

# Data Visualization - 1.

# Looking at Data (part 1)

Kieran Healy  
Code Horizons

December 10, 2023

# Data Visualization with R and ggplot2

# Housekeeping

10:00am to 12:30pm US EST

Break from 12:30pm to 1:30pm

1:30pm to 3:30pm

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

In between class sessions



# 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



R version 4.1.1 (2021-08-10) -- "Kick Things"  
Copyright (C) 2021 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin20 (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

> |

The screenshot shows the RStudio interface with the title bar "dataviz\_xaringan - main - RStudio". The left panel contains the R console output. The right panel has several panes: Environment, Files, Plots, Packages, Help, Viewer, and Presentation. The Environment pane shows the "Global Environment" with a "Functions" section containing "set". The Files pane displays a file tree for the "dataviz\_xaringan" directory, listing files like .gitignore, code, course\_notes.Rmd, data, data\_visualization.Rproj, LICENSE, pdf\_slides, README.md, README.Rmd, and slides, all modified on Feb 18, 2022, at 5:01 PM.

dataviz\_xaringan - main - RStudio

course\_notes.Rmd 01-looking-at-data.Rmd

Knit on Save Knit Run

Source Visual Outline

```
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.
```

Environment History Connections Git Tutorial

Import Dataset 271 MiB

Global Environment

Functions set function (name, value)

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 rendering your notes



The screenshot shows the RStudio interface with the following details:

- Title Bar:** The file name "course\_notes.Rmd" is displayed in the title bar.
- Toolbar:** Includes icons for back, forward, save, knit on save (unchecked), ABC (with a green checkmark), search, knit (with a blue knitting icon), and settings.
- Source/Visual Tabs:** The "Source" tab is selected.
- Code Editor:** 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.

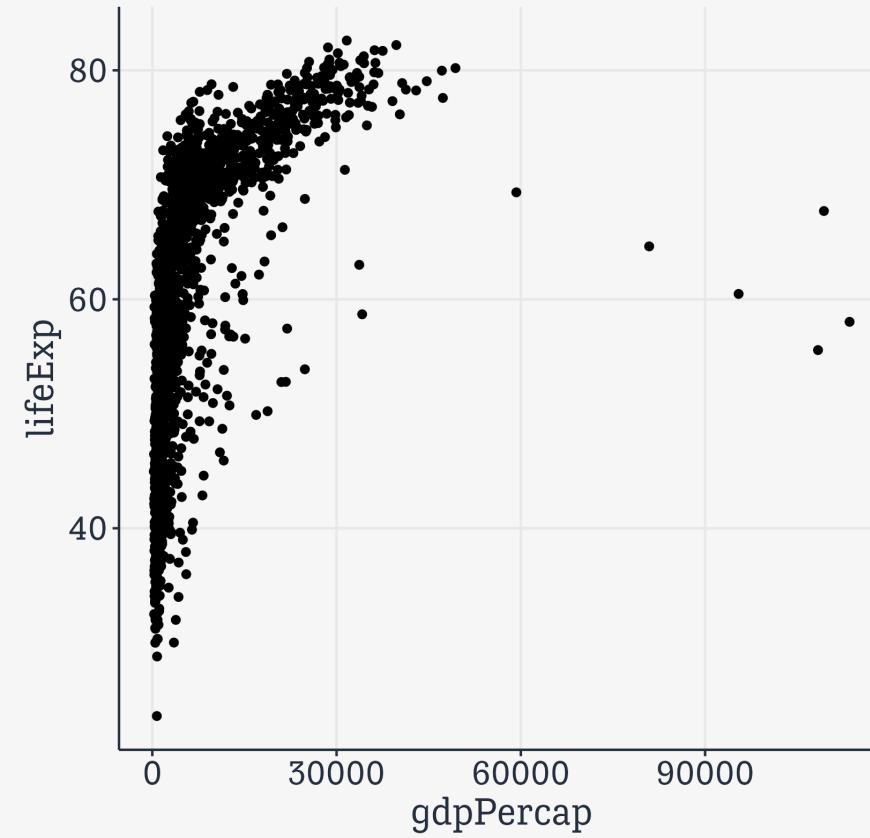
# Now write the following code

Write this out inside the “code chunk” in your notes.

```
library(tidyverse)
library(gapminder)

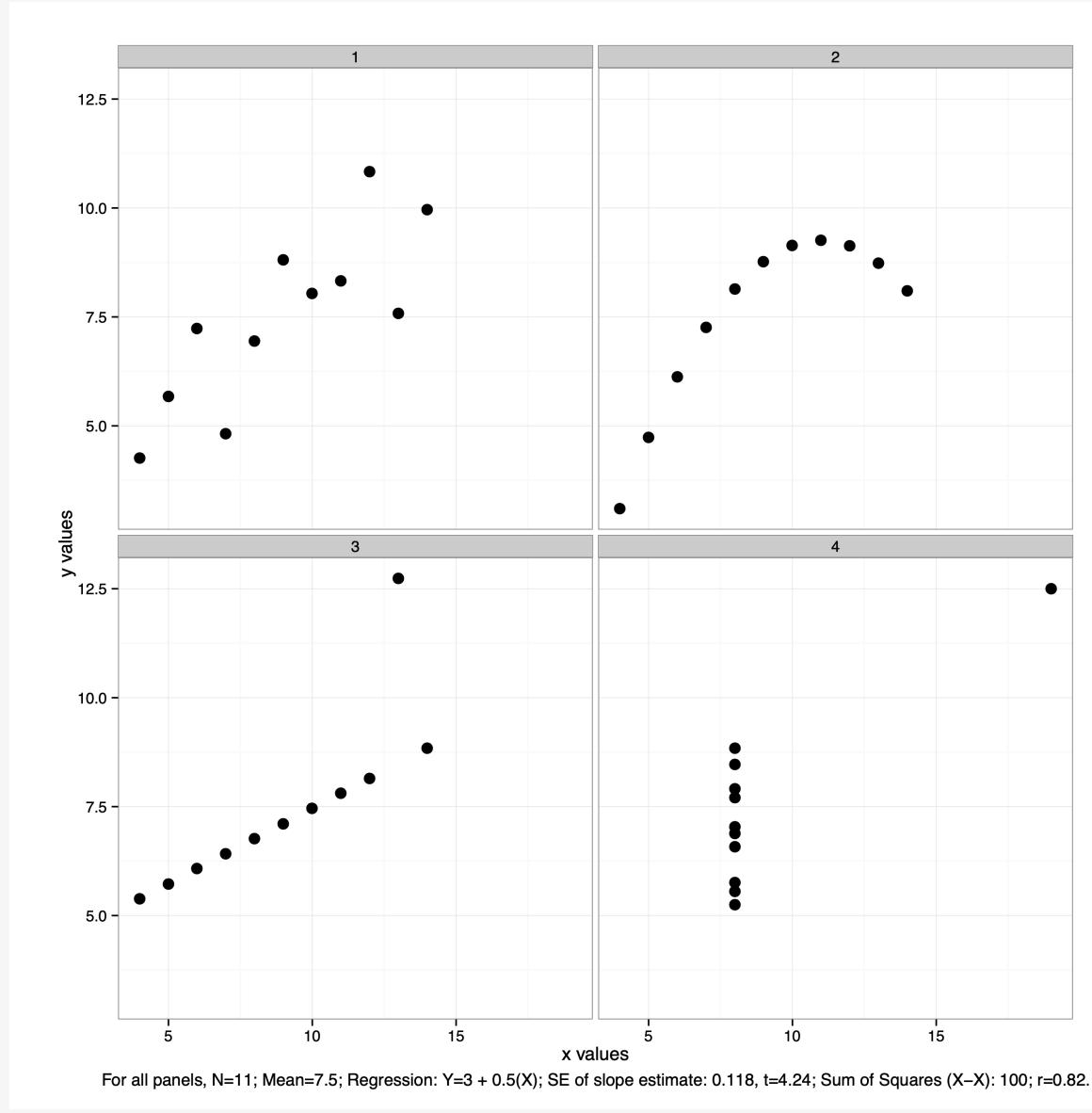
p ← ggplot(data = gapminder,
            mapping = aes(x = gdpPercap,
                           y = lifeExp))

p + geom_point()
```

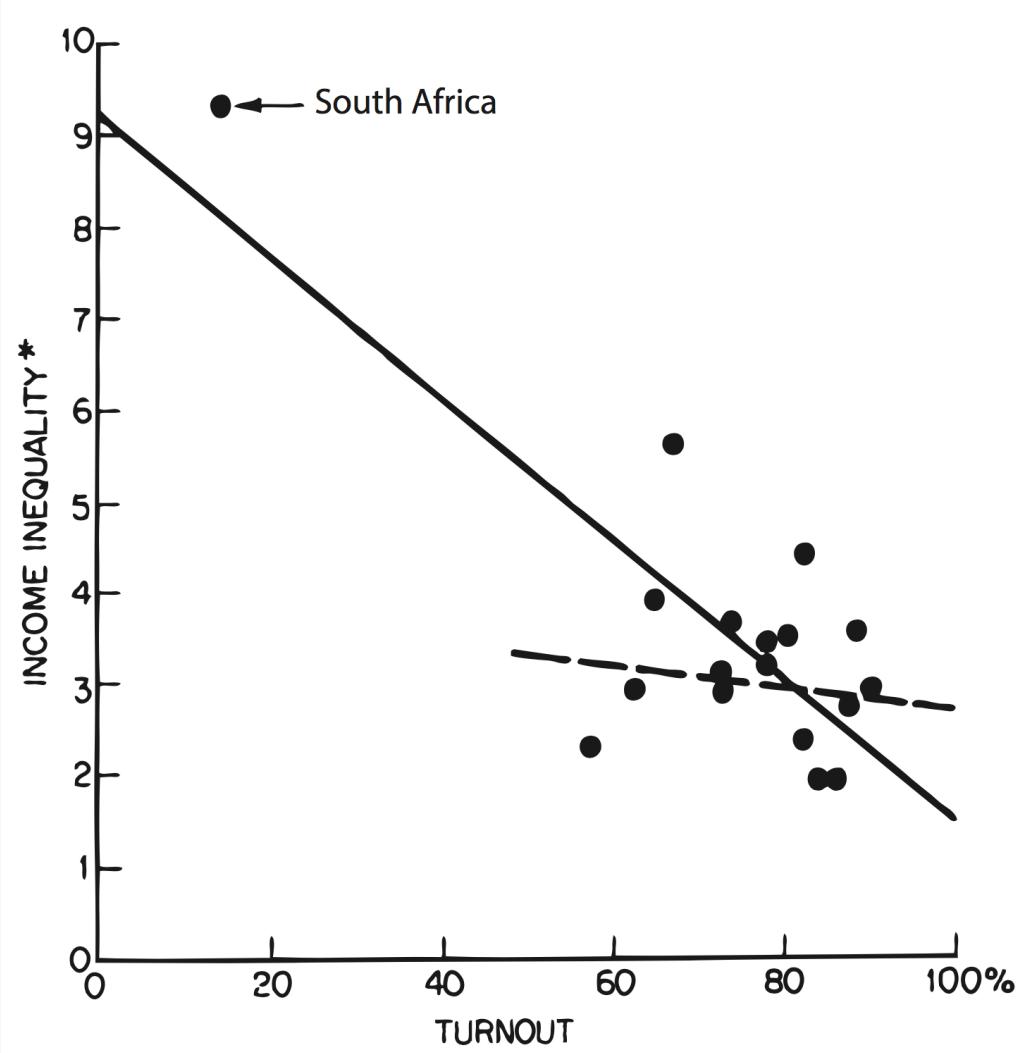


You should  
.kjh-yellowlook at  
your data

# Seeing things

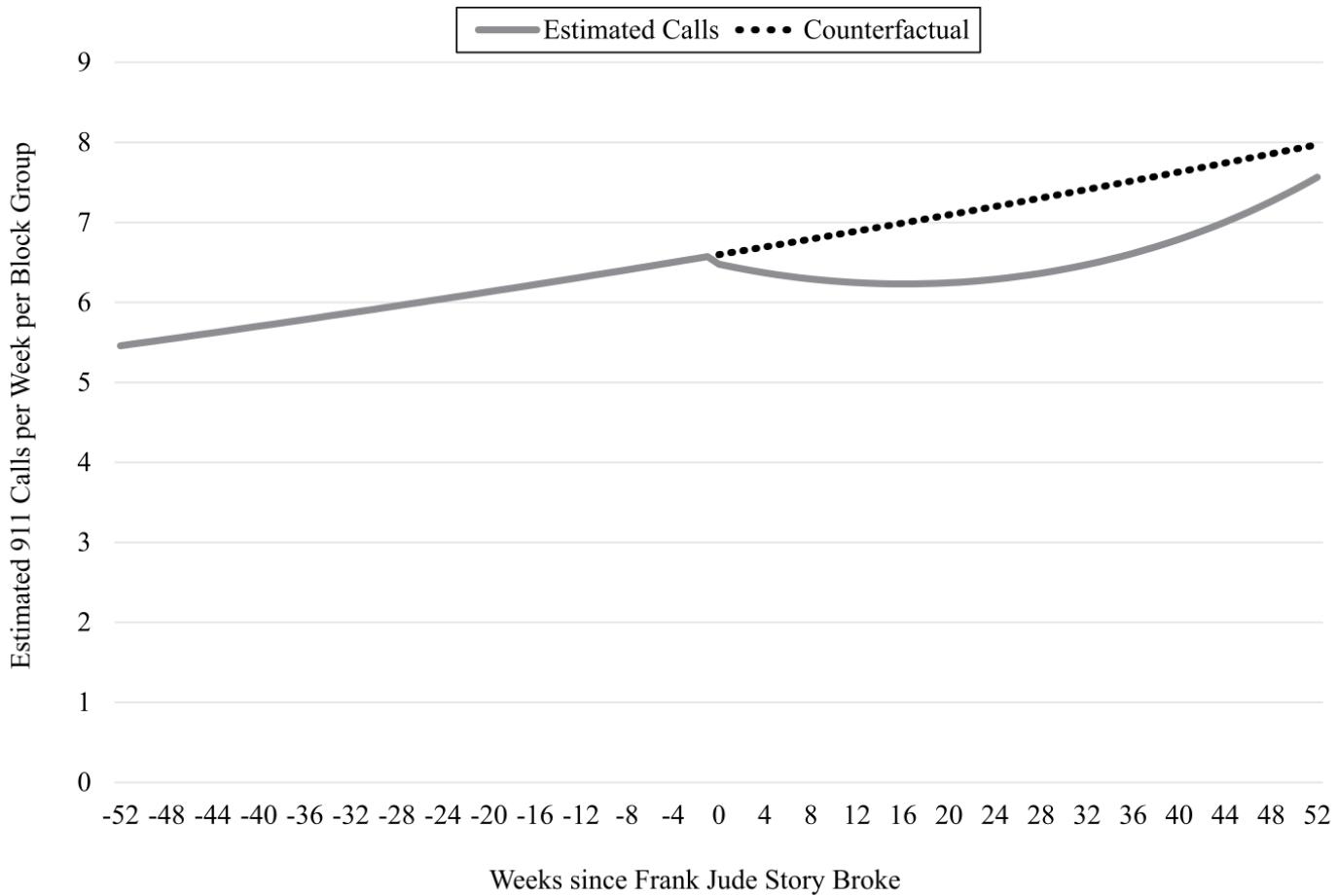


Anscombe's Quartet

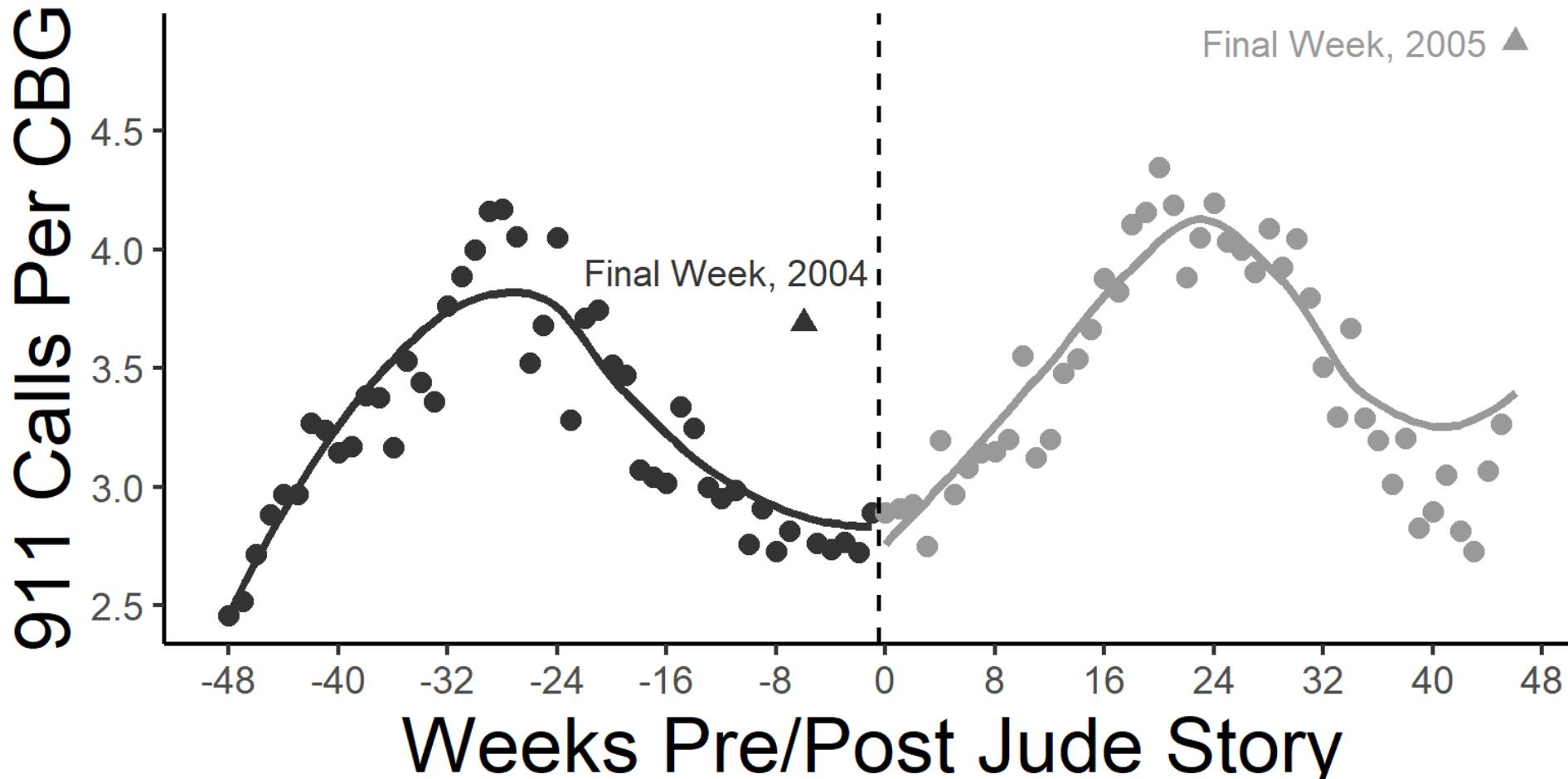


Key. — Bivariate slope including South Africa ( $N = 18$ )

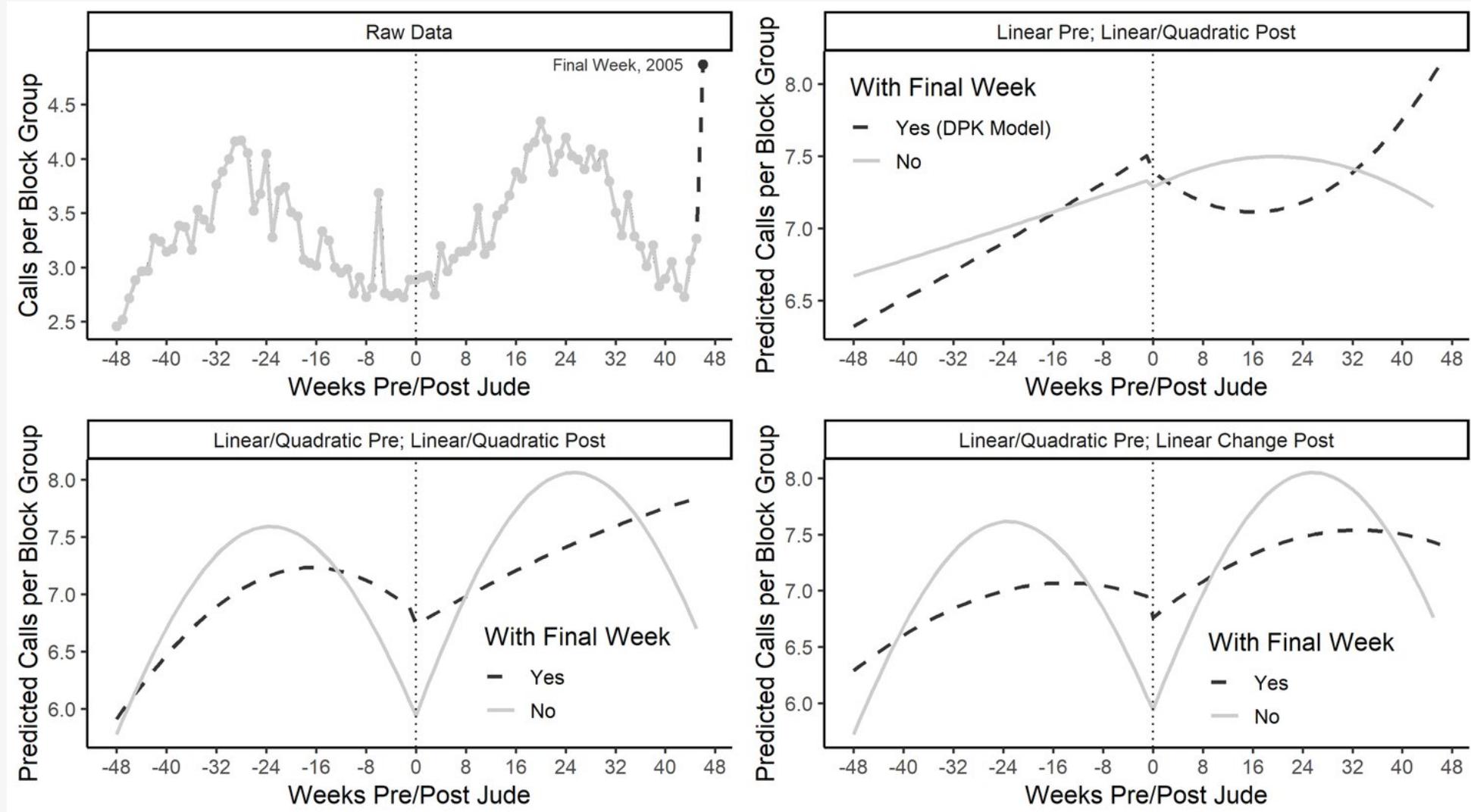
- - - Bivariate slope excluding South Africa ( $N = 17$ )



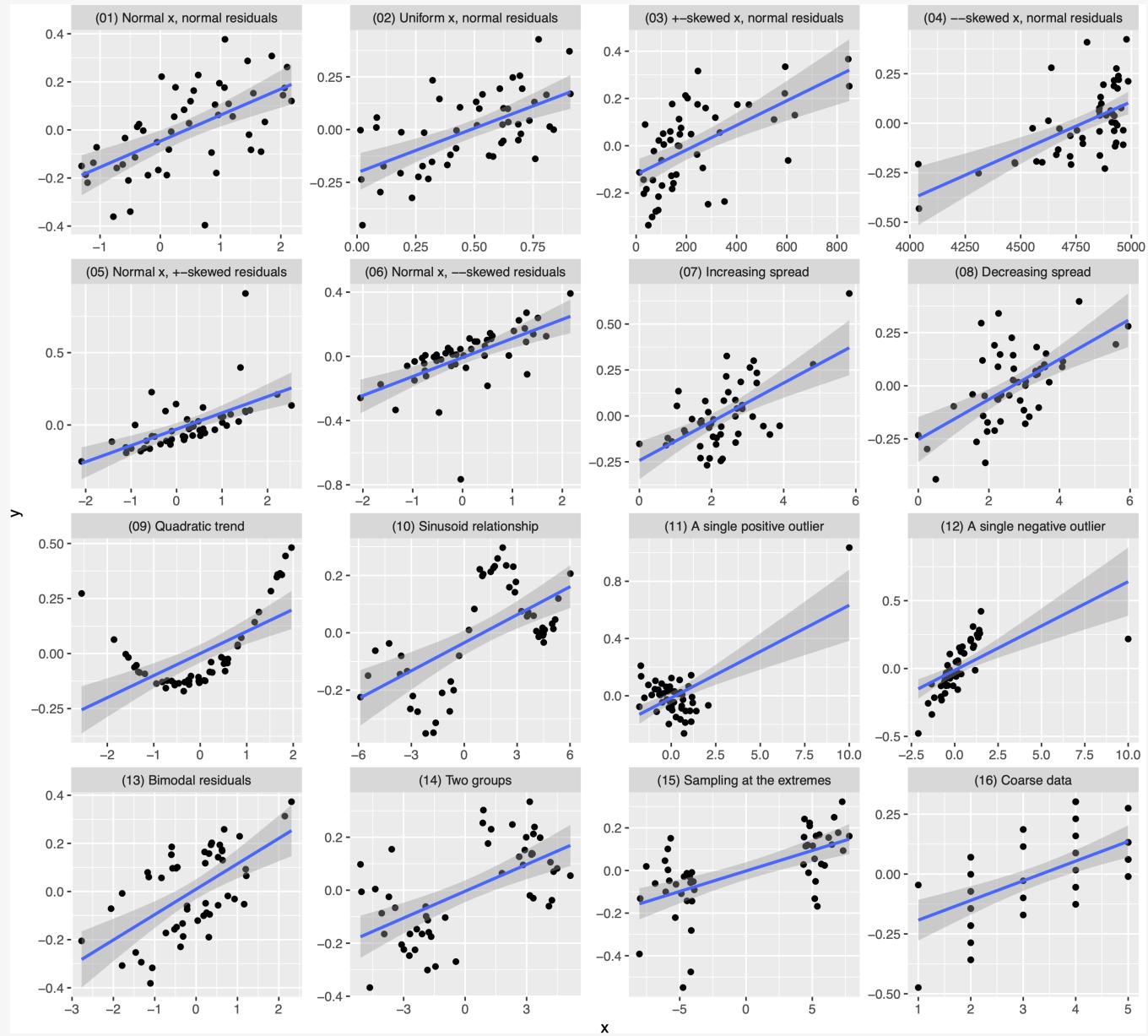
# All Neighborhoods

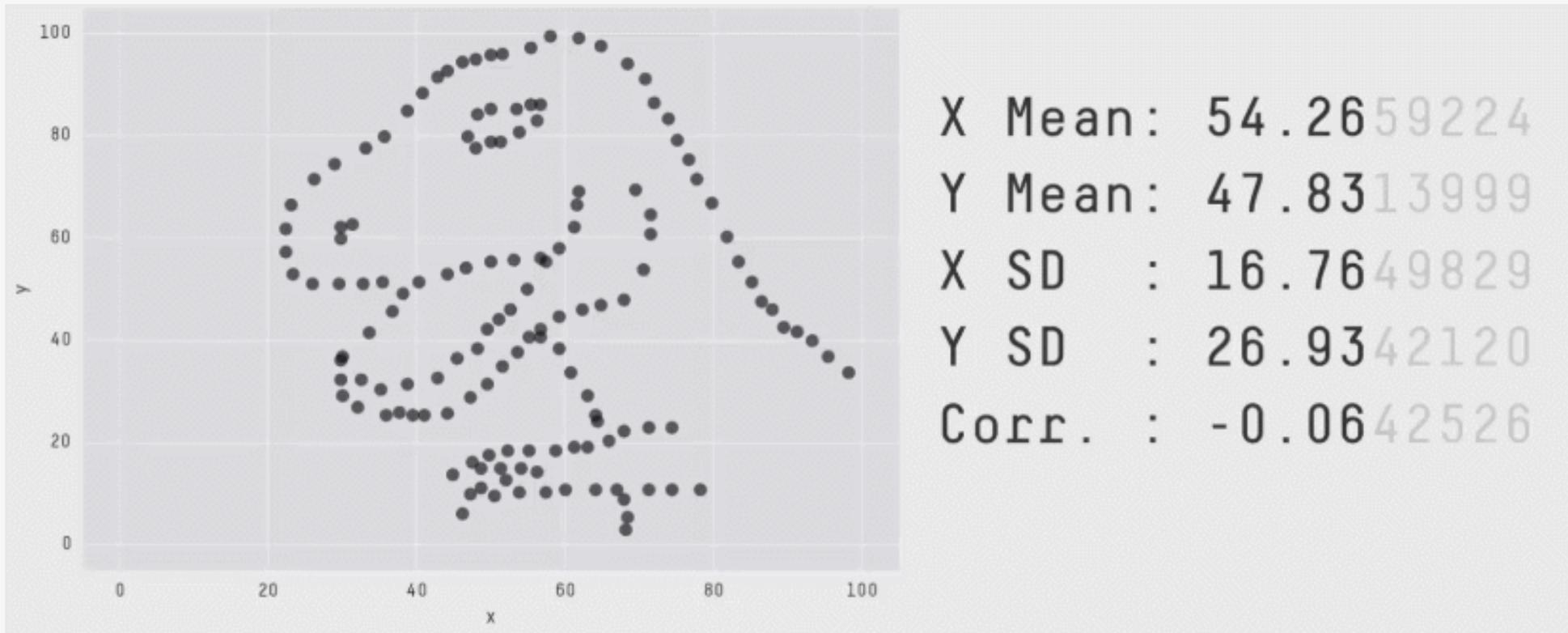


Zoorob (2020)



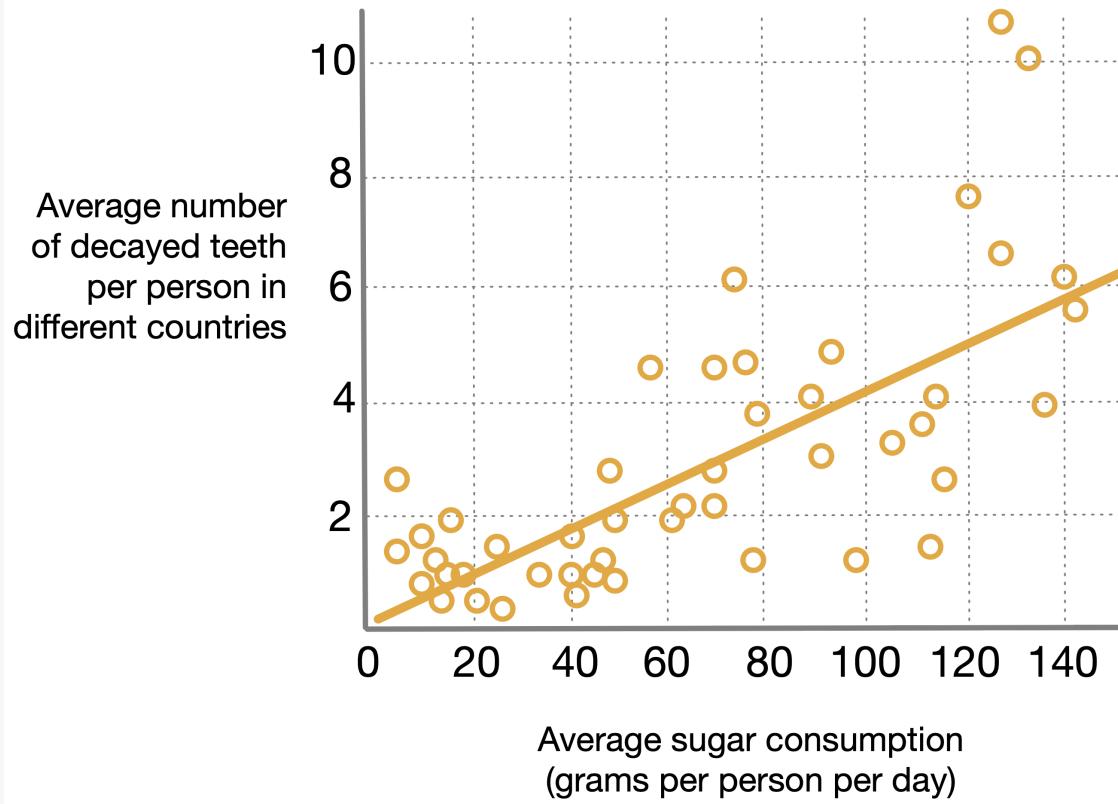
Zoorob (2020)





Cairo; Matejka & Fitzmaurice

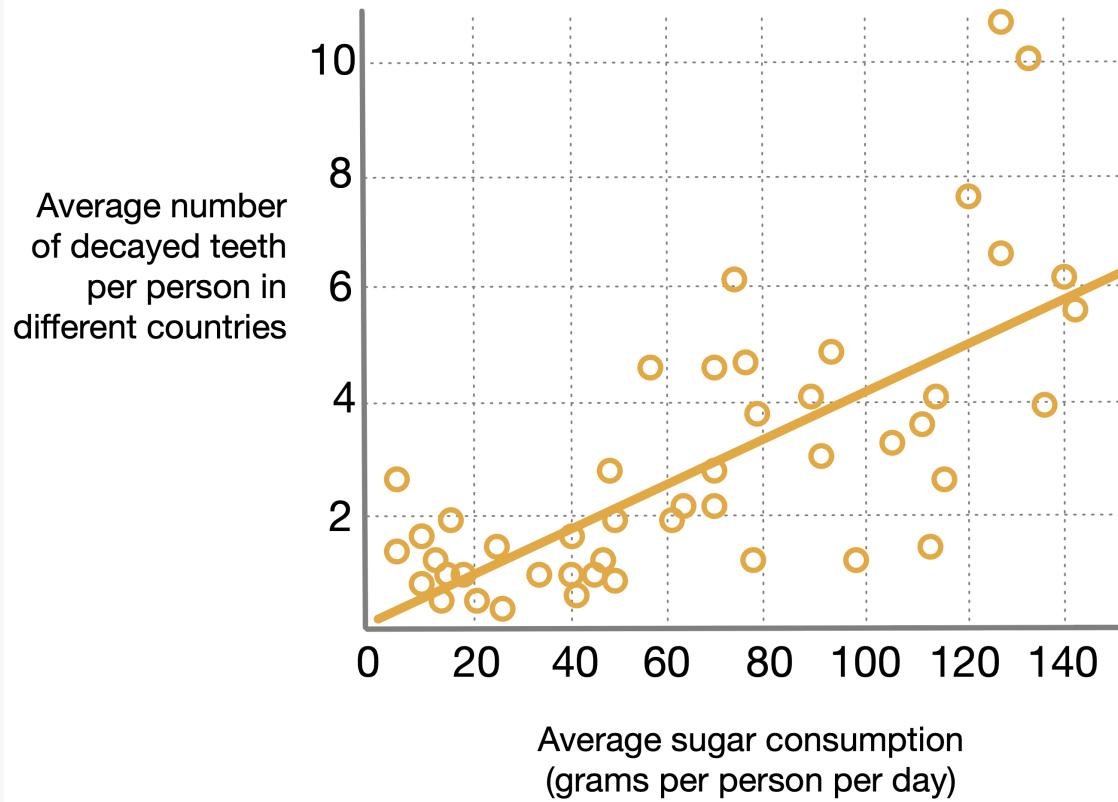
*Which of the following statements best describes the data in the graph below?*



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

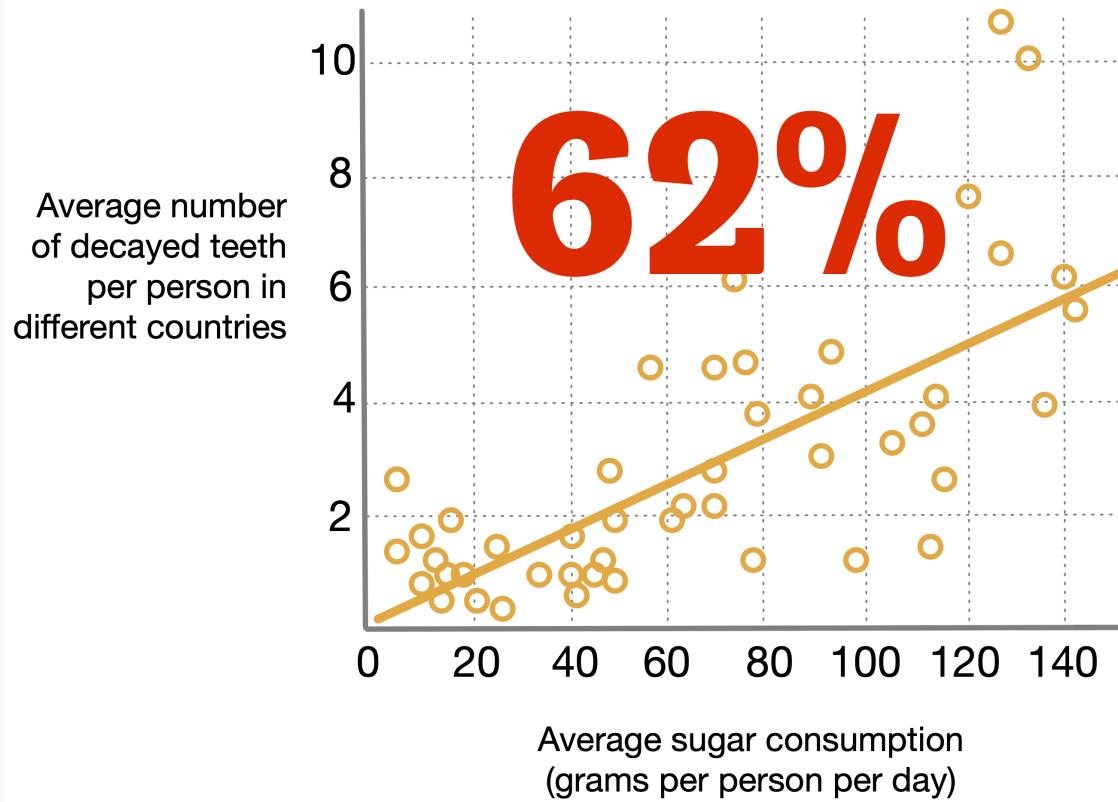
*Which of the following statements best describes the data in the graph below?*



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?*

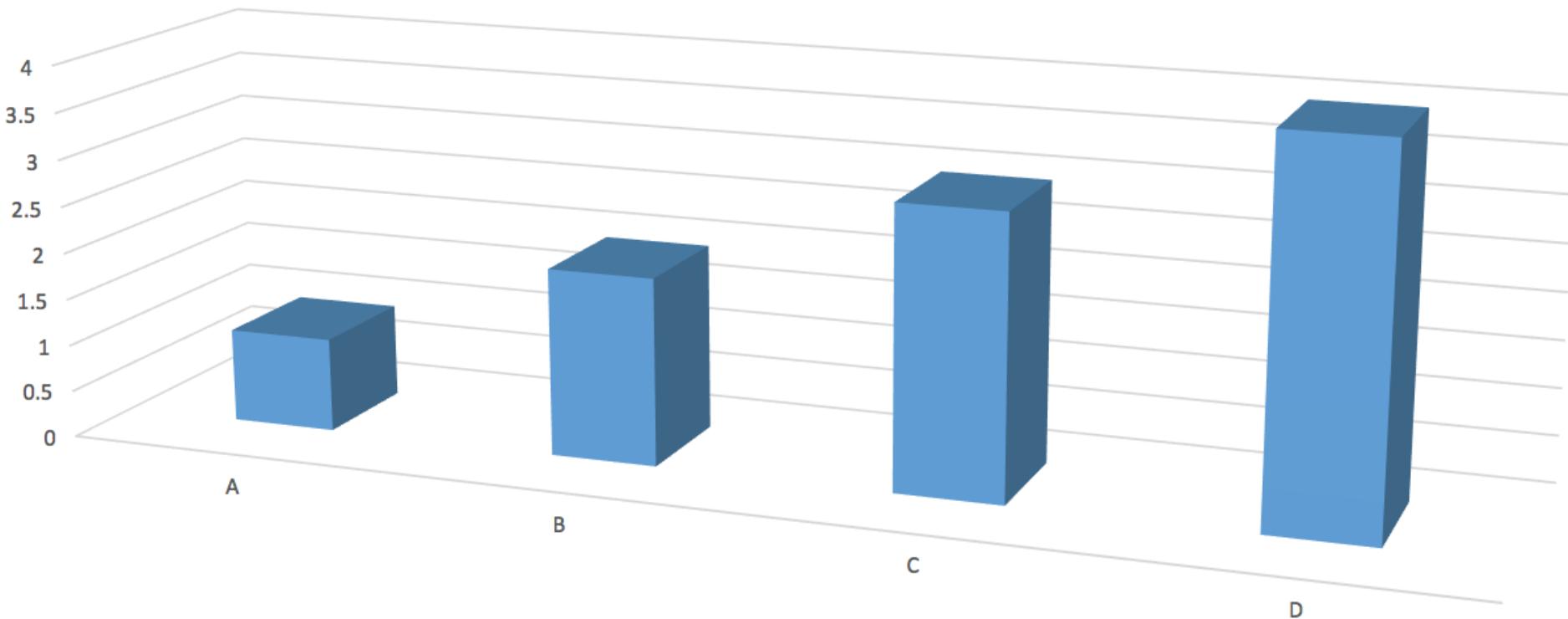


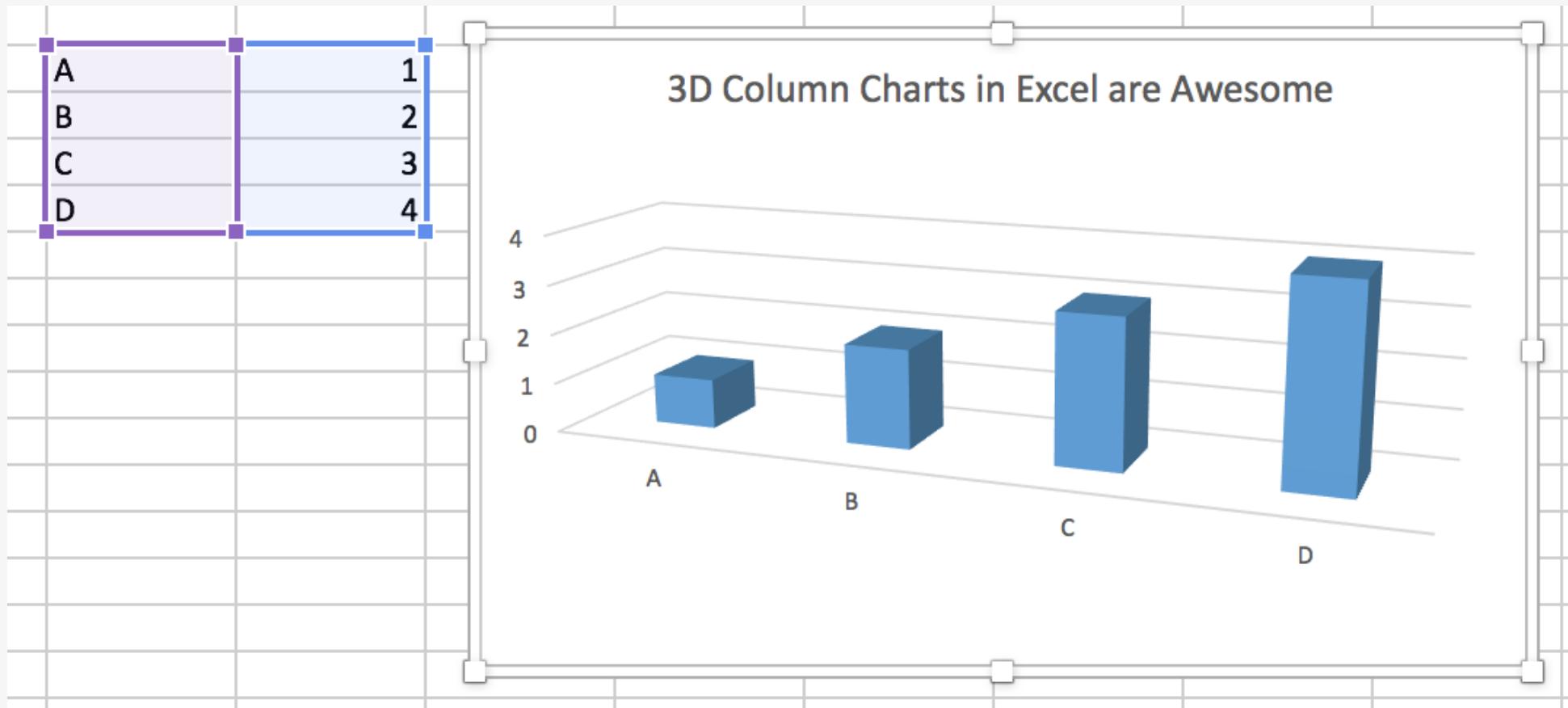
Pew Research

- 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





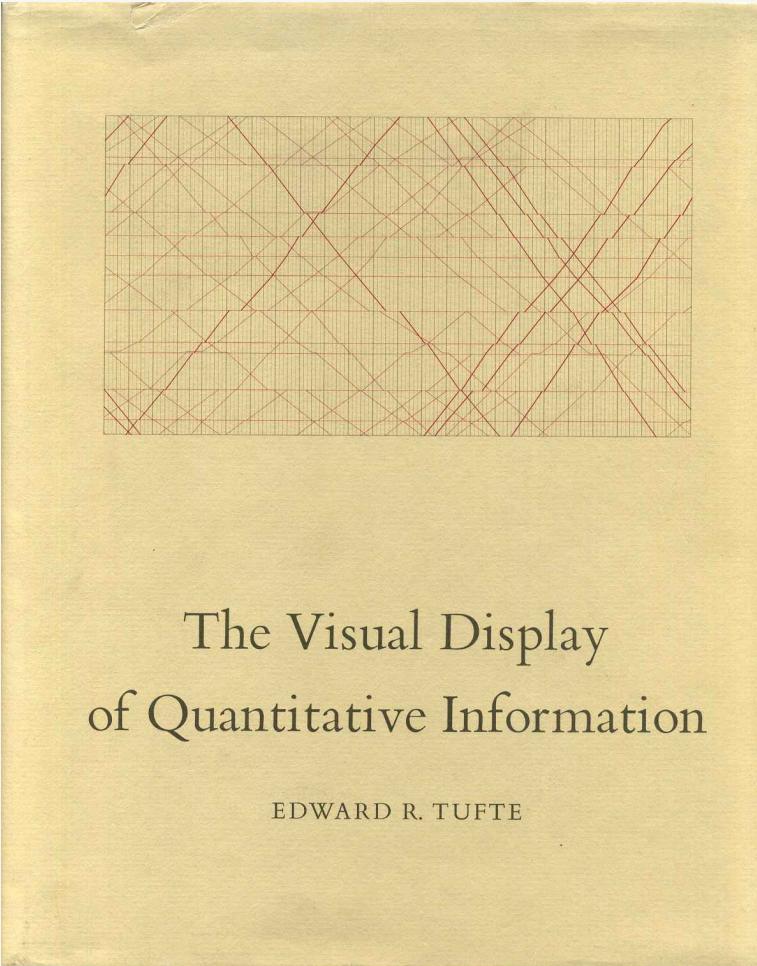
**Bad Taste**

**Bad Data**

**Bad Perception**

# Bad Taste: Simplify, Simplify?

# Tufte's “Data to Ink Ratio”





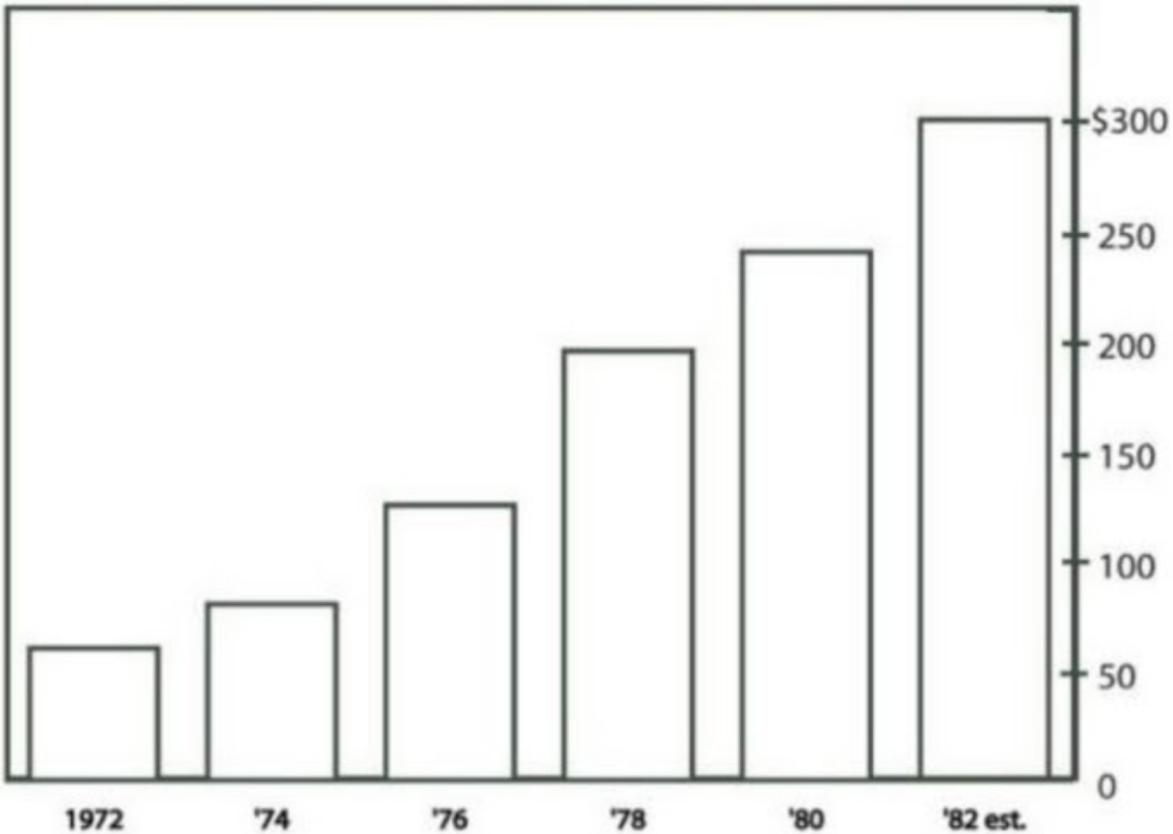
## MONSTROUS COSTS

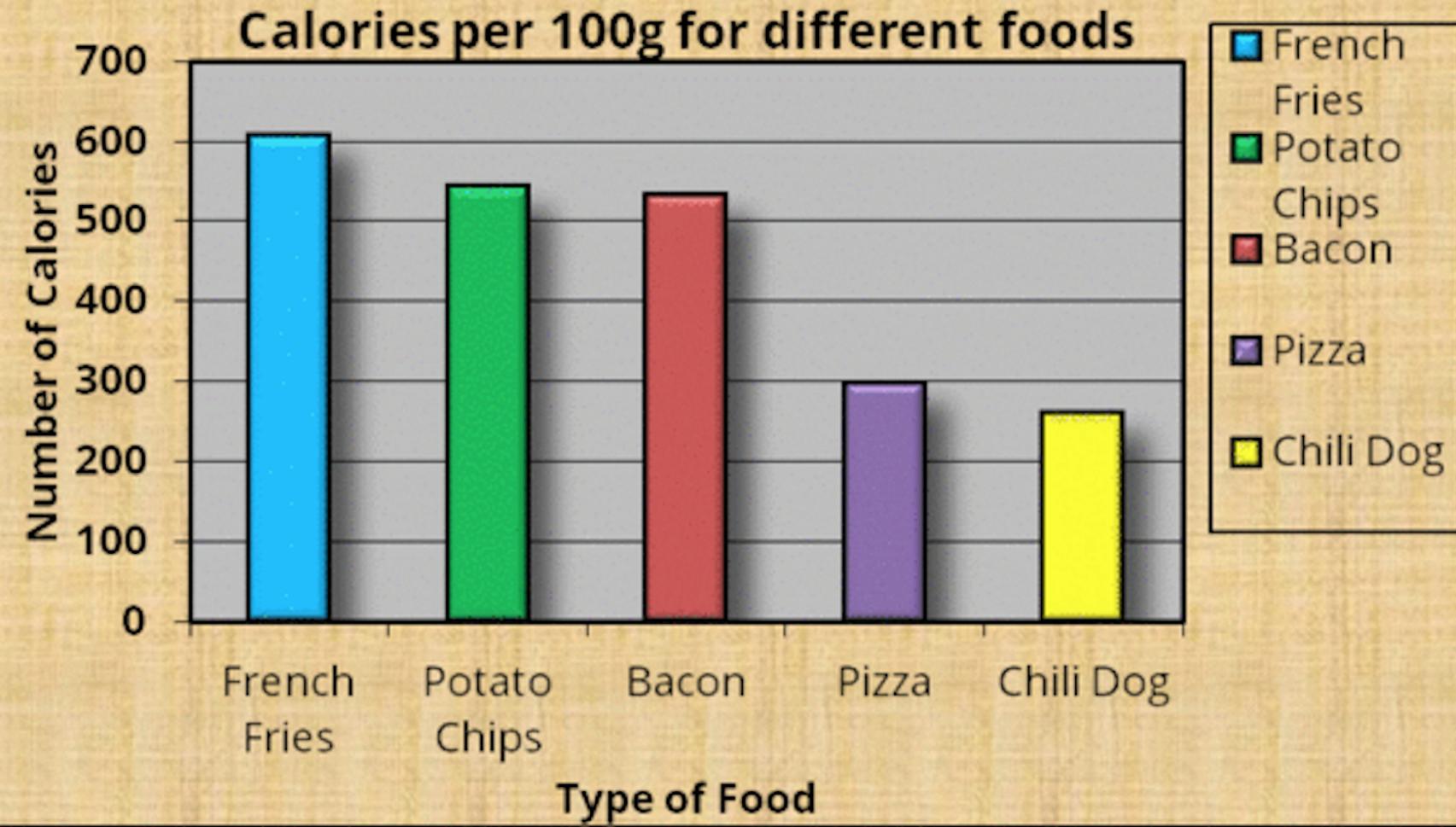
Total House and Senate campaign expenditures,  
in millions



Nigel Holmes

**MONSTROUS COSTS**  
Total House and Senate campaign expenditures, in millions

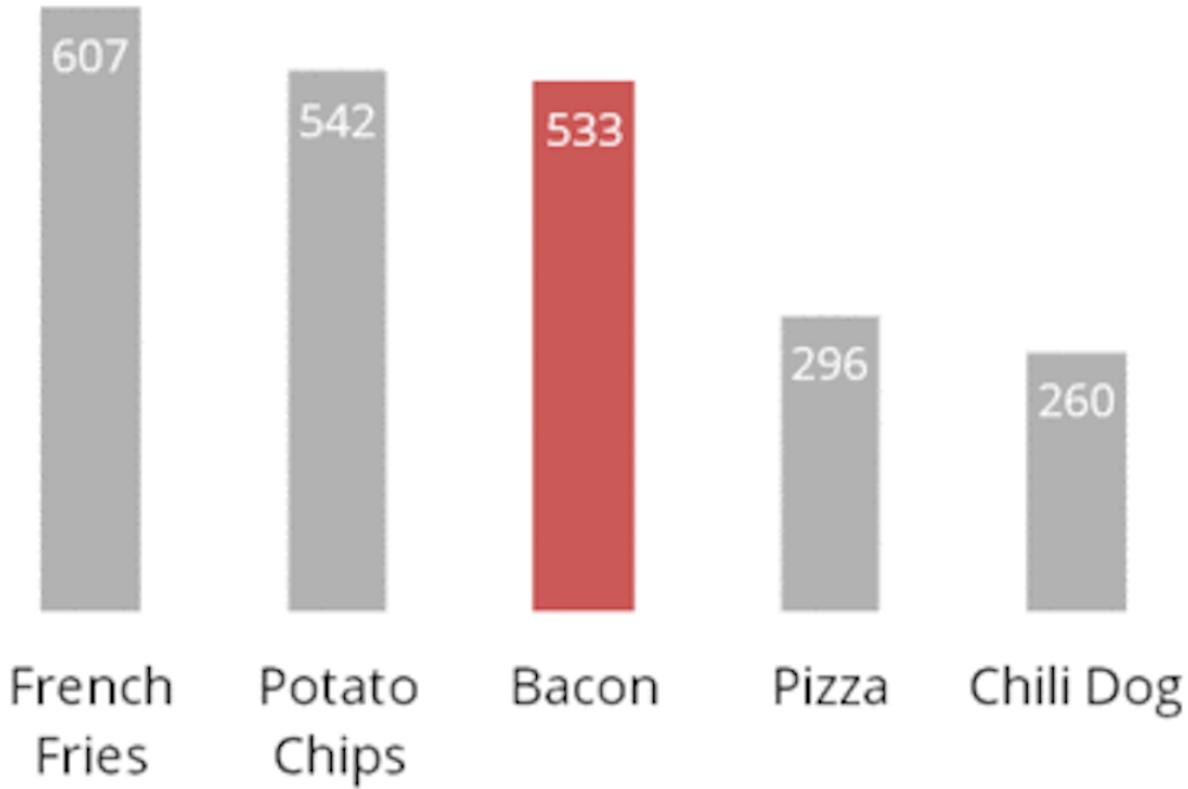




Created by Darkhorse Analytics

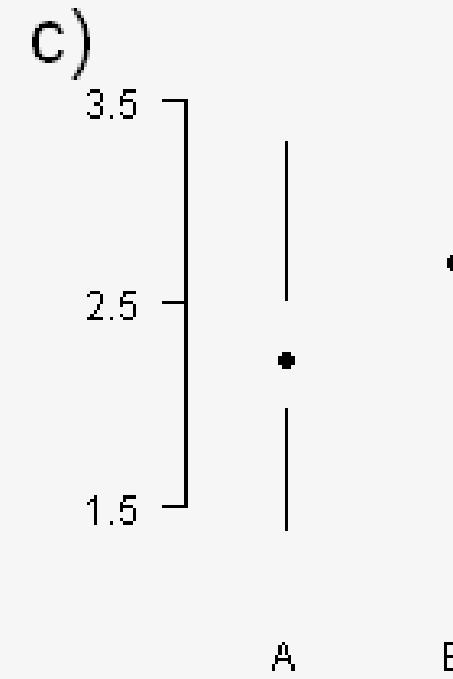
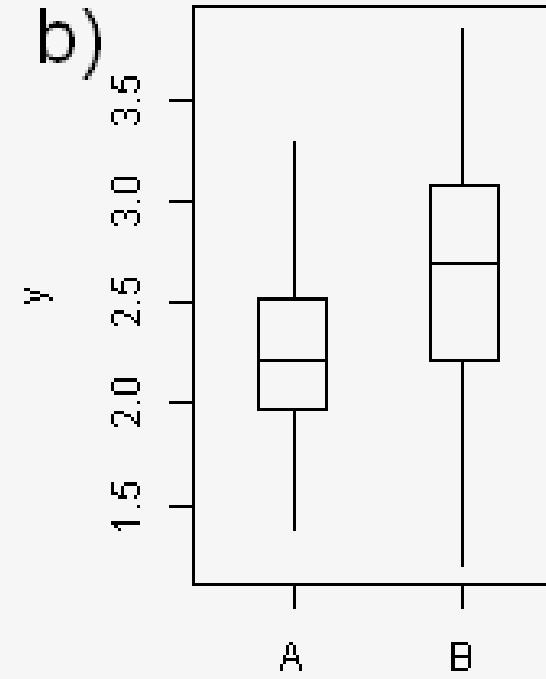
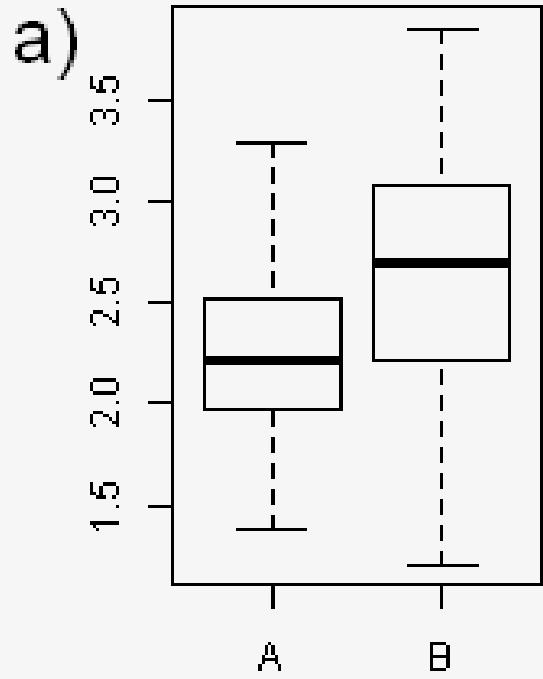
[www.darkhorseanalytics.com](http://www.darkhorseanalytics.com)

## Calories per 100g



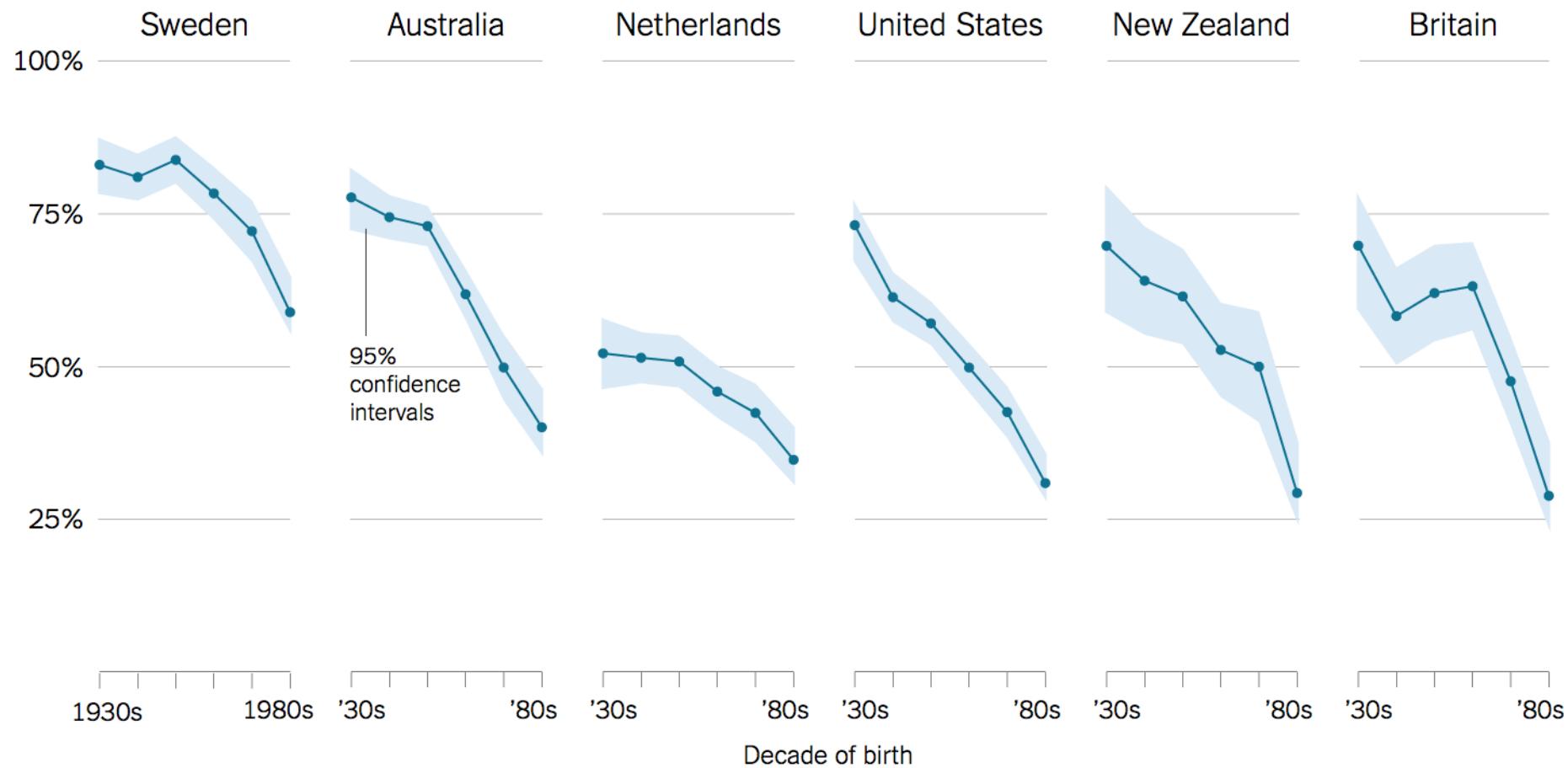
Created by Darkhorse Analytics

[www.darkhorseanalytics.com](http://www.darkhorseanalytics.com)

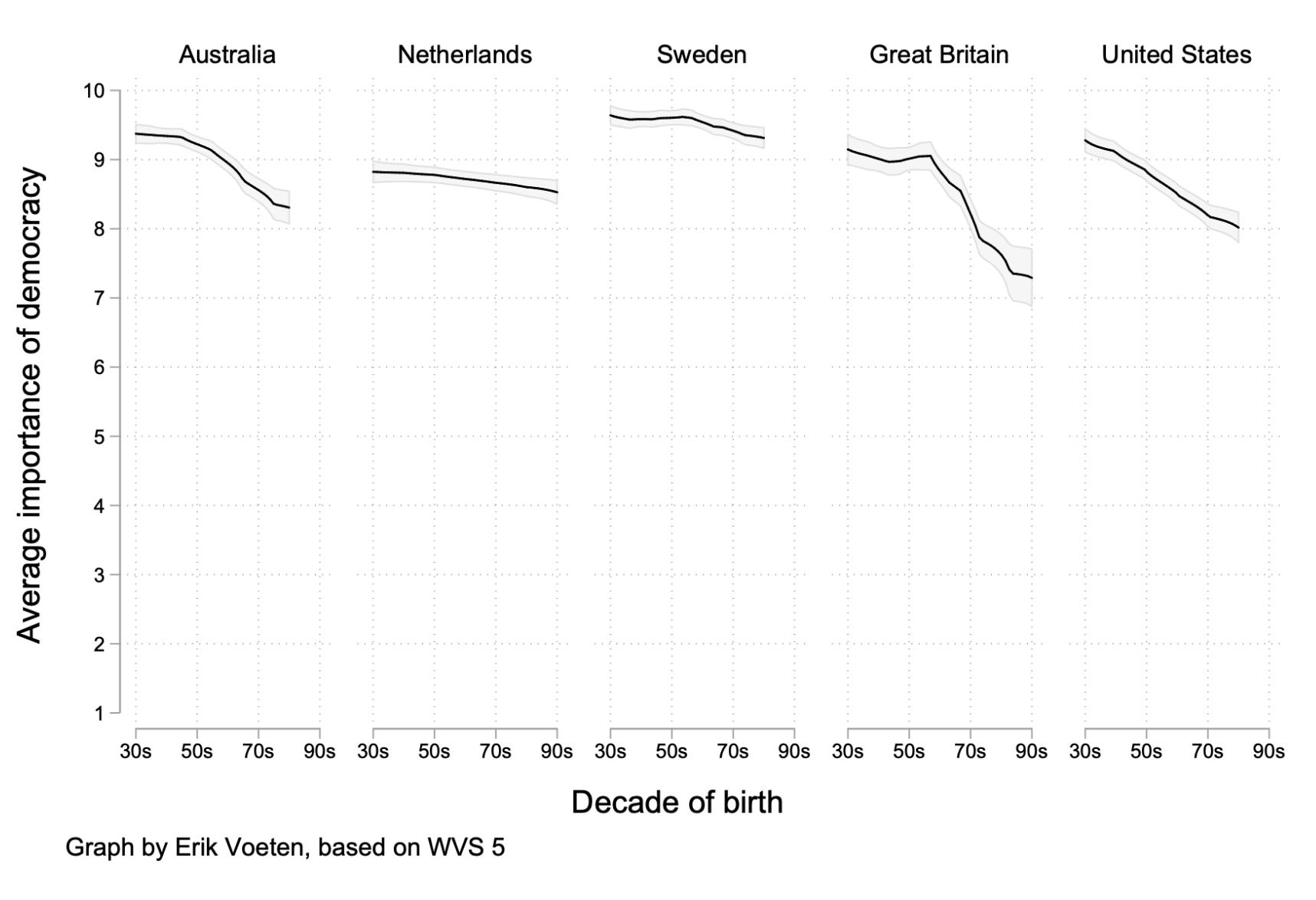


# Bad Data: Junk-Free Junk Charts

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



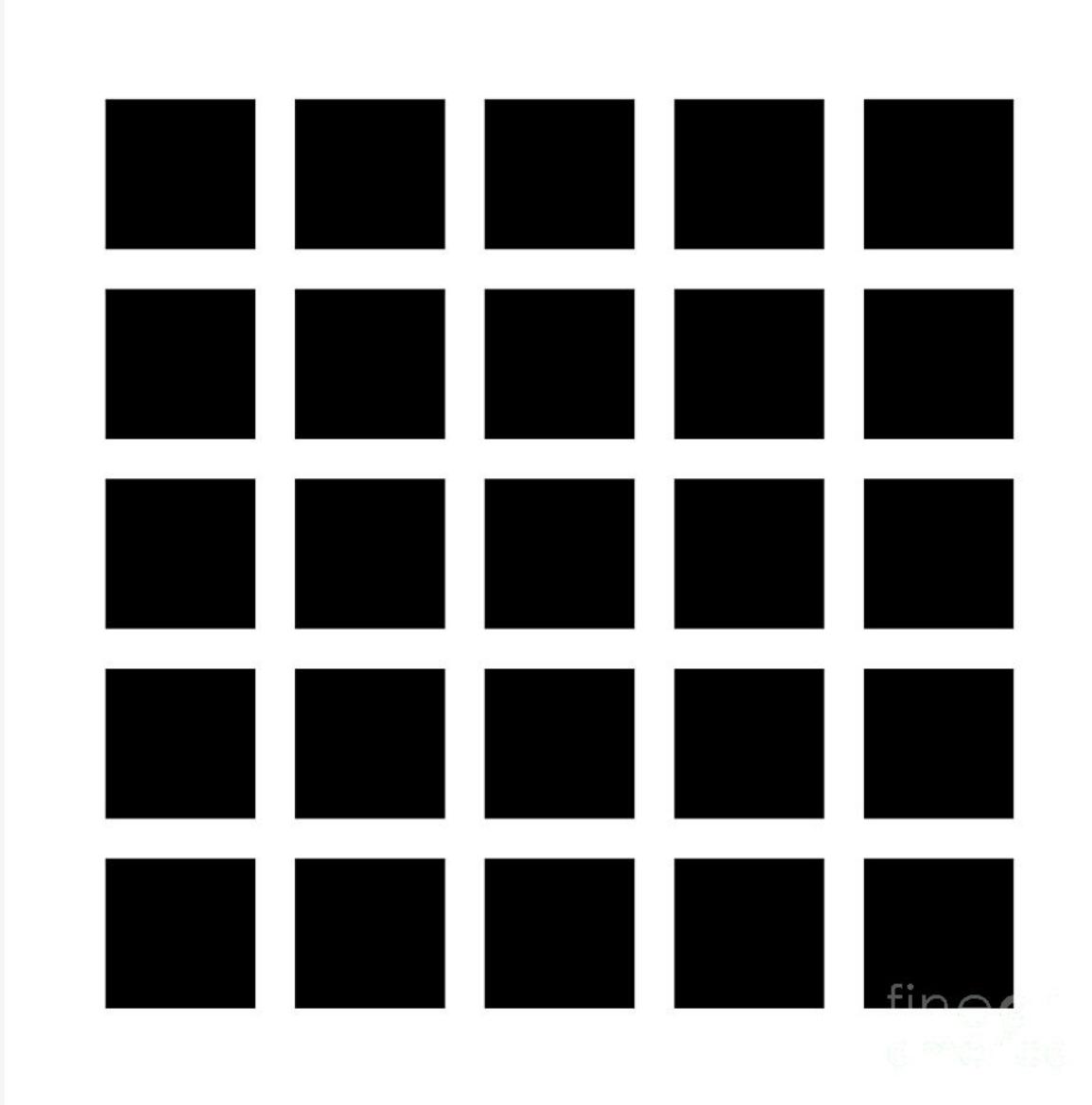
Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times



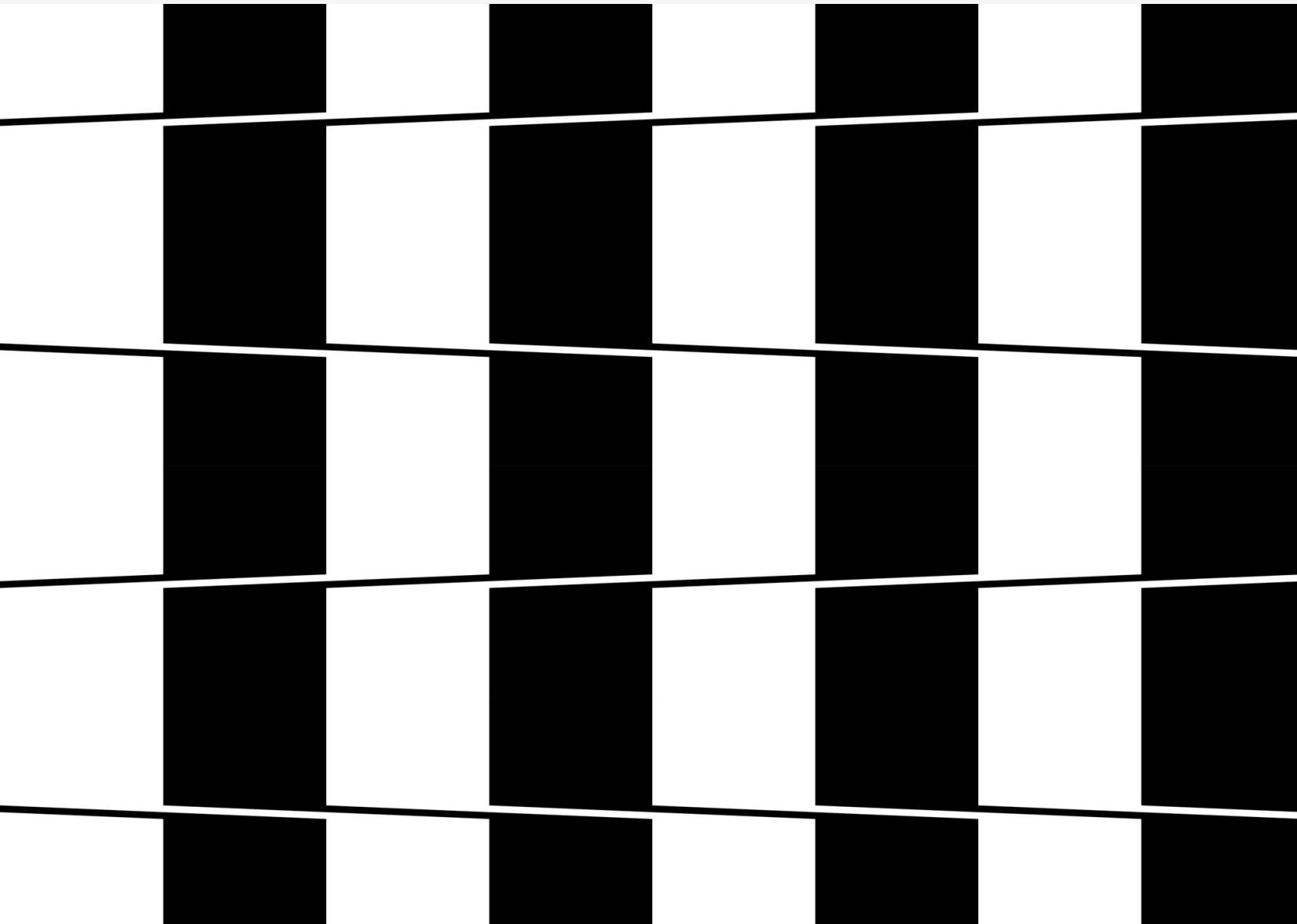
Erik Voeten

# Bad Perception: Seeing and Not Seeing

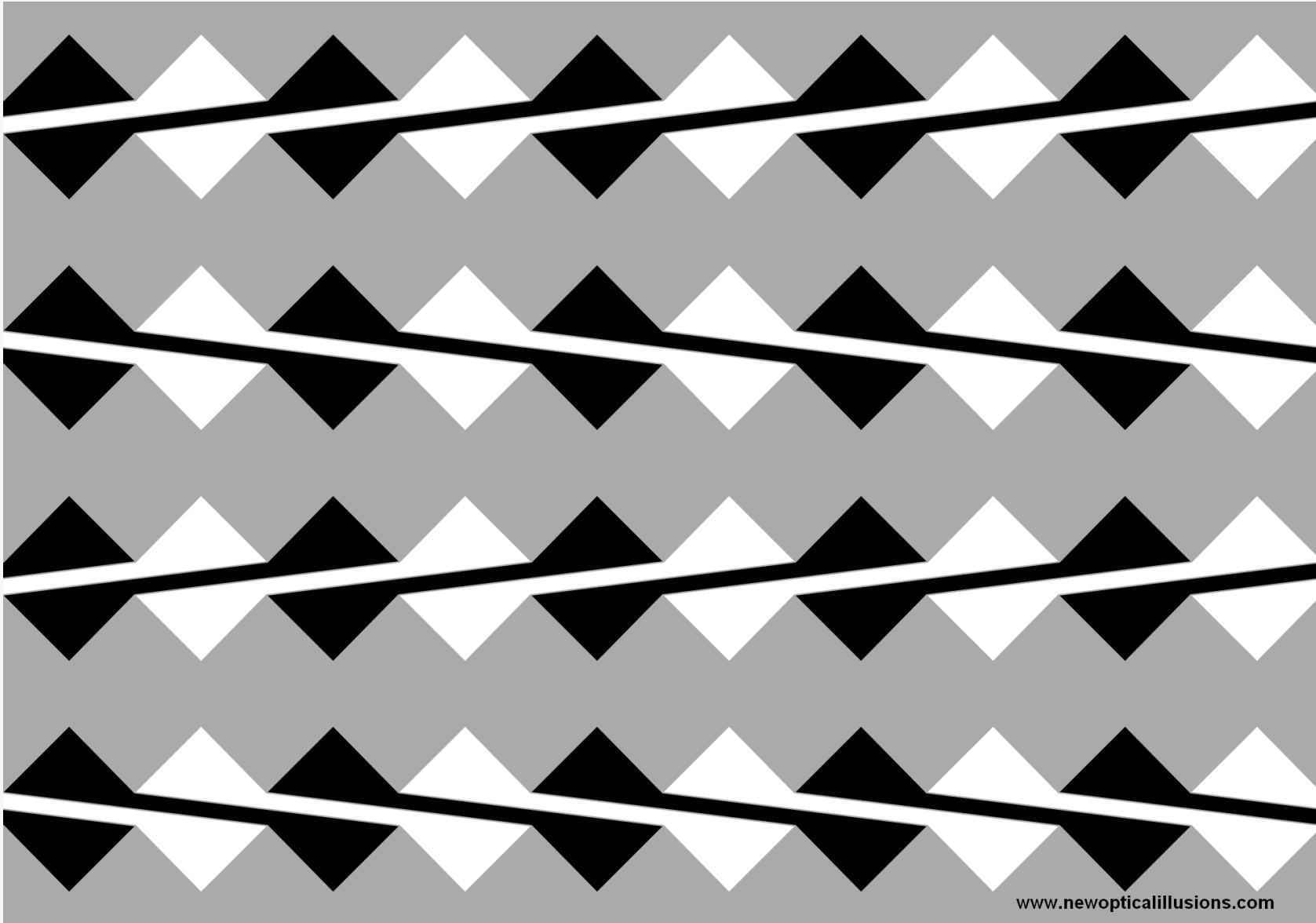
# Edges & Contrasts



Hermann Grid Effect



Fraser Columns



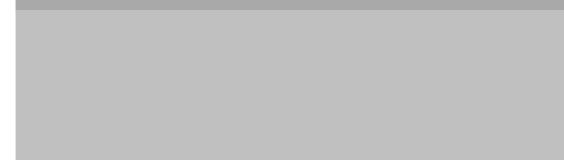
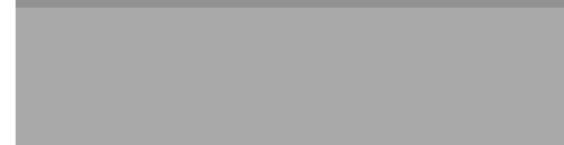
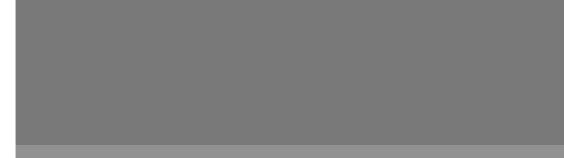
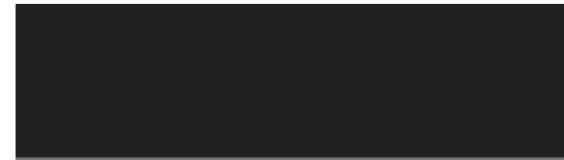
Fraser Diamonds



These are two perfectly  
geometrical circles

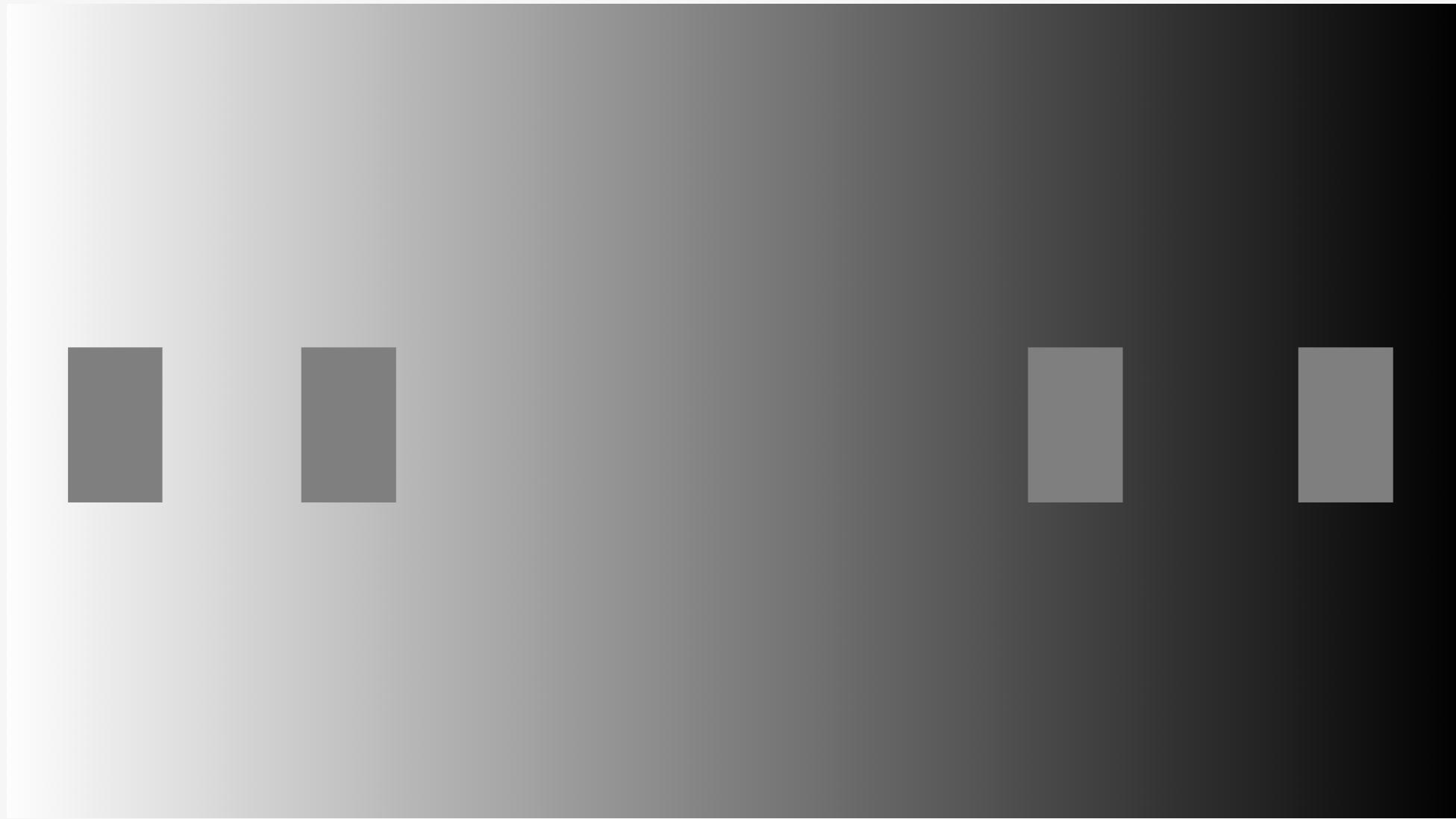


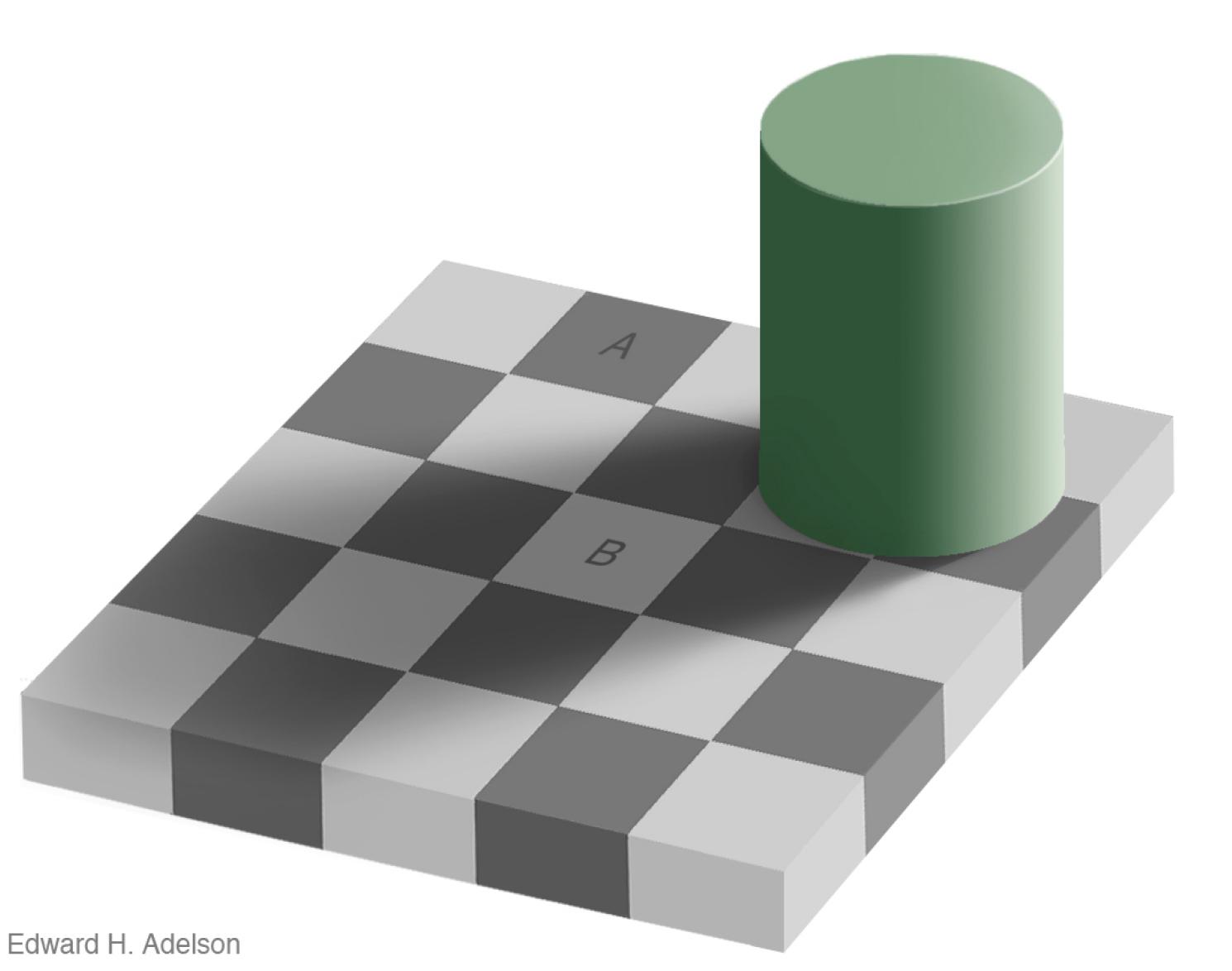
Mach Bands



Mach Bands

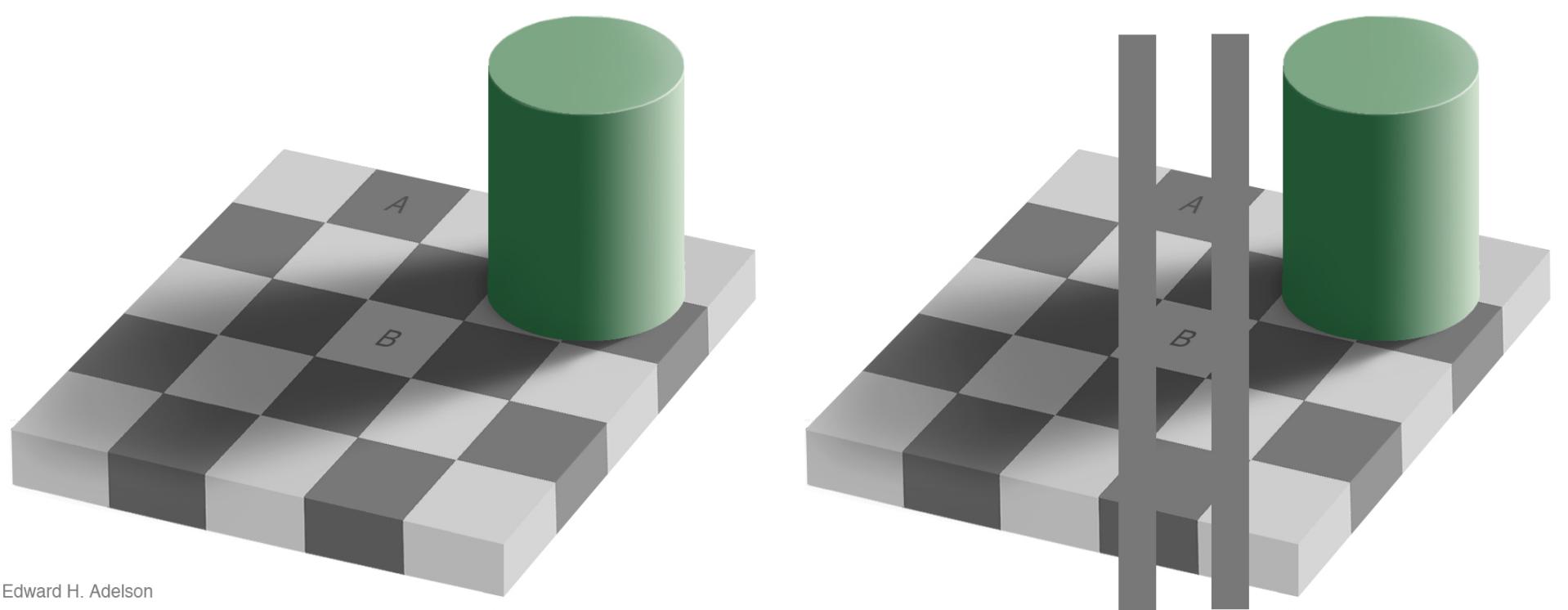






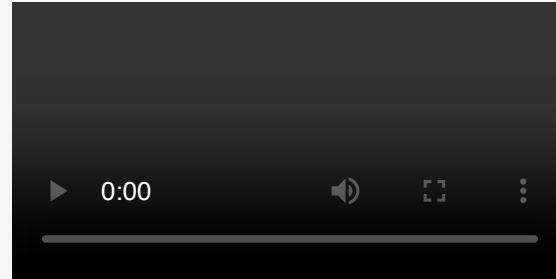
Edward H. Adelson

Edward Adelson



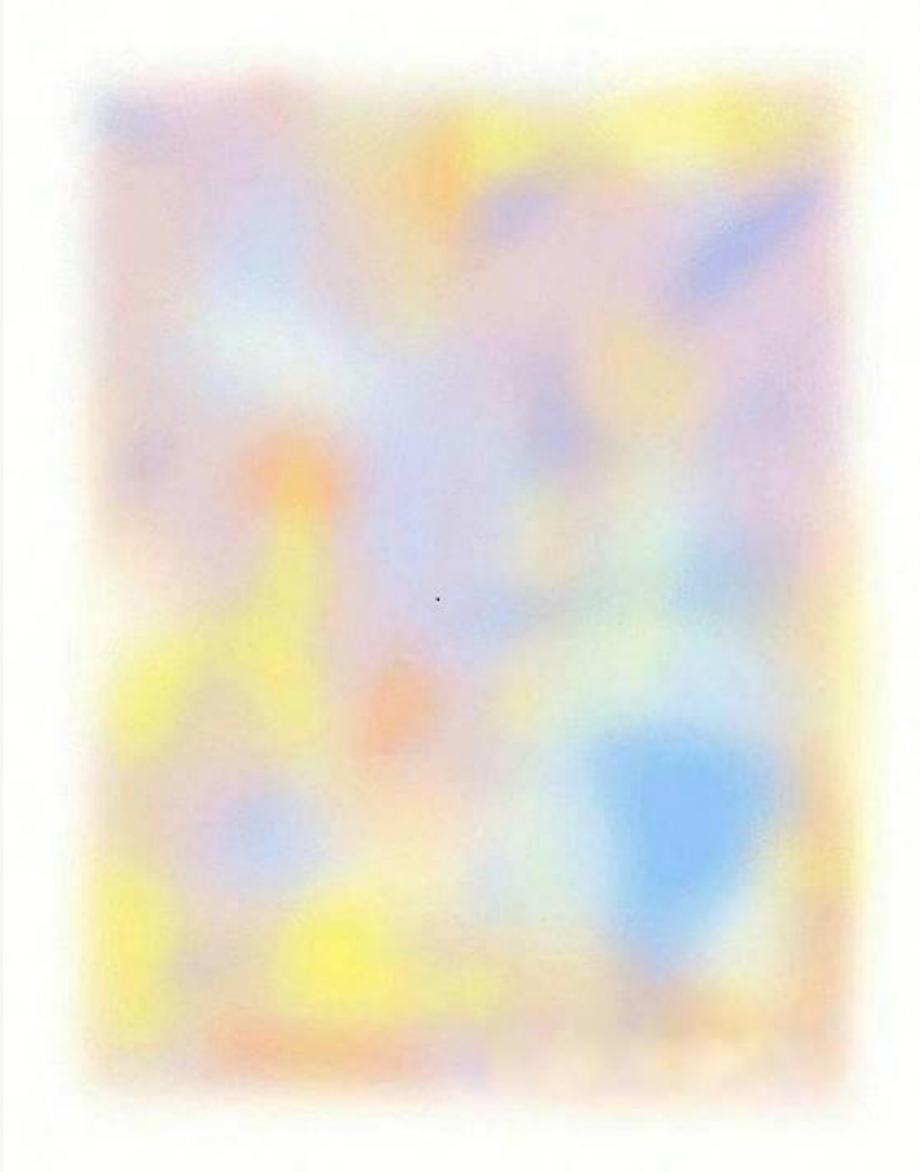
Edward H. Adelson

Edward Adelson

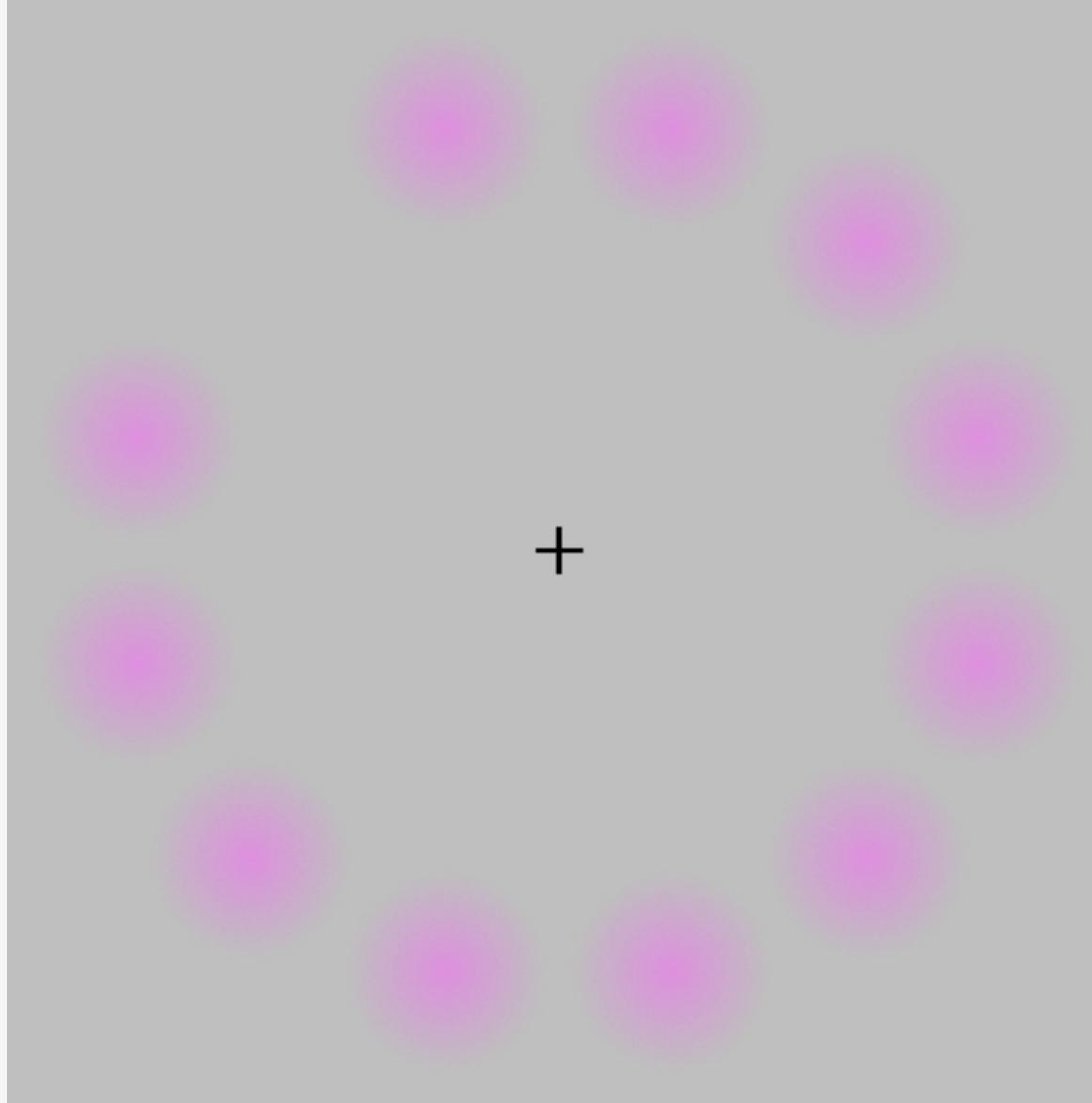


Edward Adelson

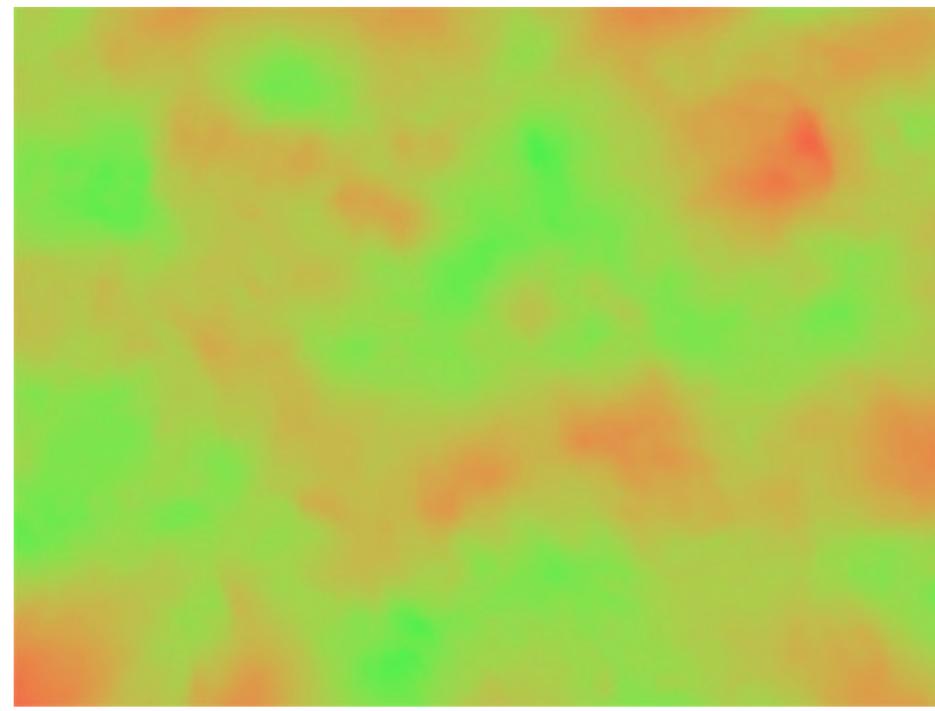
# Luminance and Color



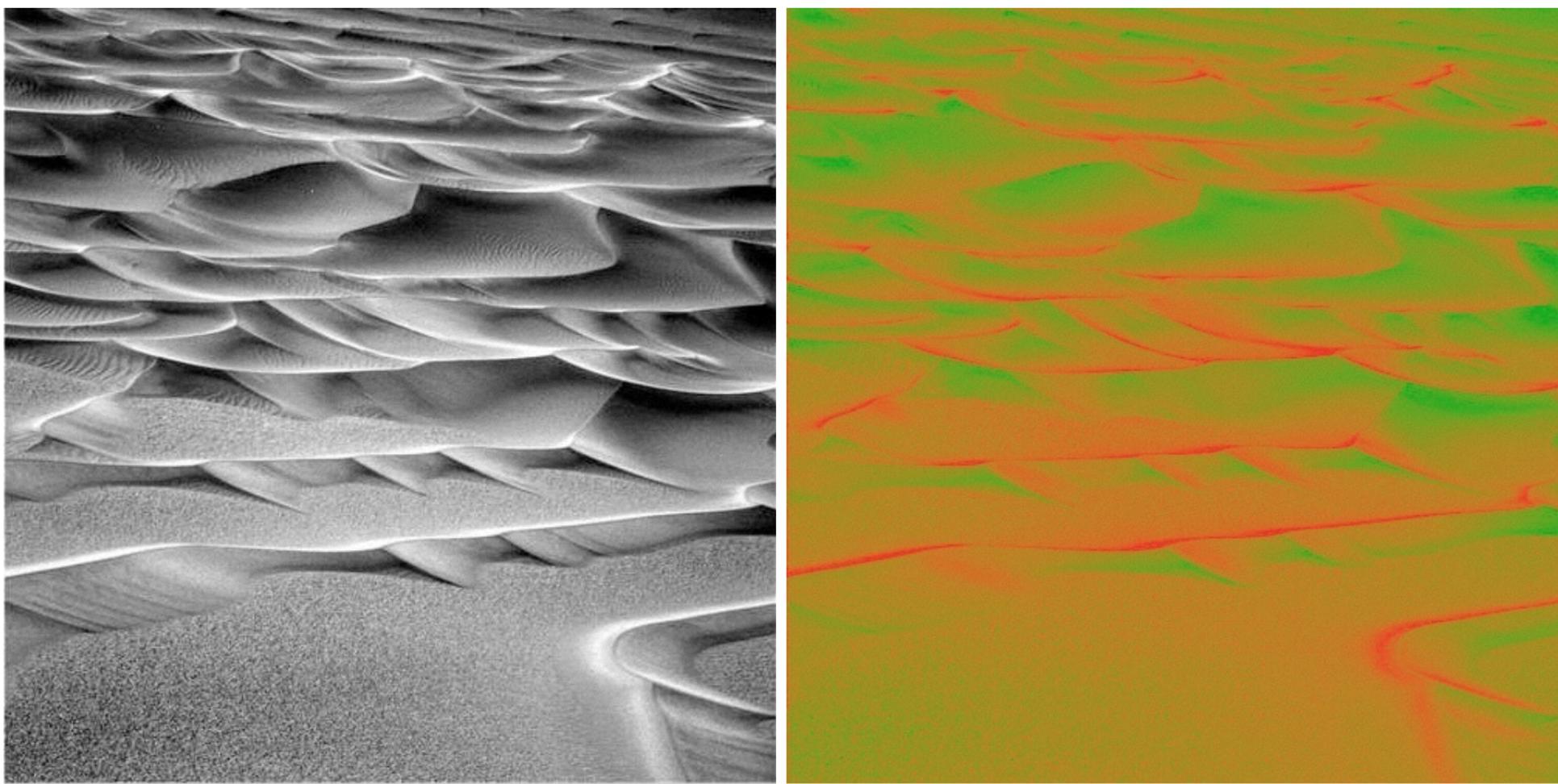
Troxler effect



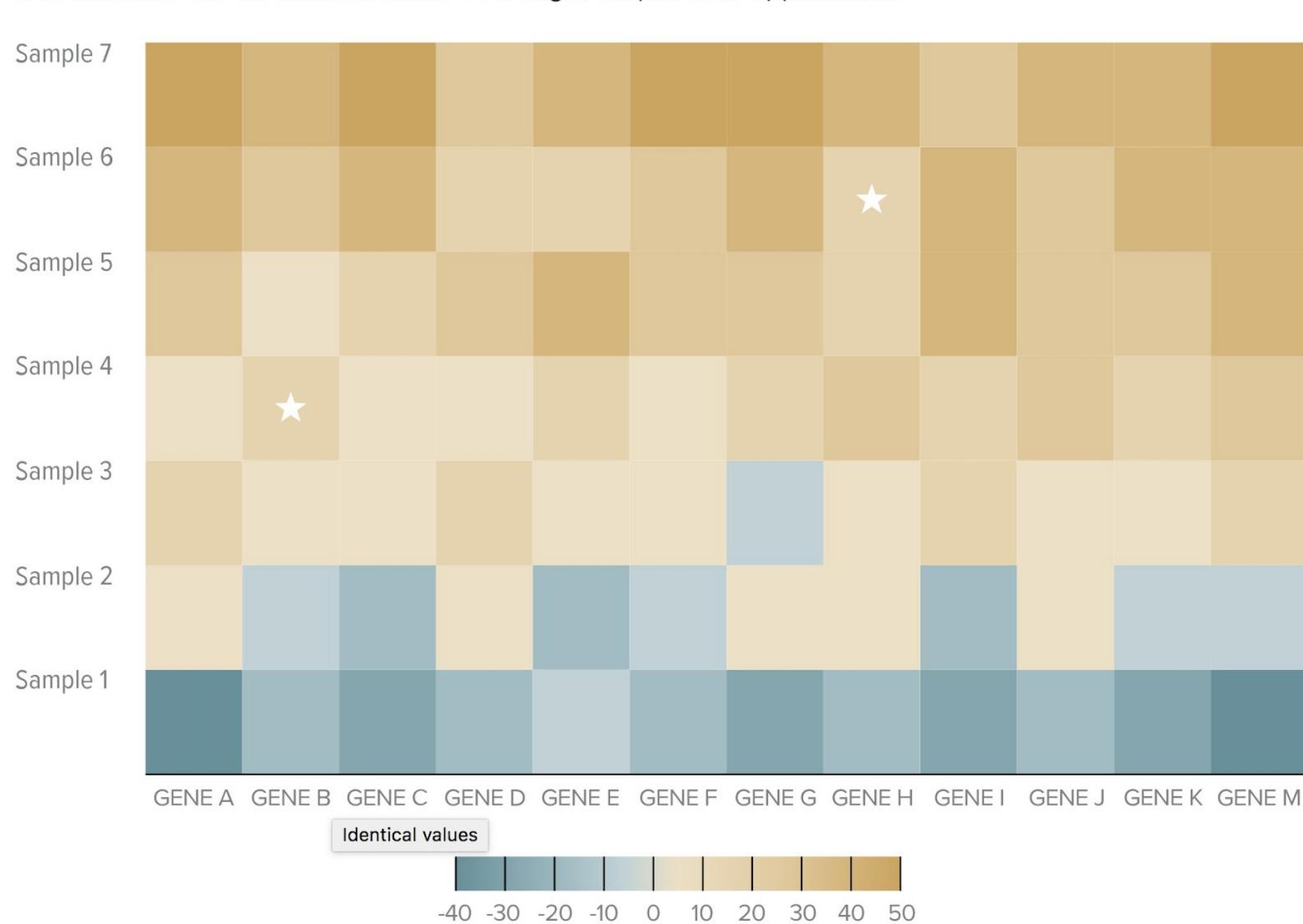
Lilac Chaser



Colin Ware

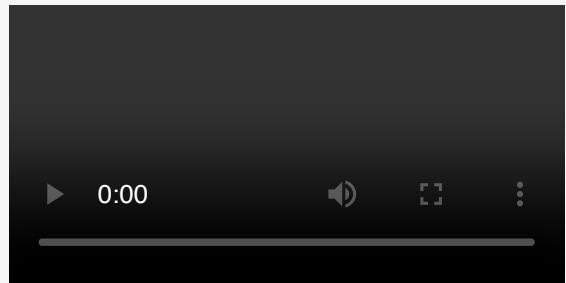


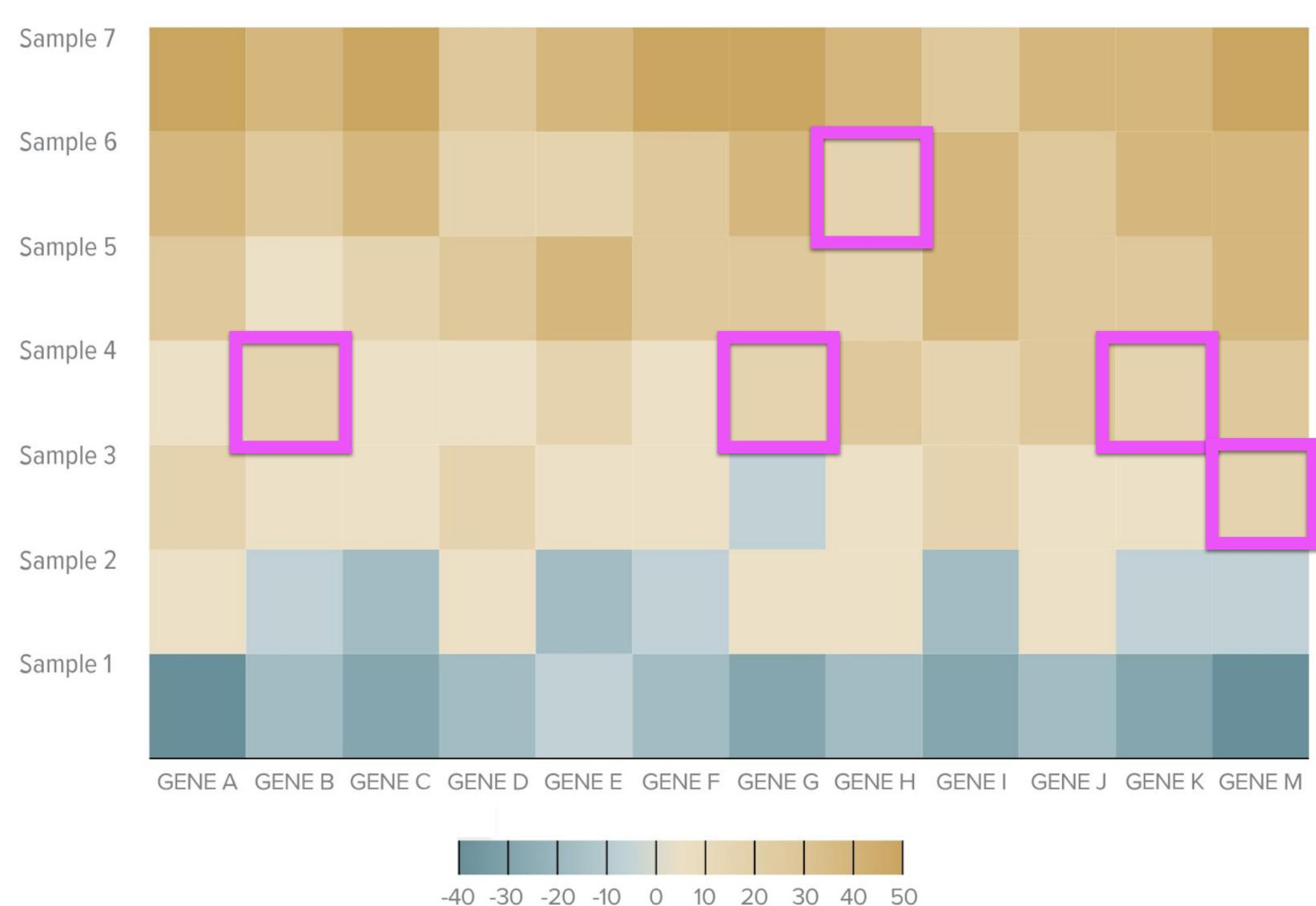
Colin Ware



SOURCE: H.E. GRECCO ET AL / NATURE METHODS 2010

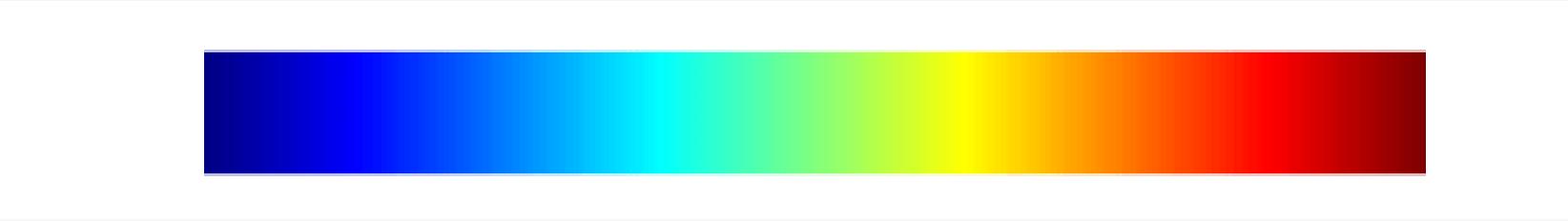
5W INFOGRAPHIC / KNOWABLE

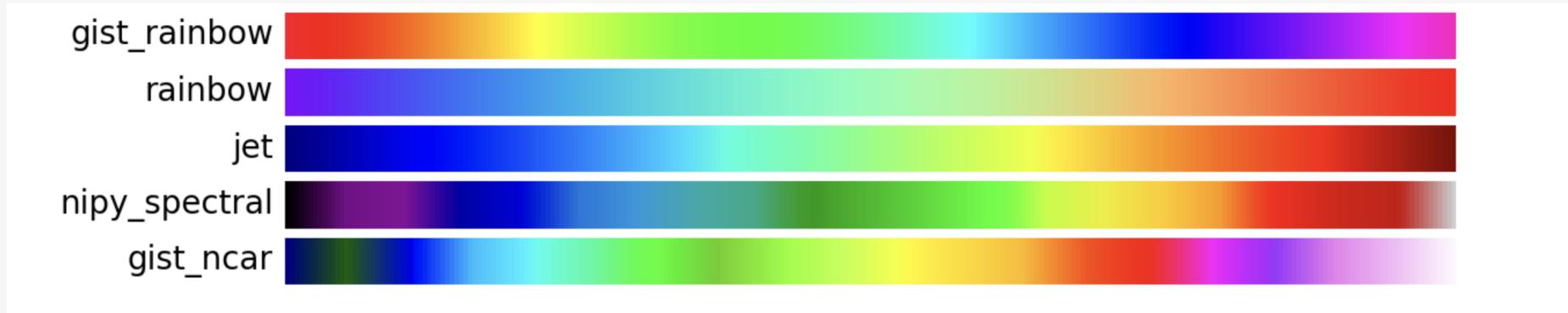


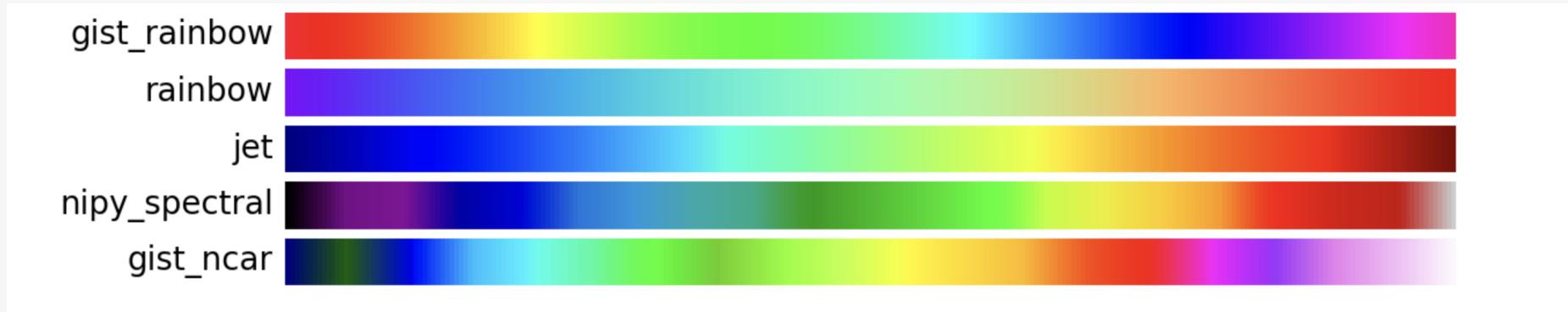


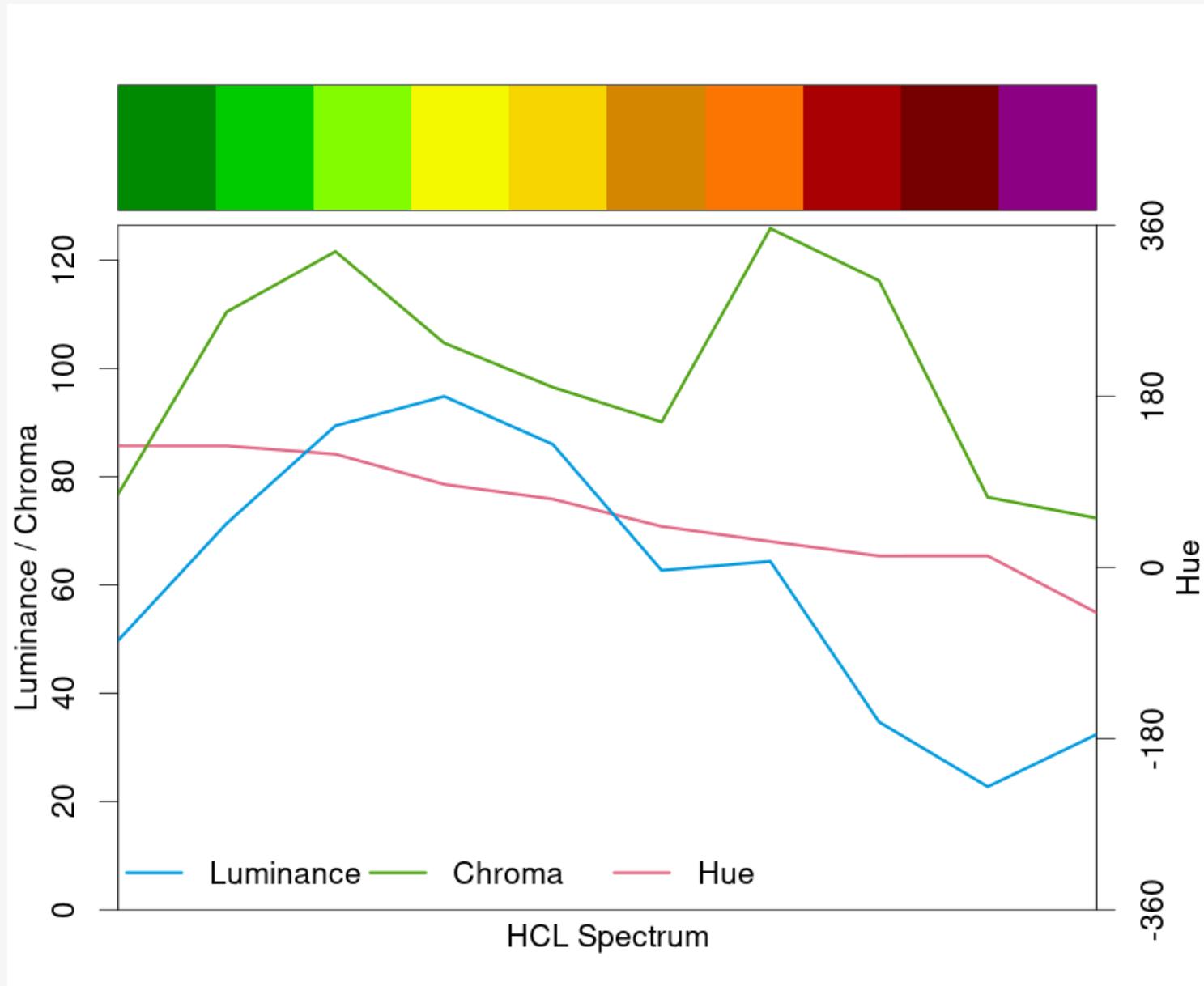
SOURCE: H.E. GRECCO ET AL / NATURE METHODS 2010

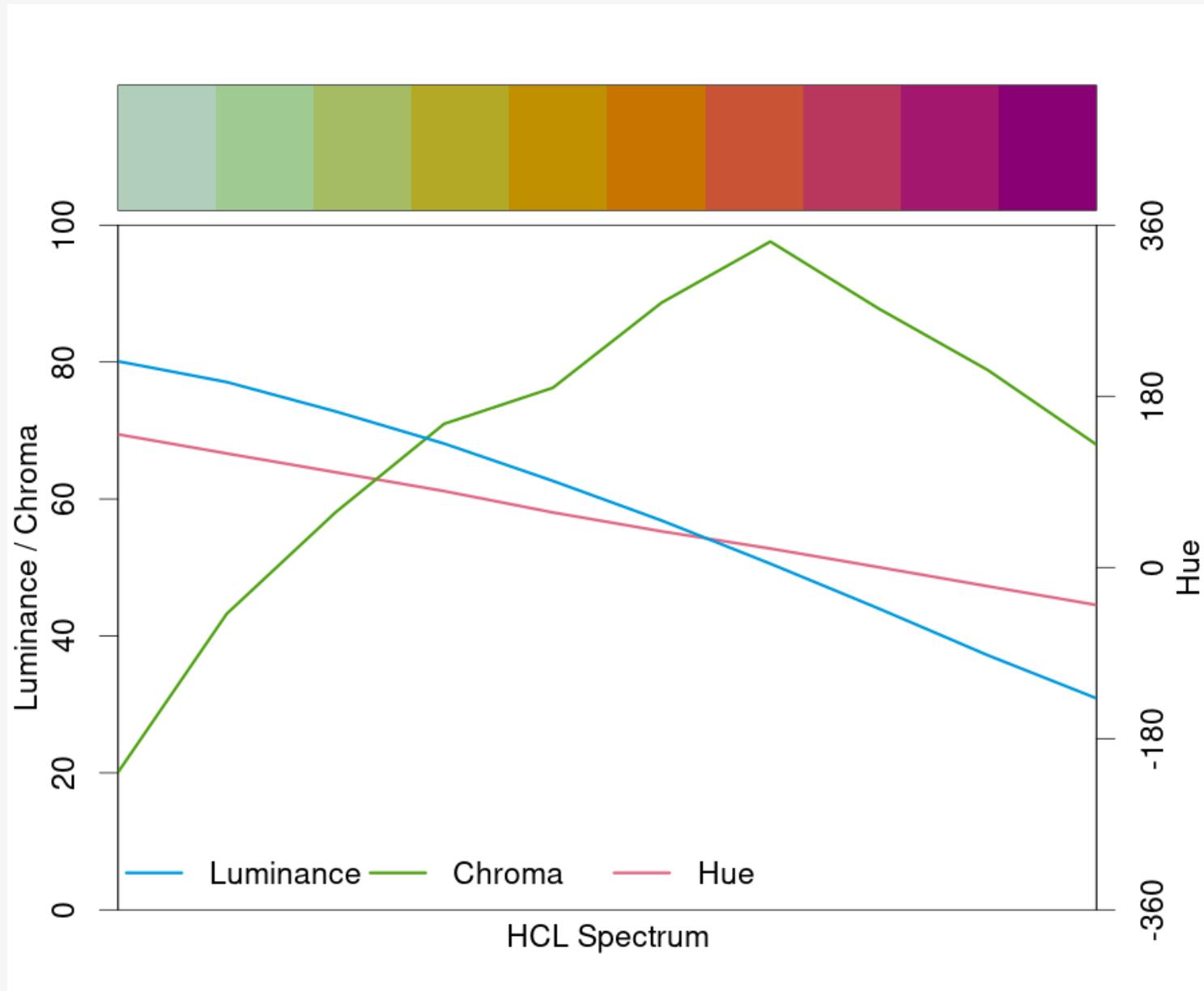
5W INFOGRAPHIC / KNOWABLE

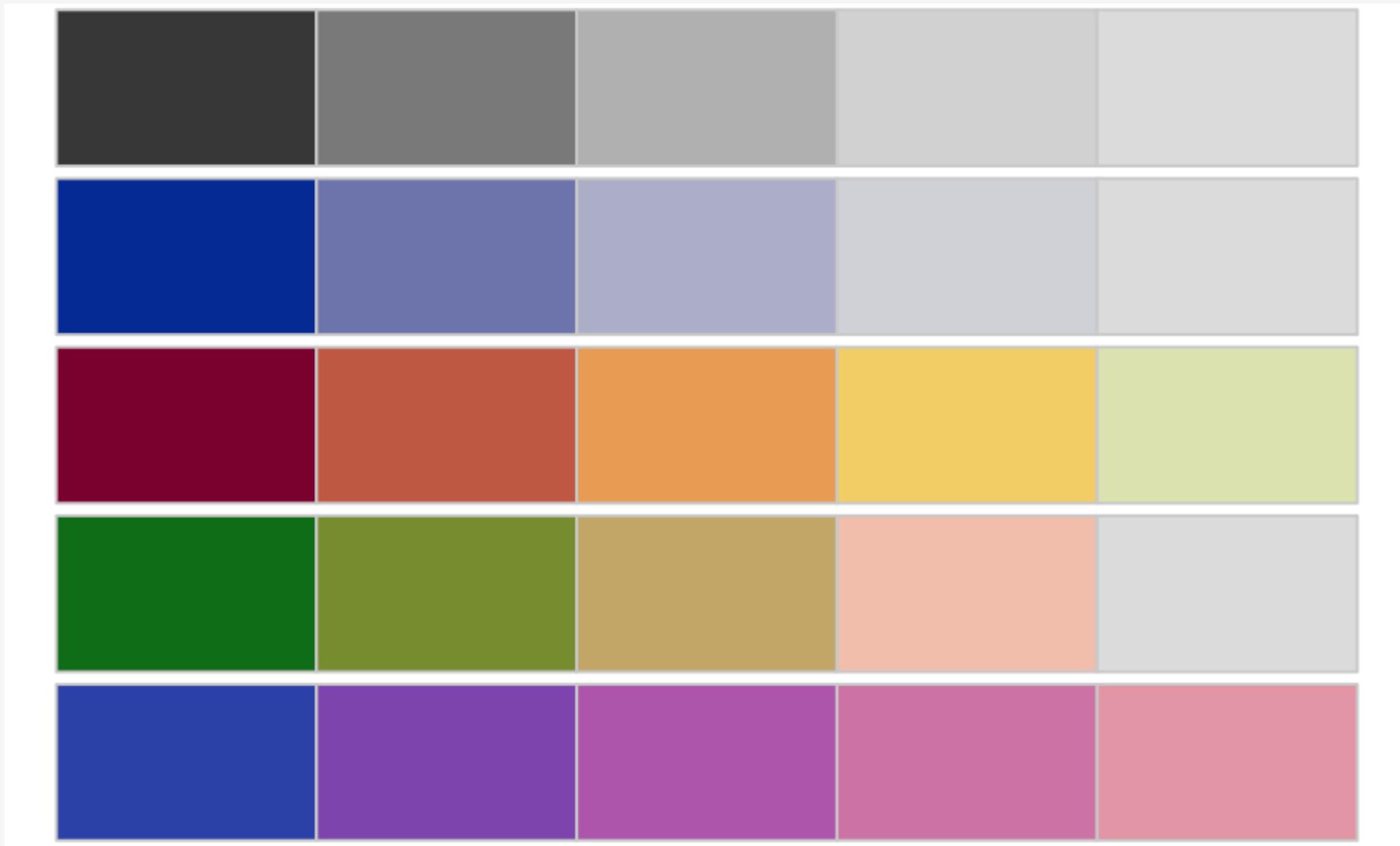


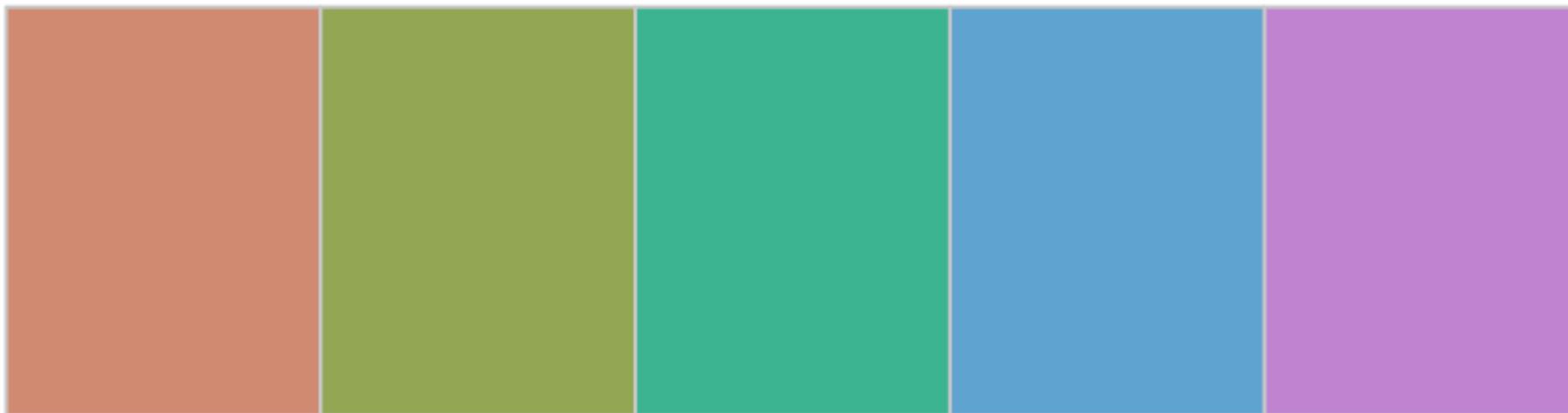
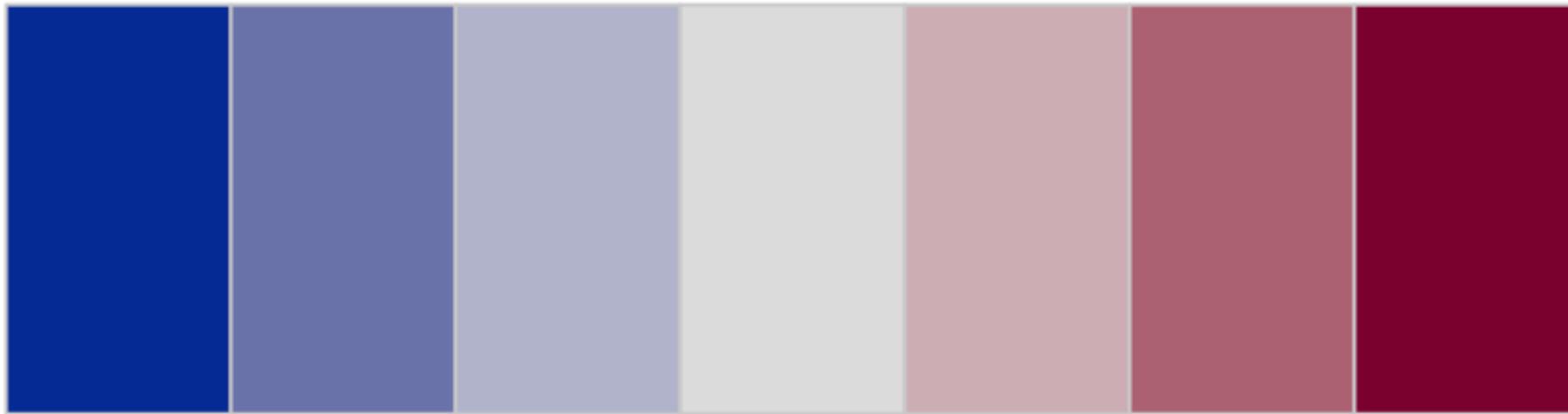


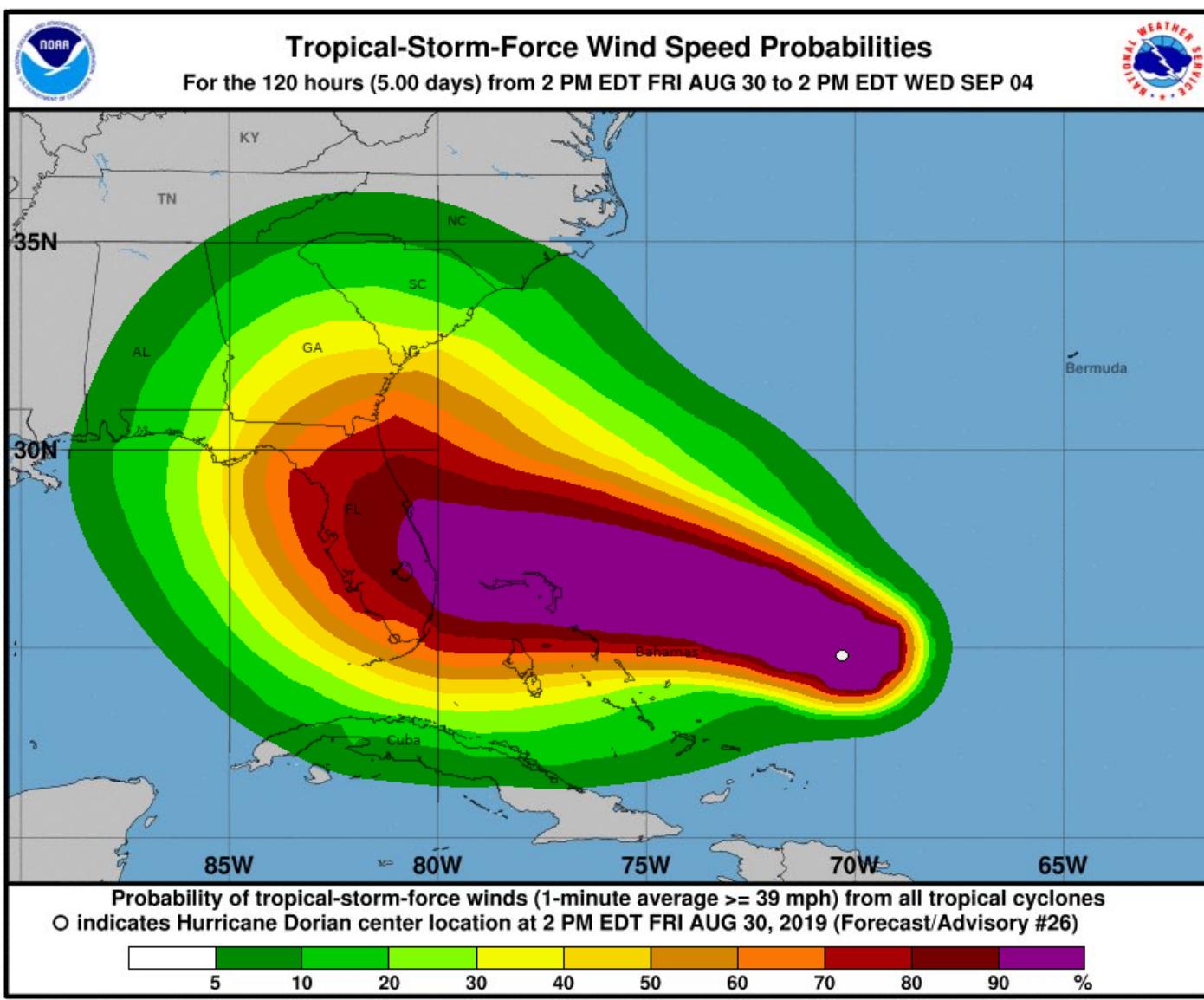




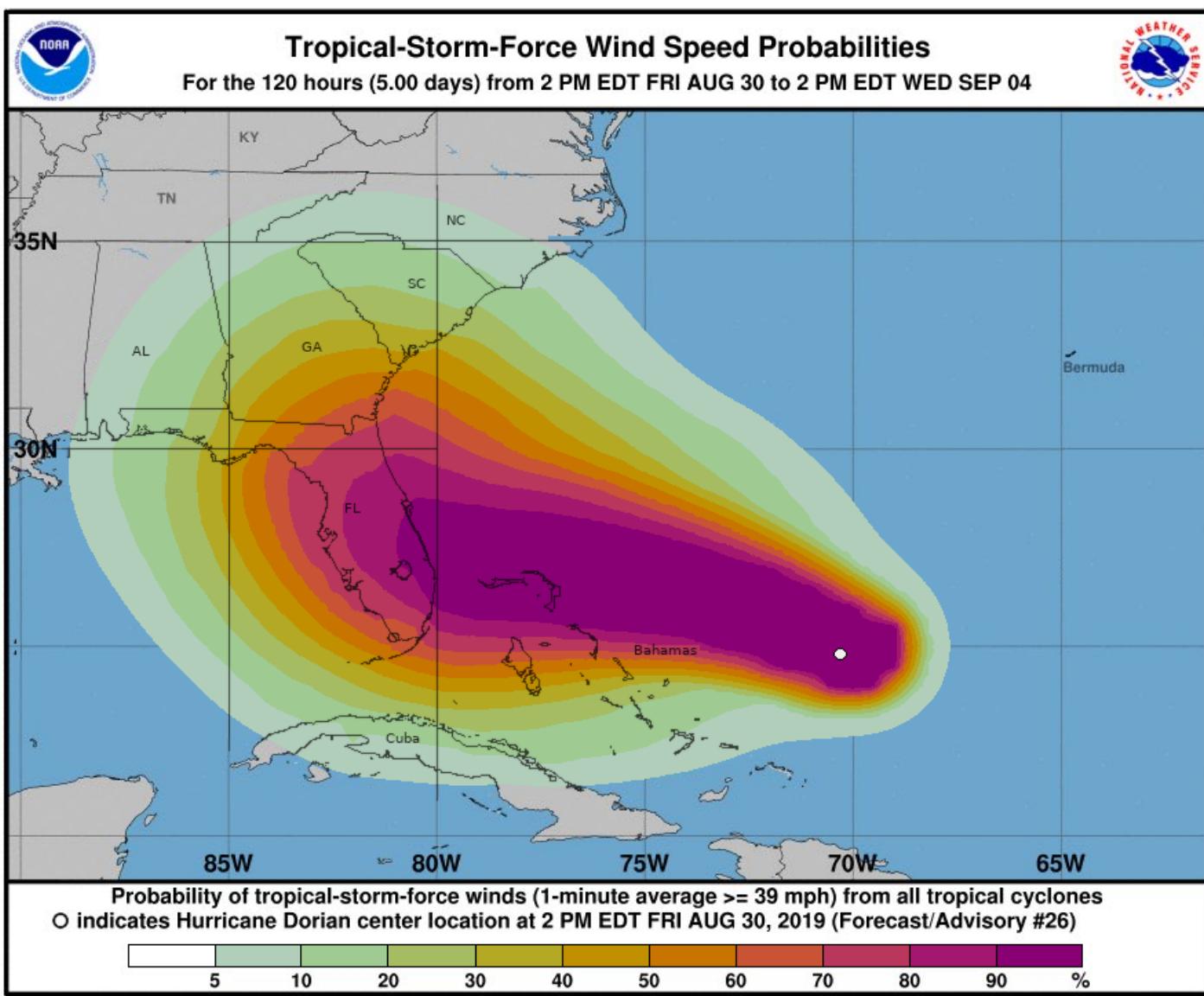




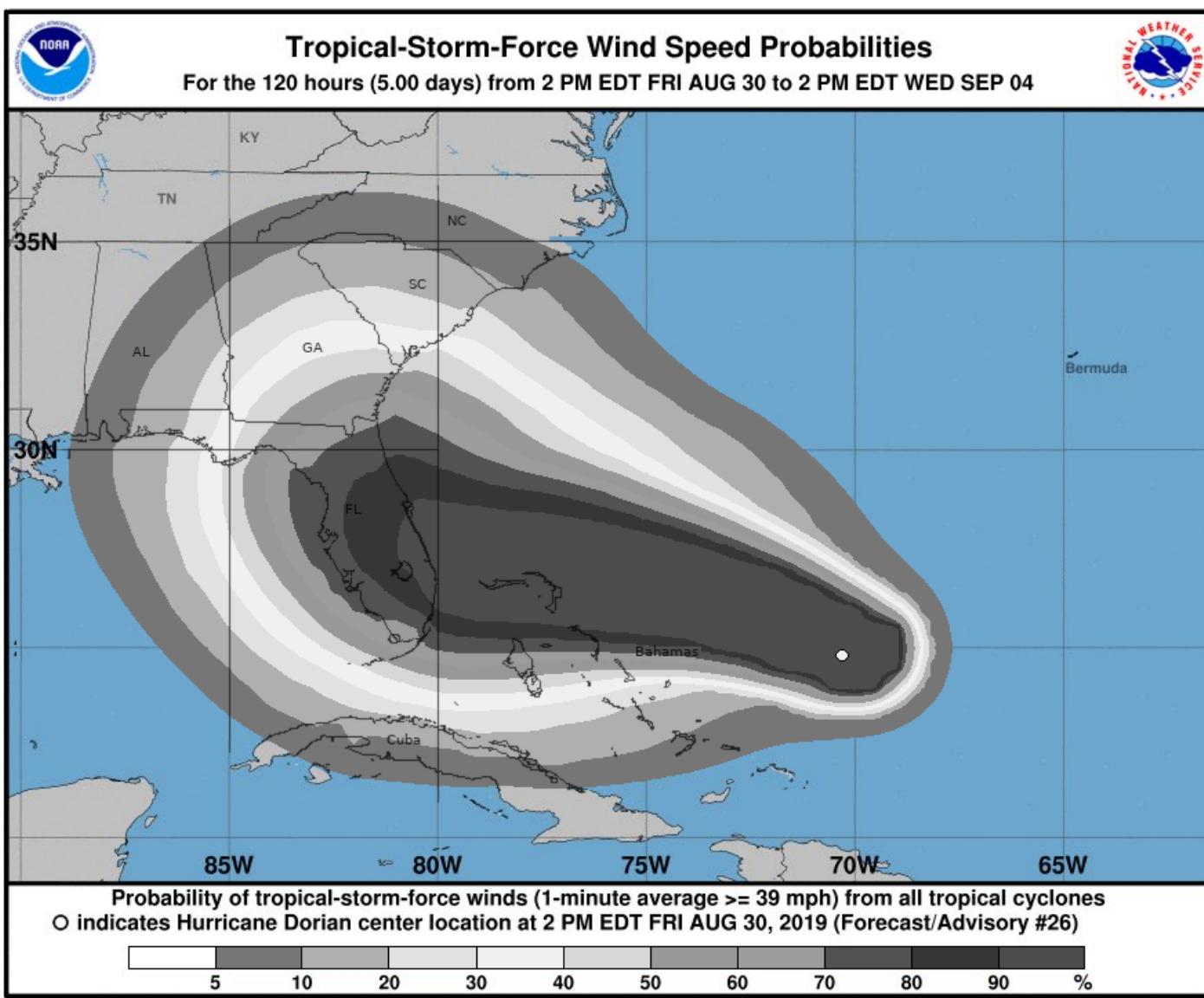




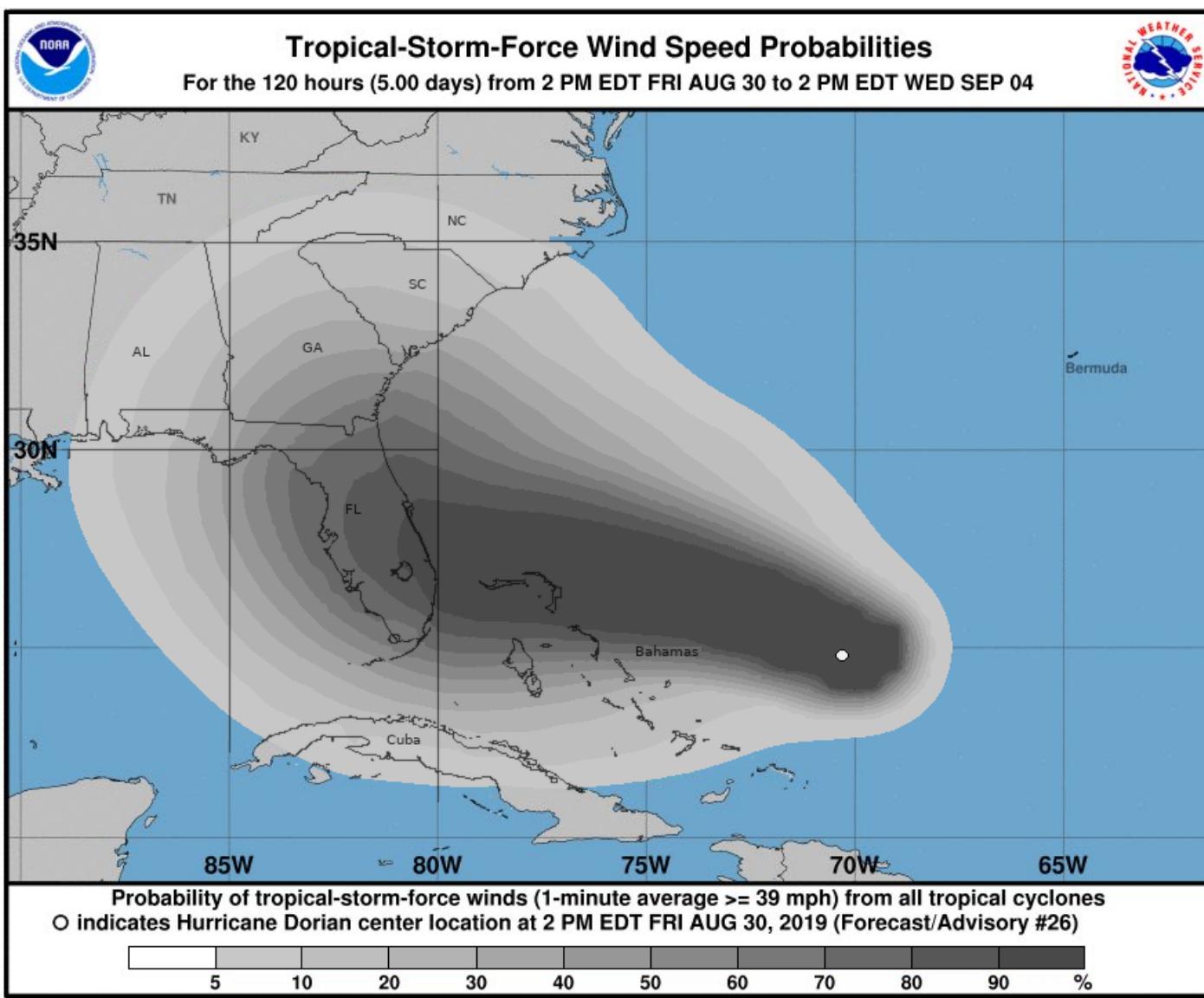
National Weather Service



Achim Zeileis



Achim Zeileis



Achim Zeileis