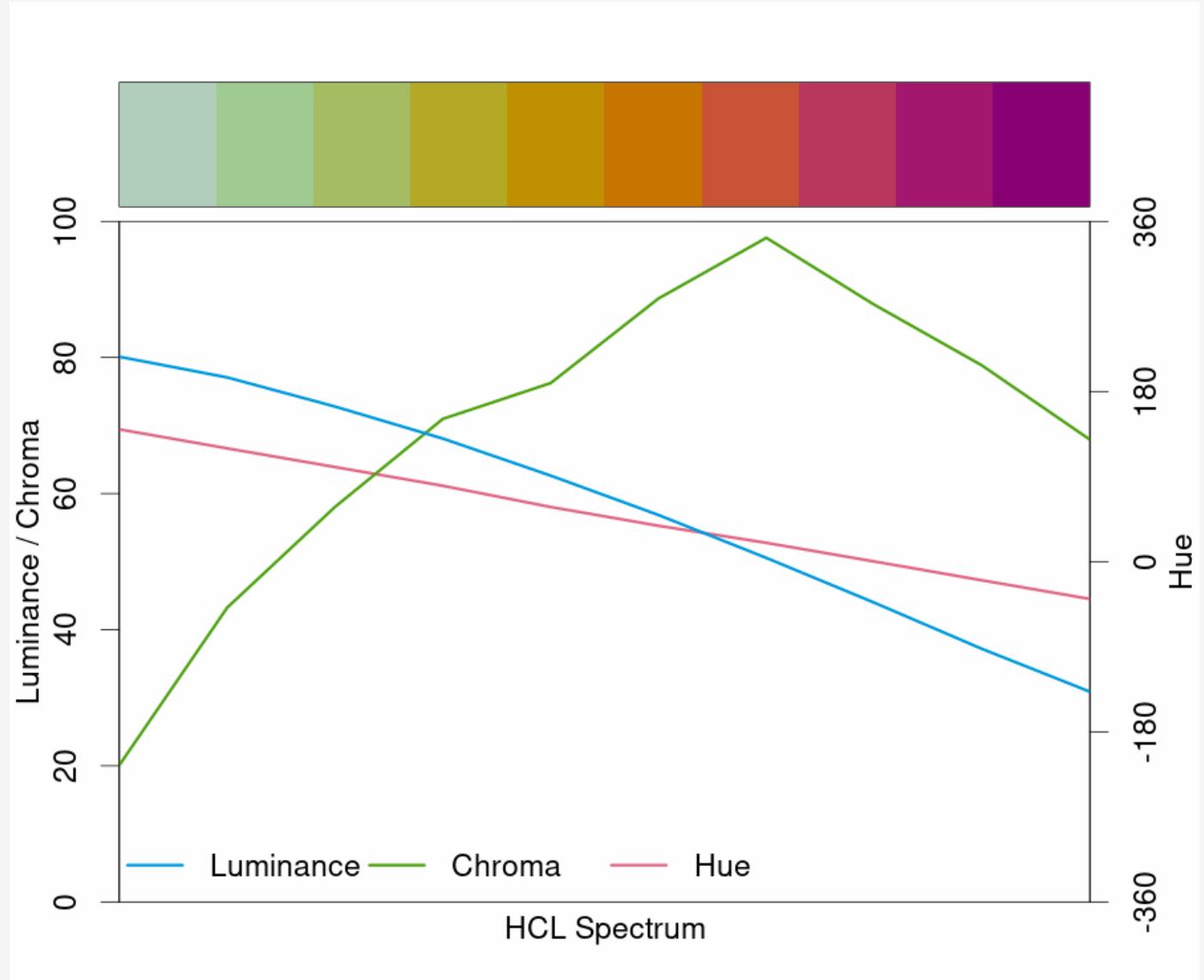




Perceptually Uniform Sequential colormaps







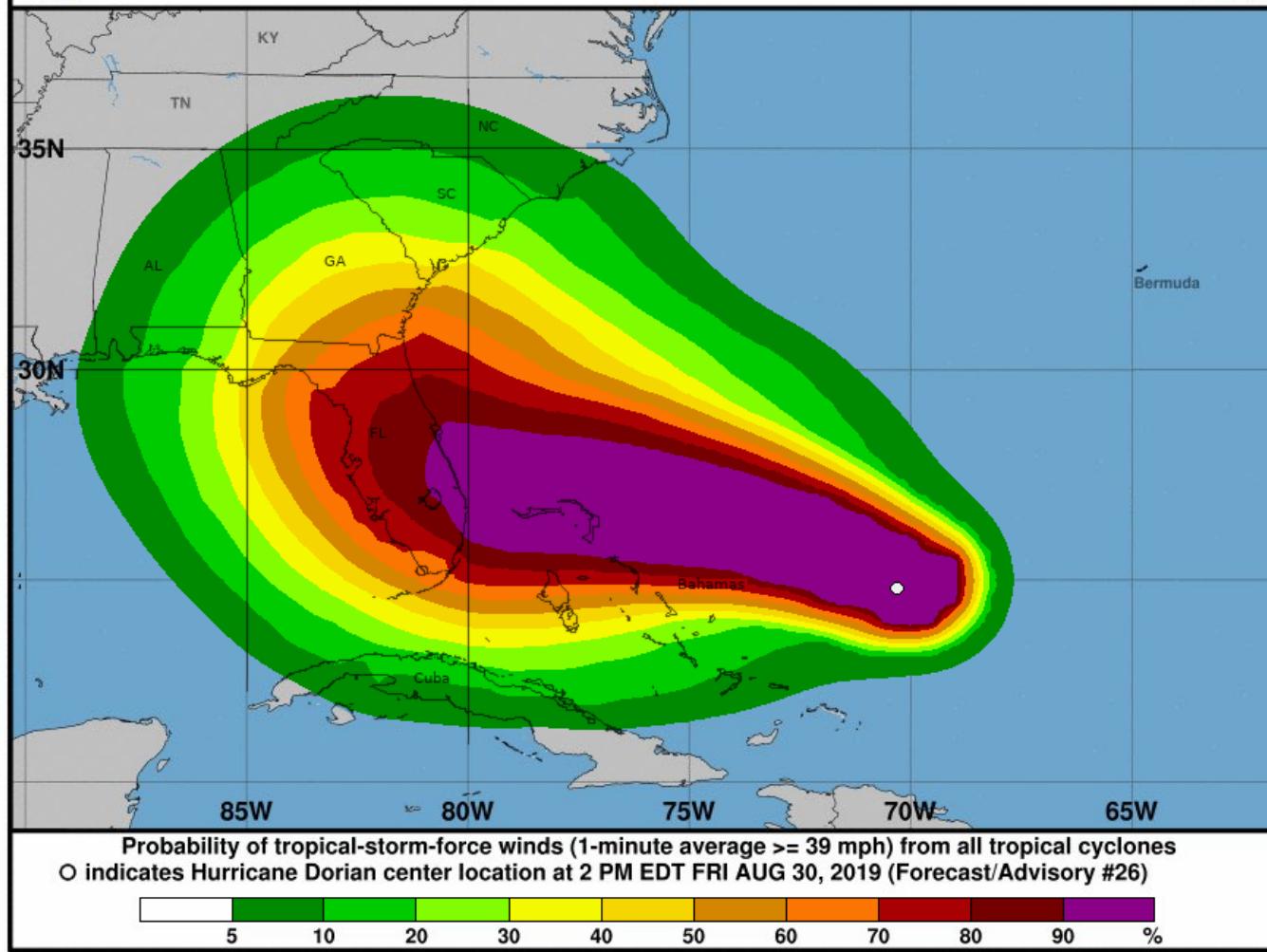






Tropical-Storm-Force Wind Speed Probabilities

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



National Weather Service

