

# Data Visualization - 1.

# Looking at Data (a)

Kieran Healy  
Code Horizons

October 2024

# 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 4.4.1 (2024-06-14) -- "Race for Your Life"  
Copyright (C) 2024 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin20

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.

> |

The screenshot shows the RStudio interface with the following components:

- Environment pane:** Shows a function named "set" with its definition: `function (name, value)`.
- Files pane:** Displays the contents of the project directory `data_visualization_notes`. The files listed are:

Name	Size	Modified
_quarto.yml	48 B	Oct 2, 2024, 3:17 PM
.gitignore	40 B	Oct 2, 2024, 3:17 PM
.Rhistory	0 B	Oct 3, 2024, 7:15 PM
code		
data		
data_visualization_notes.qmd	892 B	Oct 3, 2024, 7:15 PM
data_visualization_notes.Rproj	258 B	Oct 3, 2024, 7:17 PM
pdf_slides		
README.md	260 B	Oct 2, 2024, 3:20 PM

data\_visualization\_notes - main - RStudio

data\_visualization\_notes.qmd

Source Visual

```
1 ---  
2 title: "Data Visualization Notes"  
3 author: "Your Name Here"  
4 format: html  
5 ---  
6 |  
7 ## Quarto  
8 |  
9 This is a quarto file. Quarto enables you to weave together content and executable  
code into a finished document. To learn more about Quarto see <https://quarto.org>.  
10 |  
11 ## Setup  
12 |  
13 Here we load some packages that we'll be using.  
14 |  
15 ``{r}  
16 library(here)  
17 library(socviz)  
18 library(gapminder)  
19 library(tidyverse)  
20 ```  
6:1 (Top Level) Quarto
```

Console Terminal x Background Jobs x

R 4.4.1 · ~/Documents/courses/data\_visualization\_notes/

```
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.
```

>

Environment History Connections Build Tutorial Git

Import Dataset 177 MiB Global Environment

Functions set function (name, value)

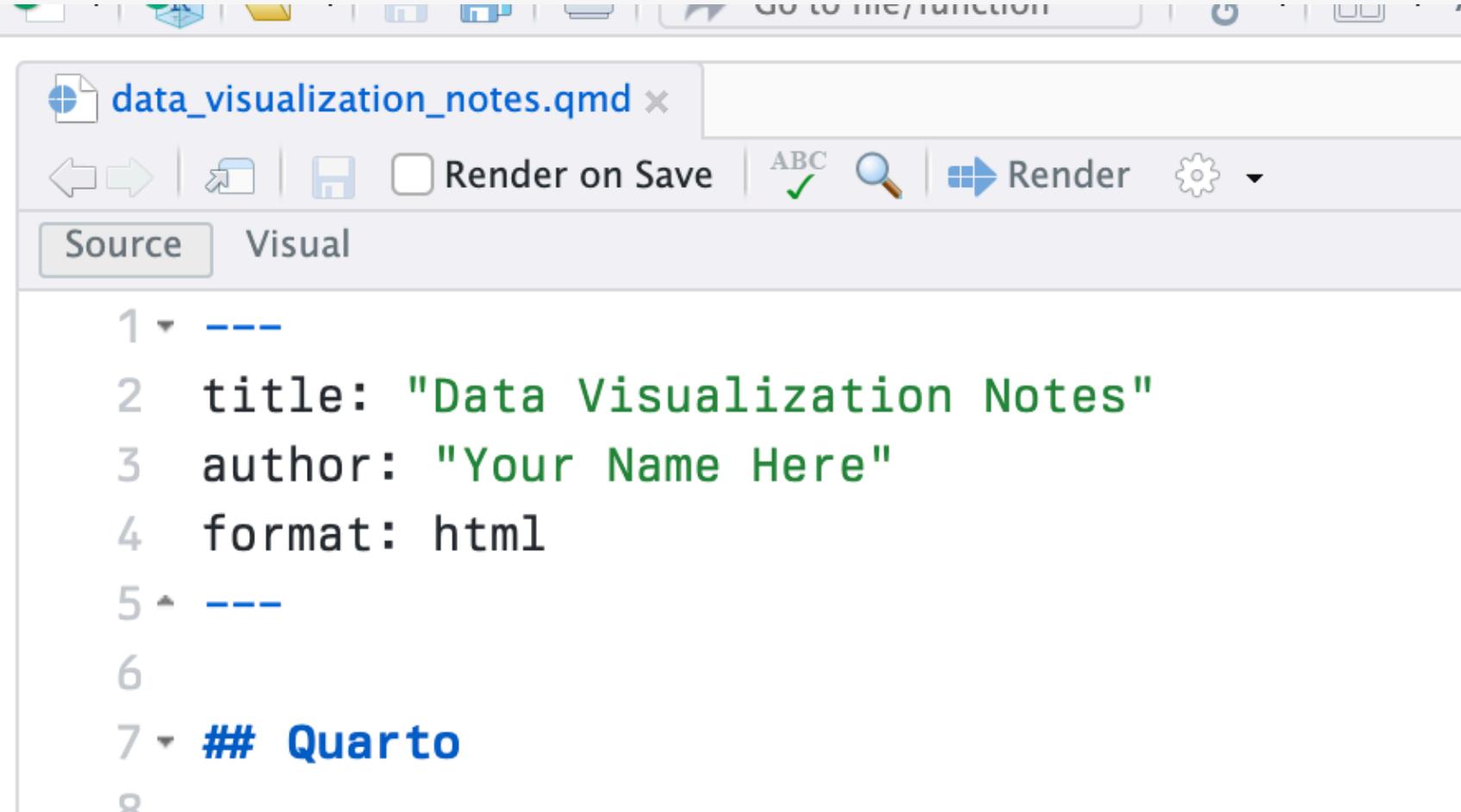
Files Plots Packages Help Viewer Presentation

New Folder New Blank File Delete Rename More

Home Documents courses data\_visualization\_notes

Name	Size	Modified
..		
_quarto.yml	48 B	Oct 2, 2024, 3:17 PM
.gitignore	40 B	Oct 2, 2024, 3:17 PM
.Rhistory	0 B	Oct 3, 2024, 7:15 PM
code		
data		
data_visualization_notes.qmd	892 B	Oct 3, 2024, 7:15 PM
data_visualization_notes.Rproj	258 B	Oct 3, 2024, 7:17 PM
pdf_slides		
README.md	260 B	Oct 2, 2024, 3:20 PM

# Try rendering your notes



The screenshot shows the RStudio interface with a Quarto document open. The title bar displays the file name "data\_visualization\_notes.qmd". The toolbar includes standard file operations and a "Render" button with a checkmark icon. Below the toolbar, there are two tabs: "Source" (which is selected) and "Visual". The main pane contains the following code:

```
1 ---  
2 title: "Data Visualization Notes"  
3 author: "Your Name Here"  
4 format: html  
5 ---  
6  
7 ## Quarto
```

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.

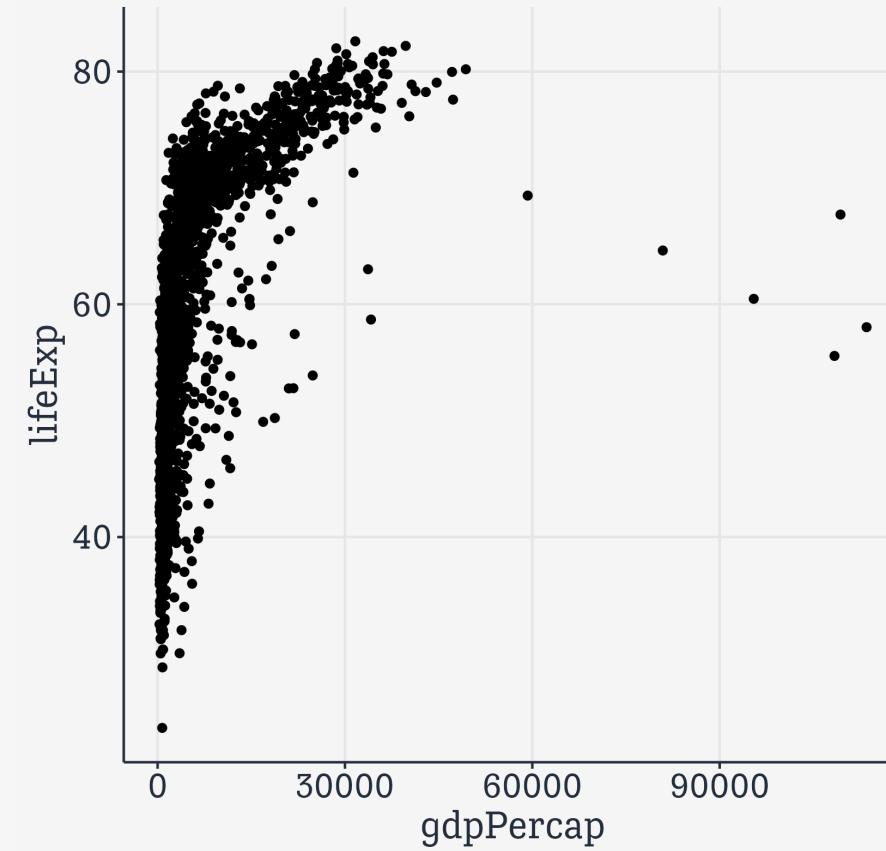
```
p ← ggplot(data = gapminder,  
            mapping = aes(x = gdpPercap,  
                           y = lifeExp))  
  
p + geom_point()
```

... And Render your document again.

# Now write the following code

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

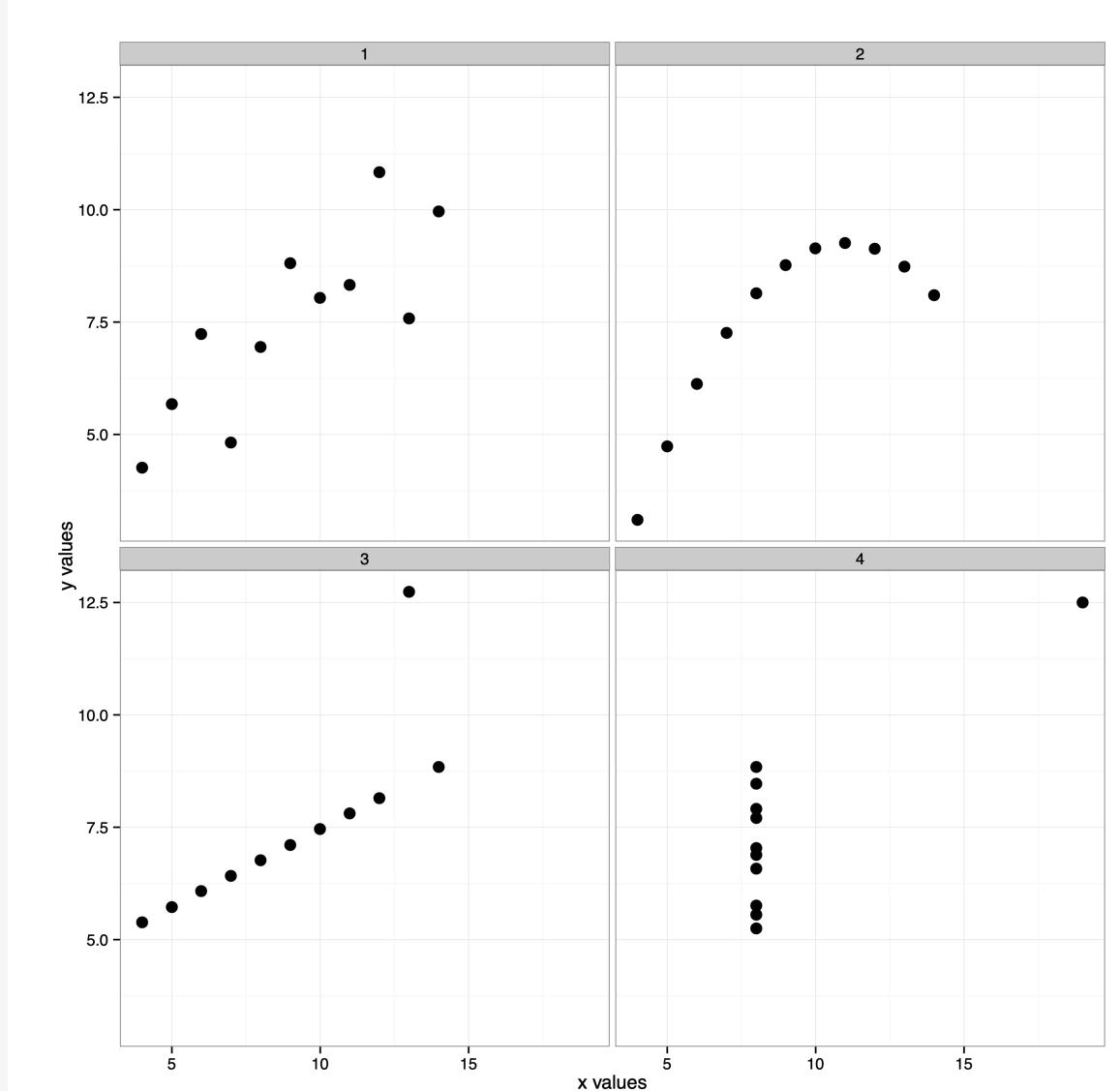
```
p ← ggplot(data = gapminder,  
            mapping = aes(x = gdpPercap,  
                           y = lifeExp))  
  
p + geom_point()
```



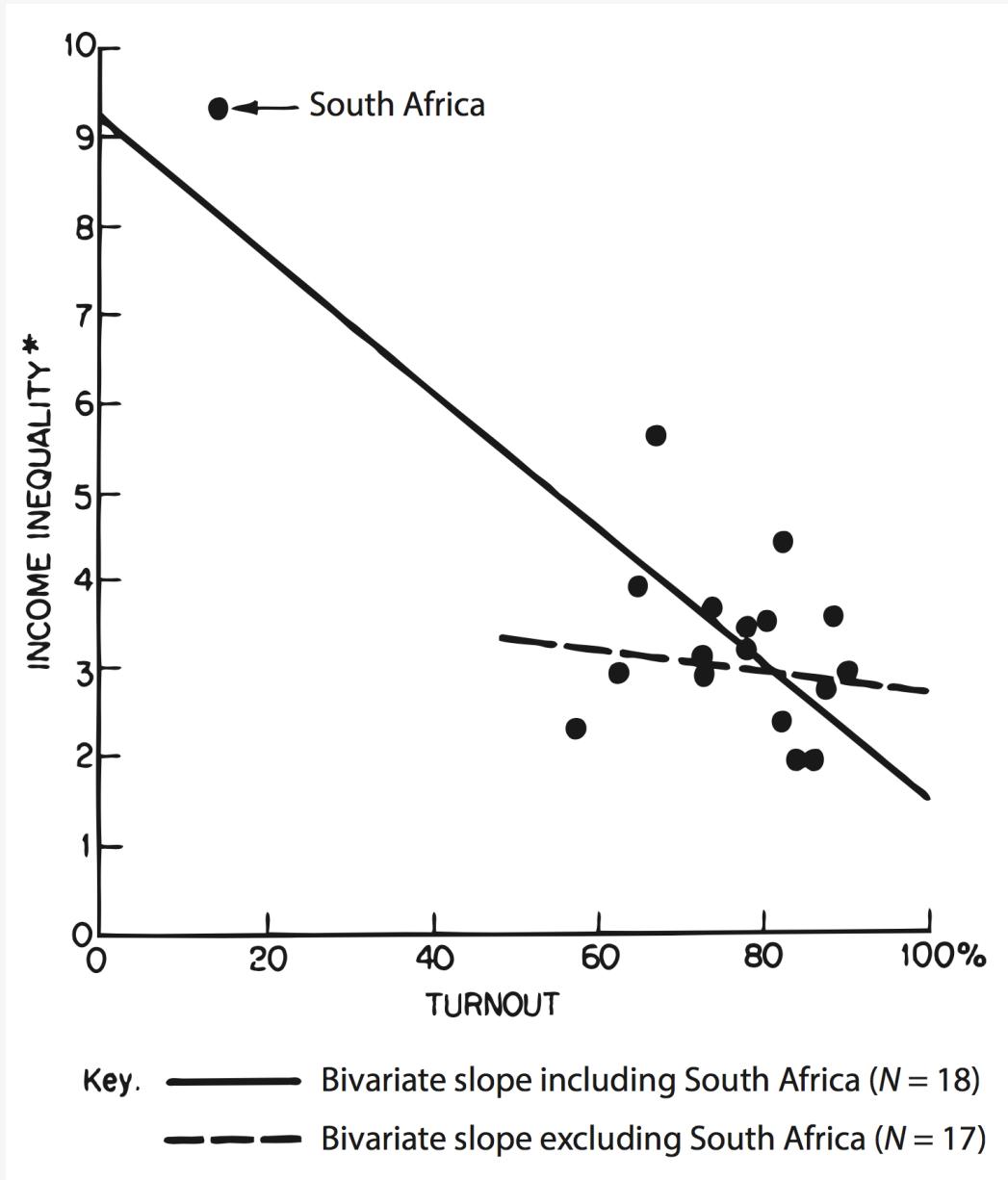
... And Render your document again.

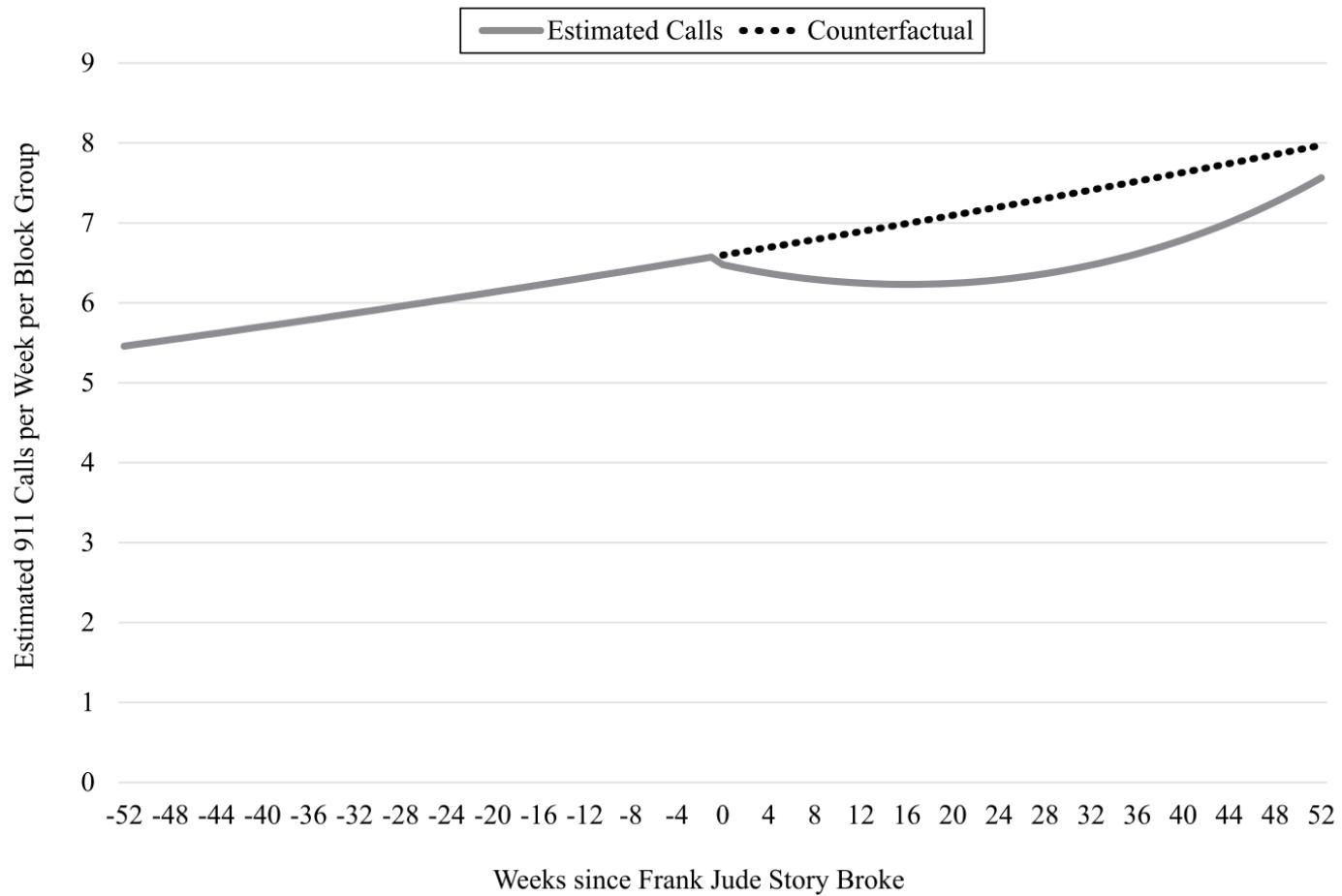
You should  
look at  
your data

# Seeing things

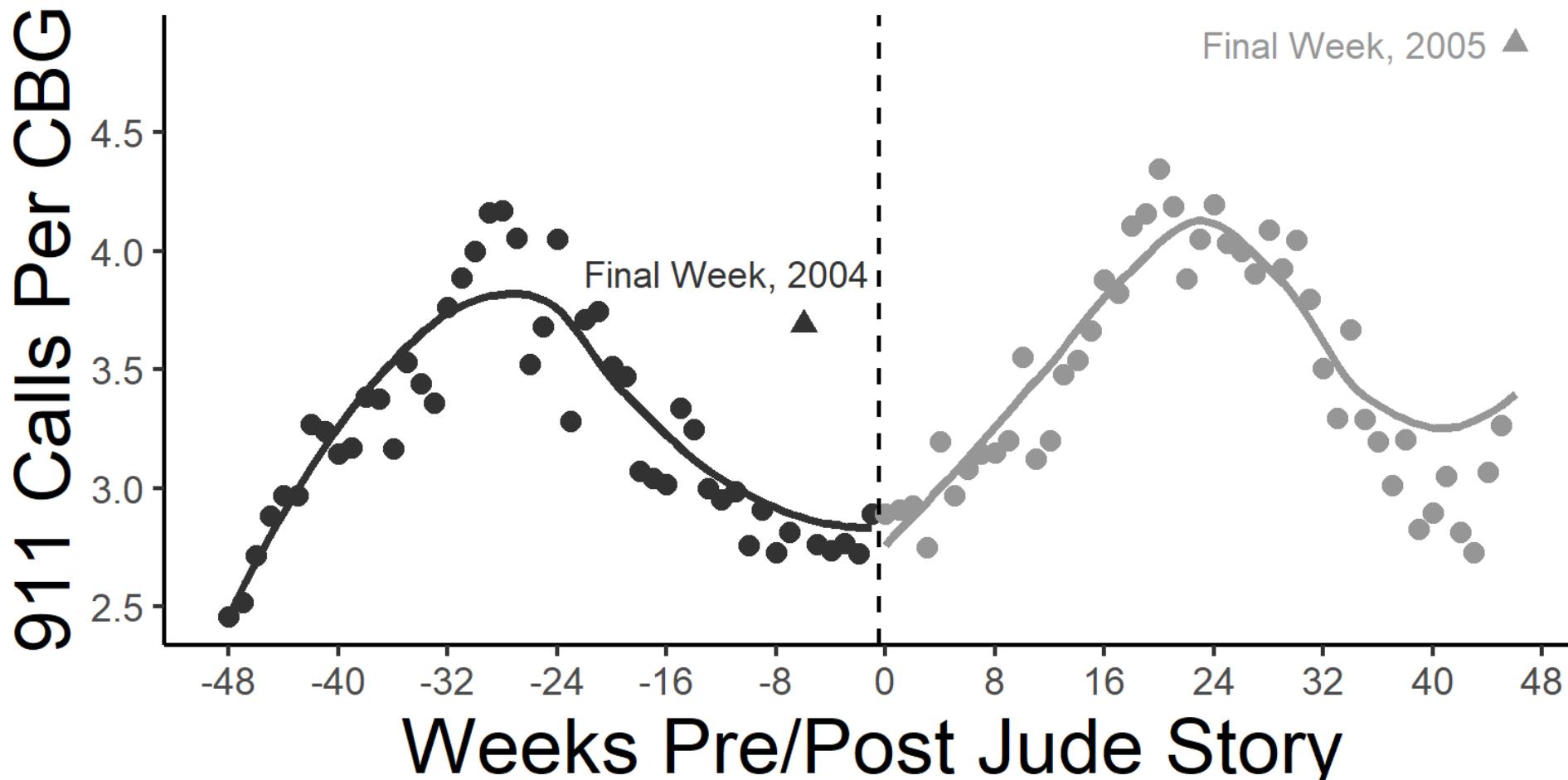


Anscombe's Quartet

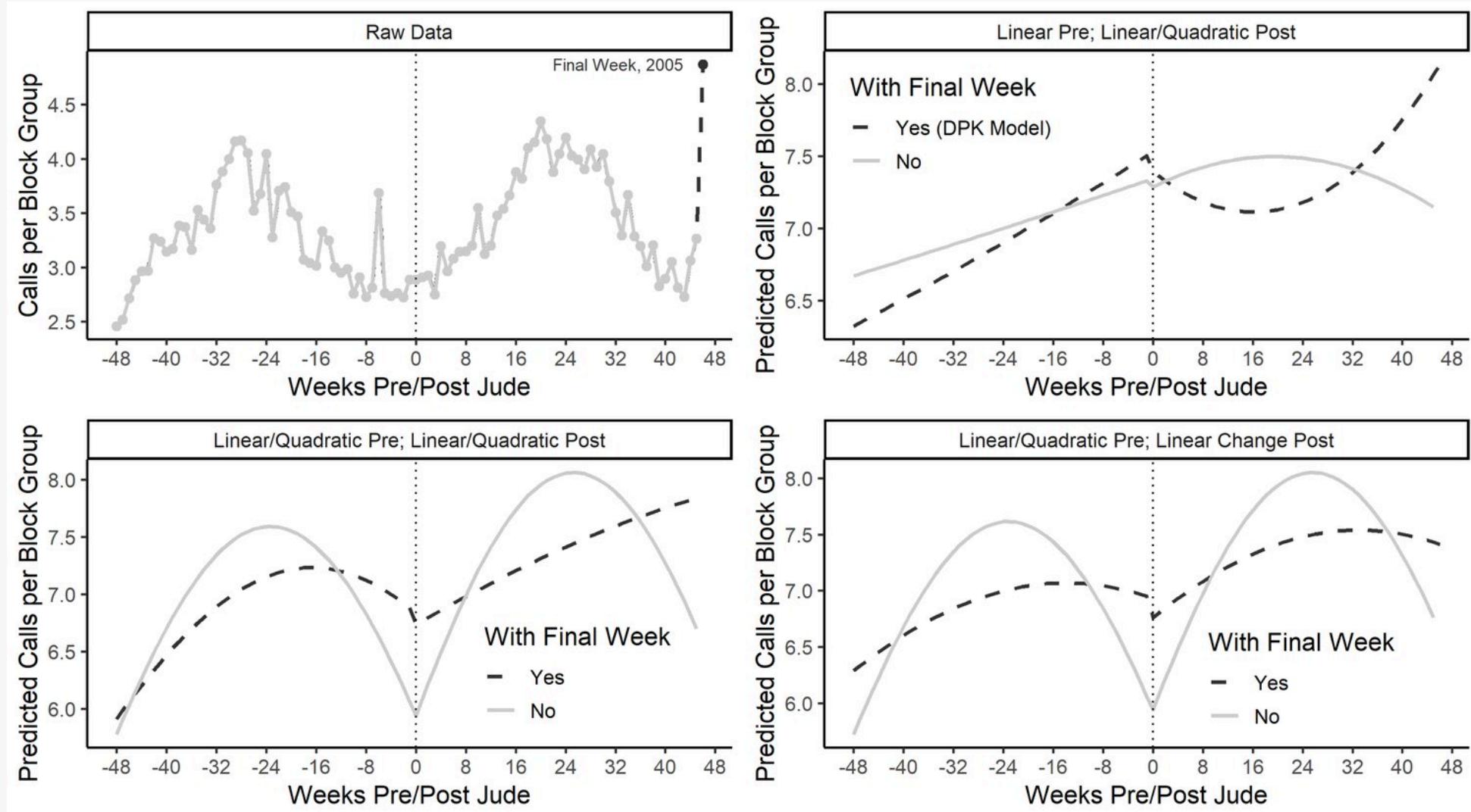


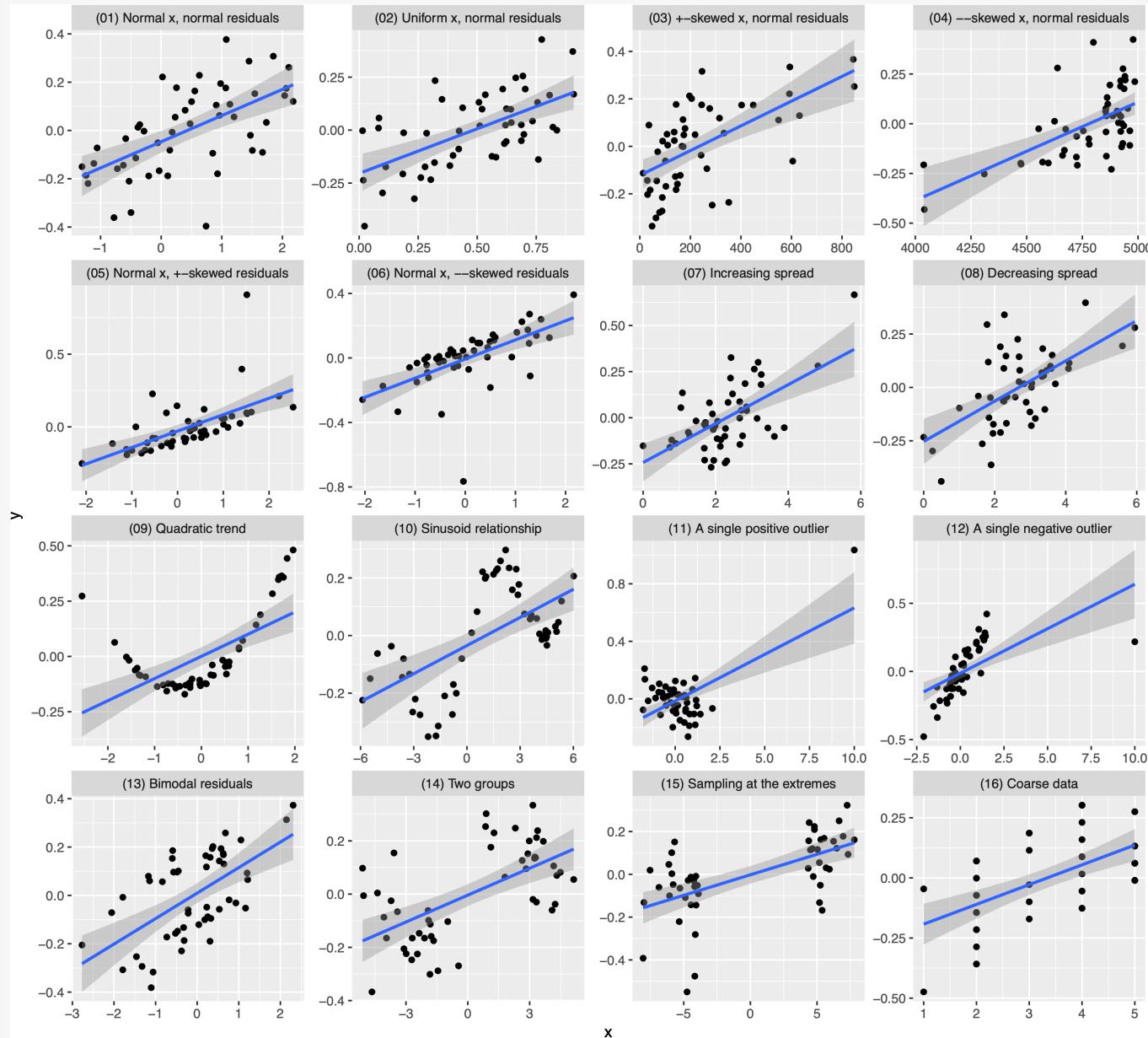


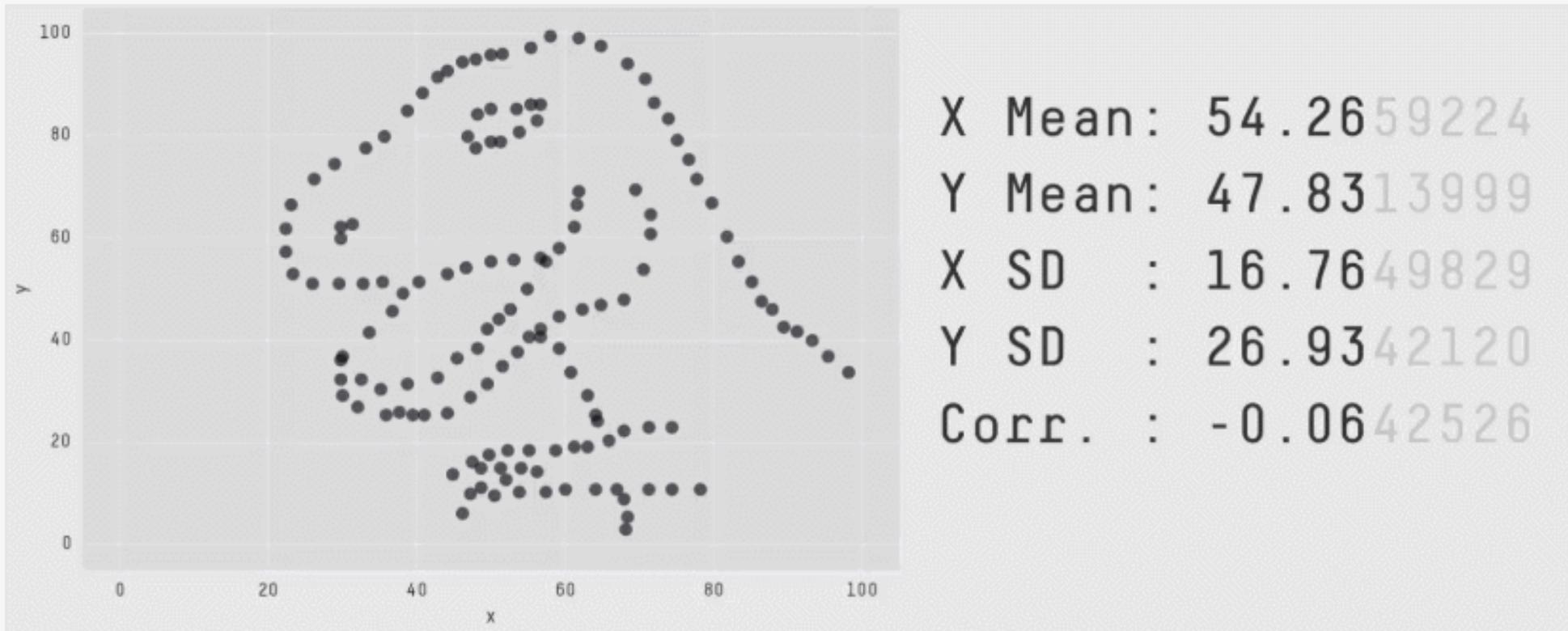
# All Neighborhoods



Zoorob (2020)

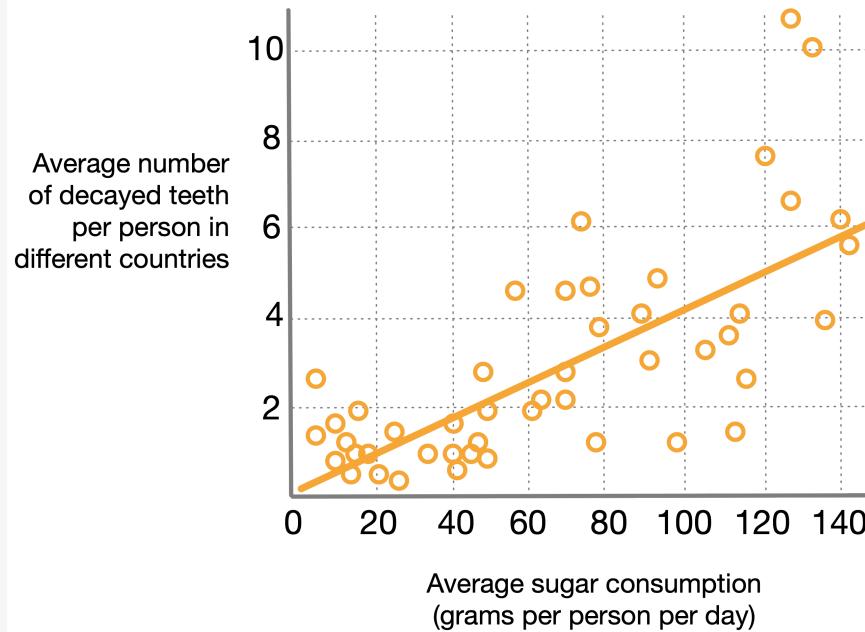






Cairo; Matejka & Fitzmaurice

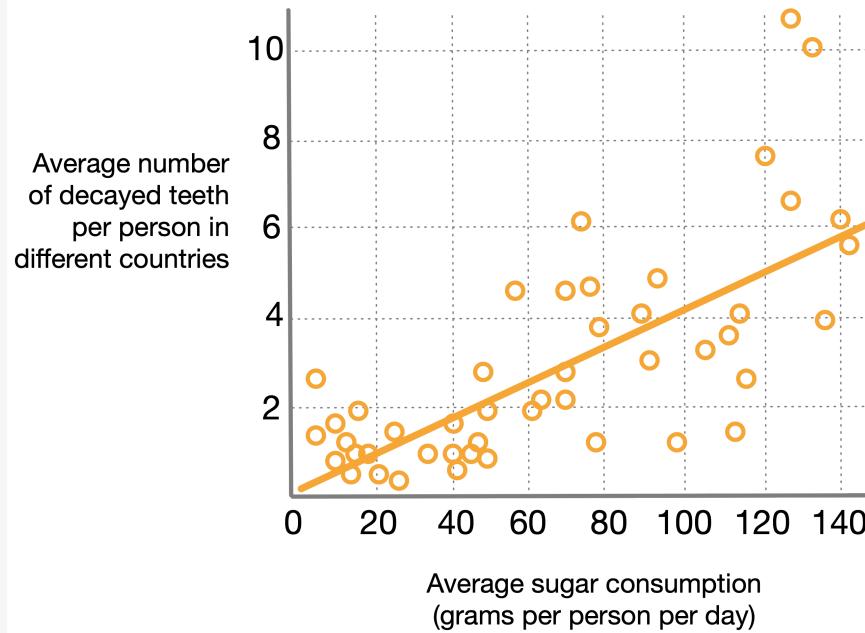
*Which of the following statements best describes the data in the graph below?*



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

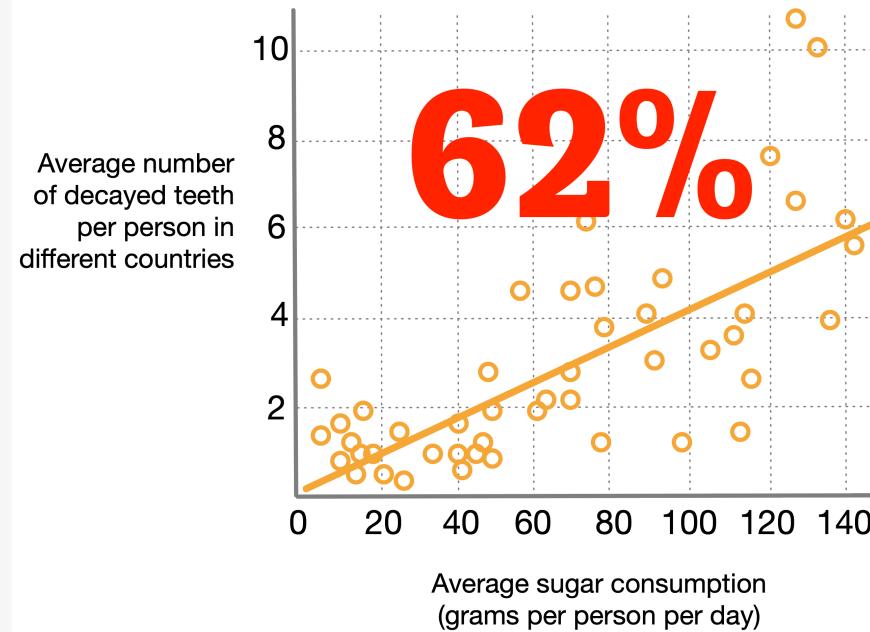
*Which of the following statements best describes the data in the graph below?*



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?*

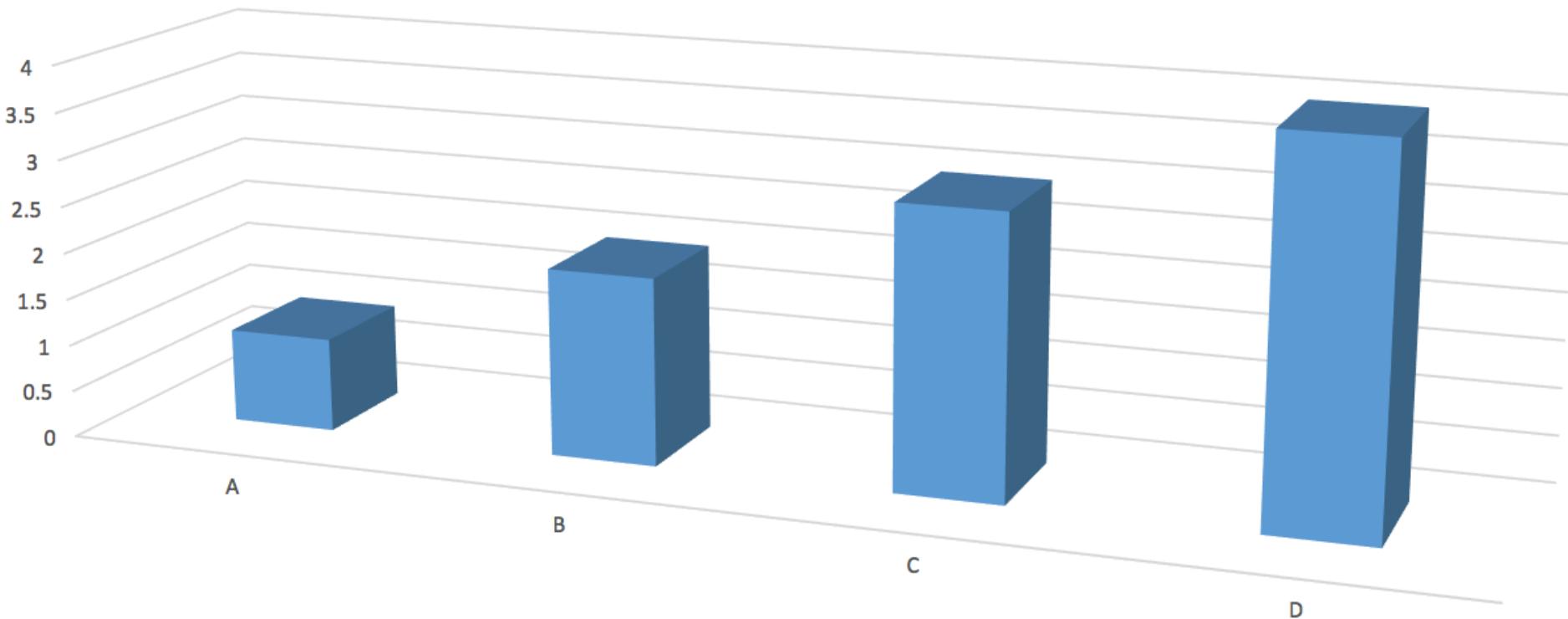


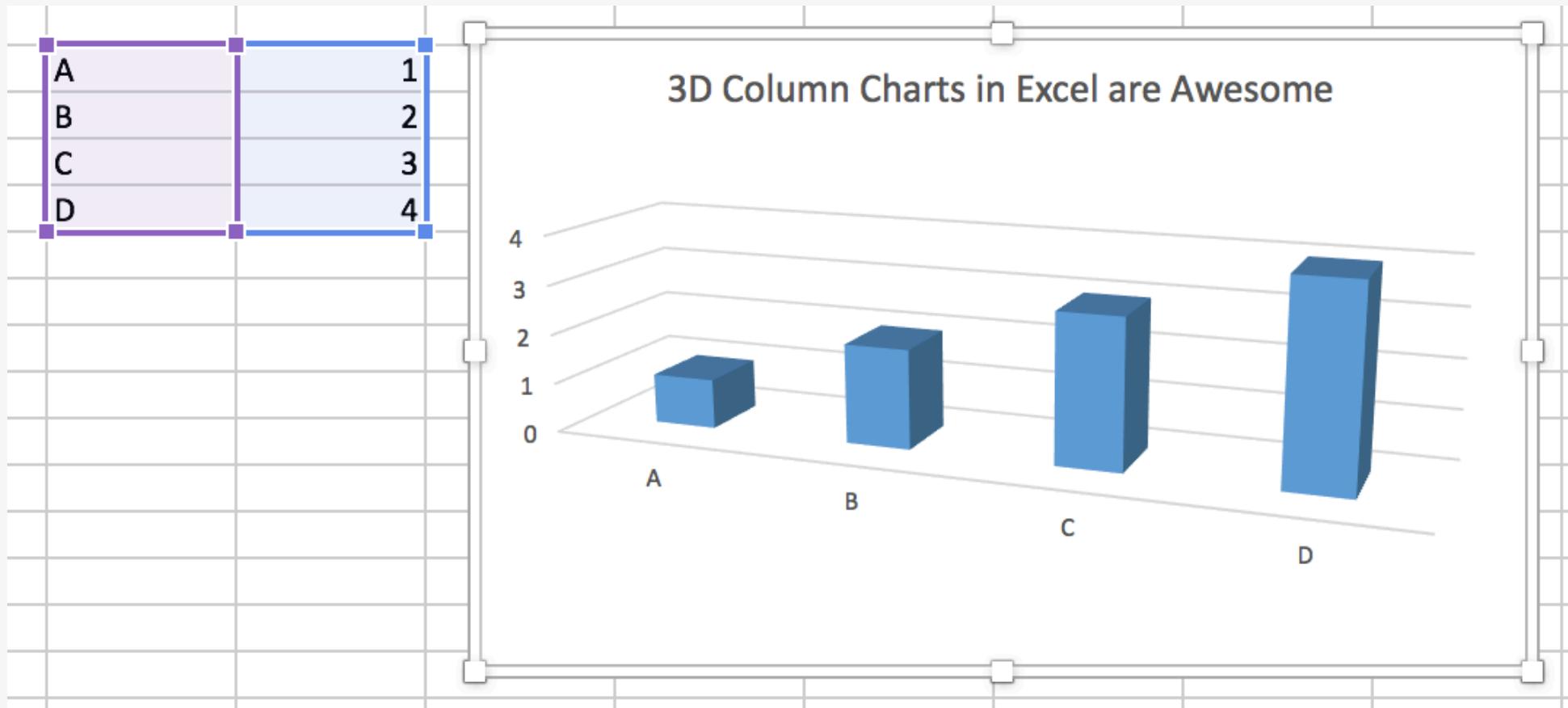
Pew Research

- 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





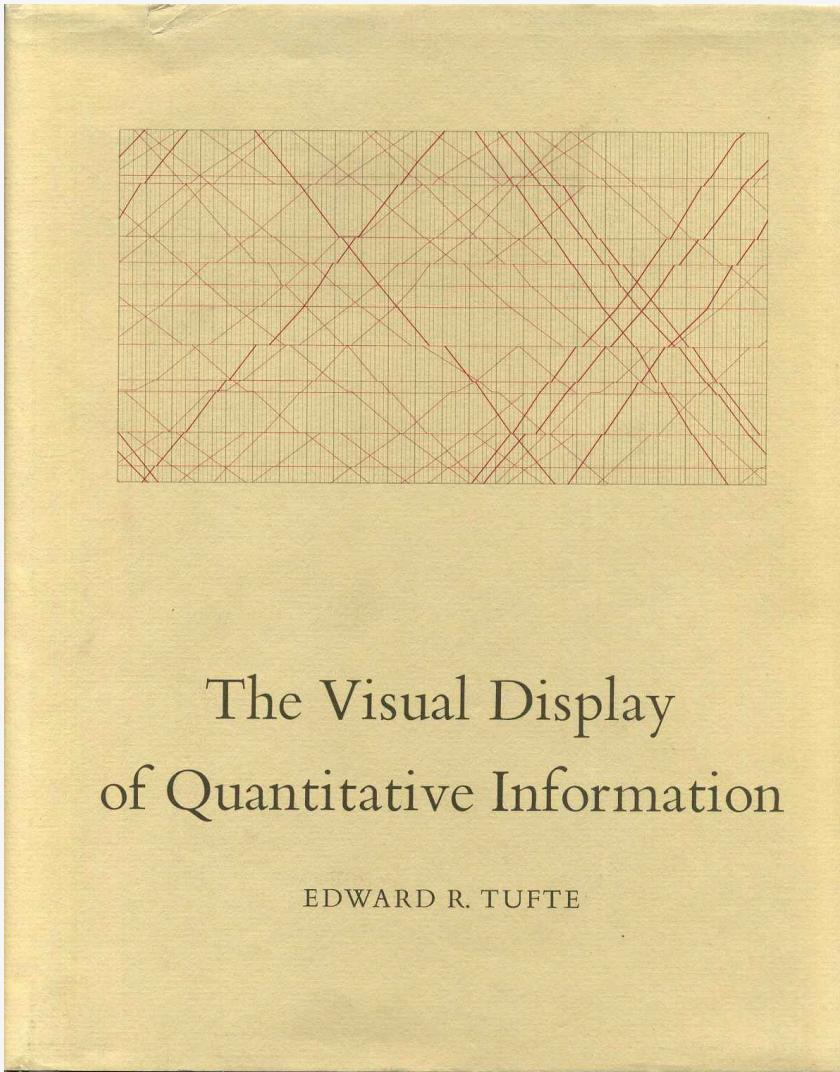
**Bad Taste**

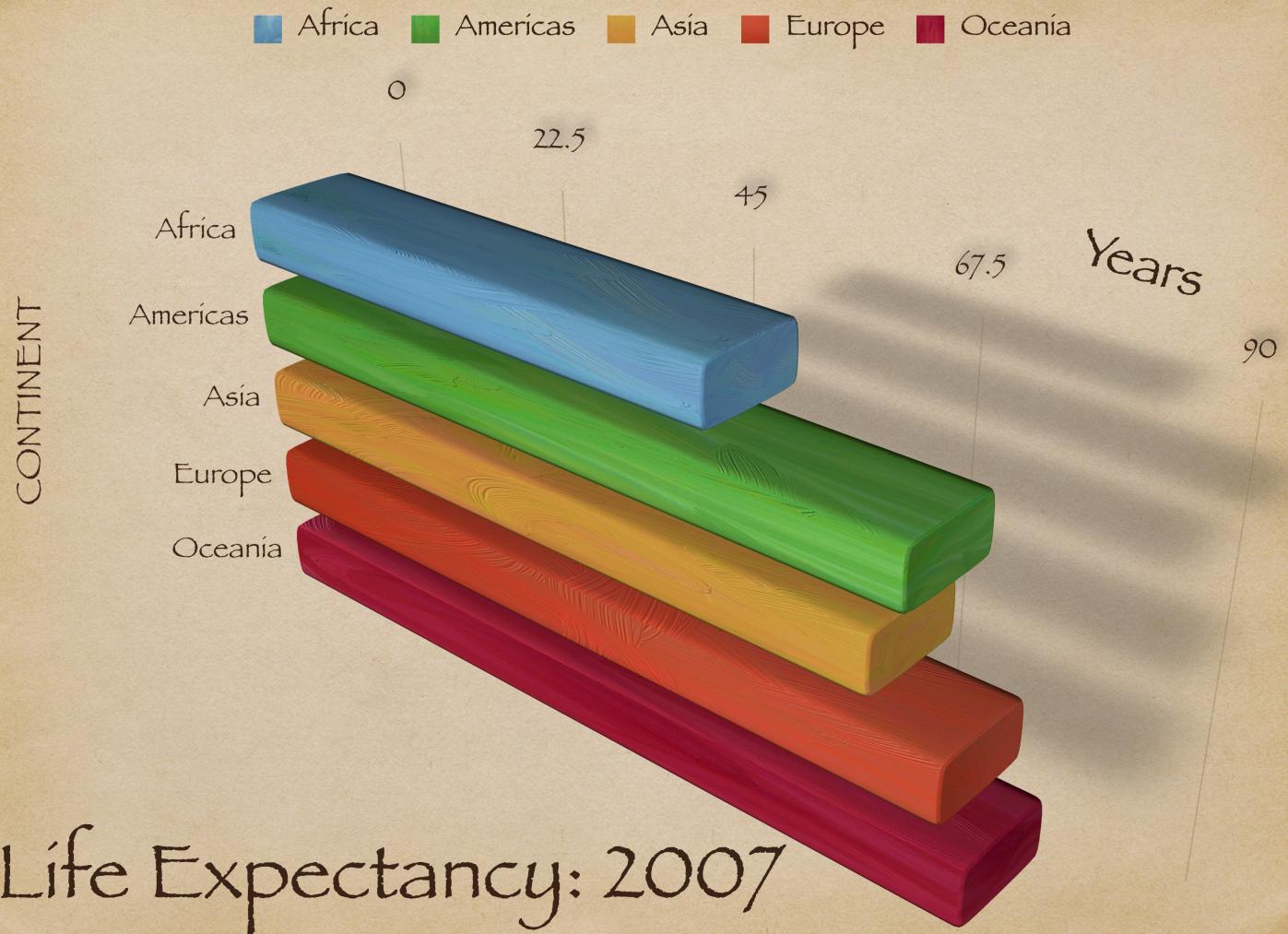
**Bad Data**

**Bad Perception**

# Bad Taste: Simplify, Simplify?

# Tufte's “Data to Ink Ratio”





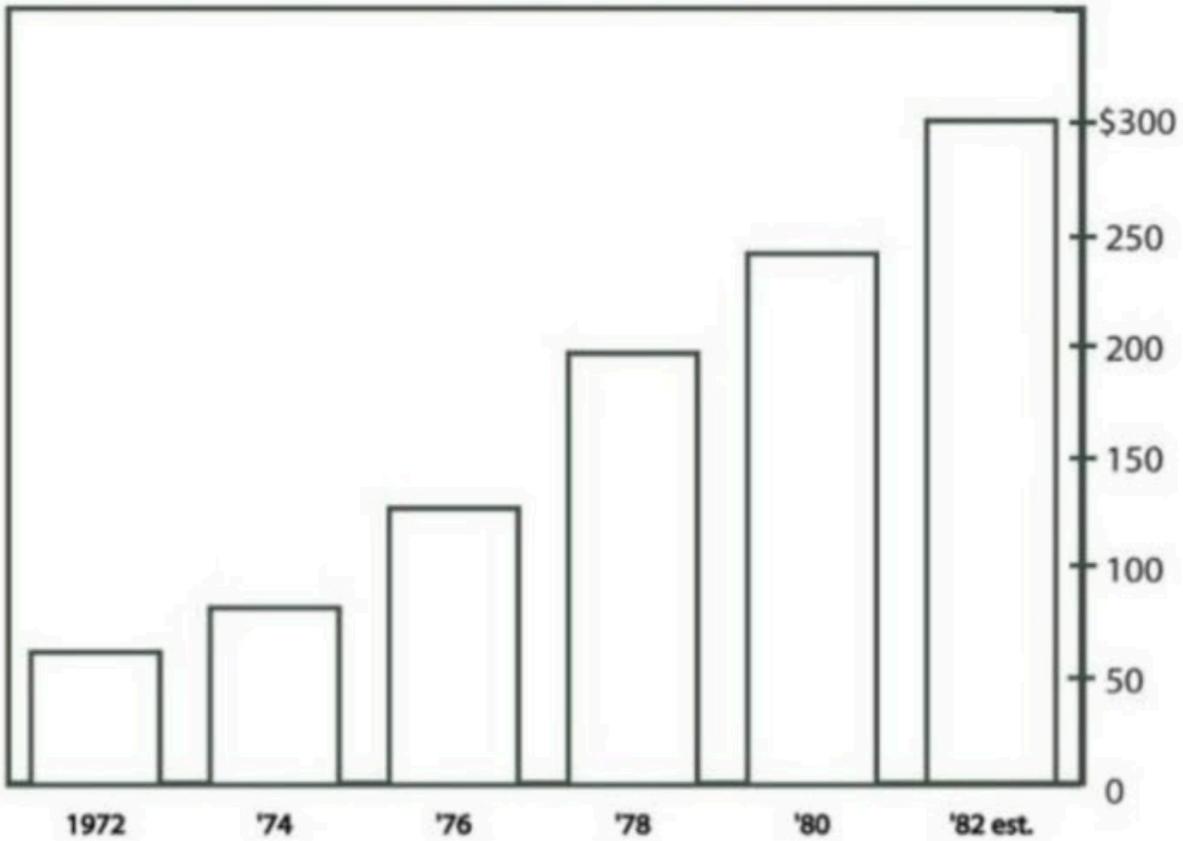
## MONSTROUS COSTS

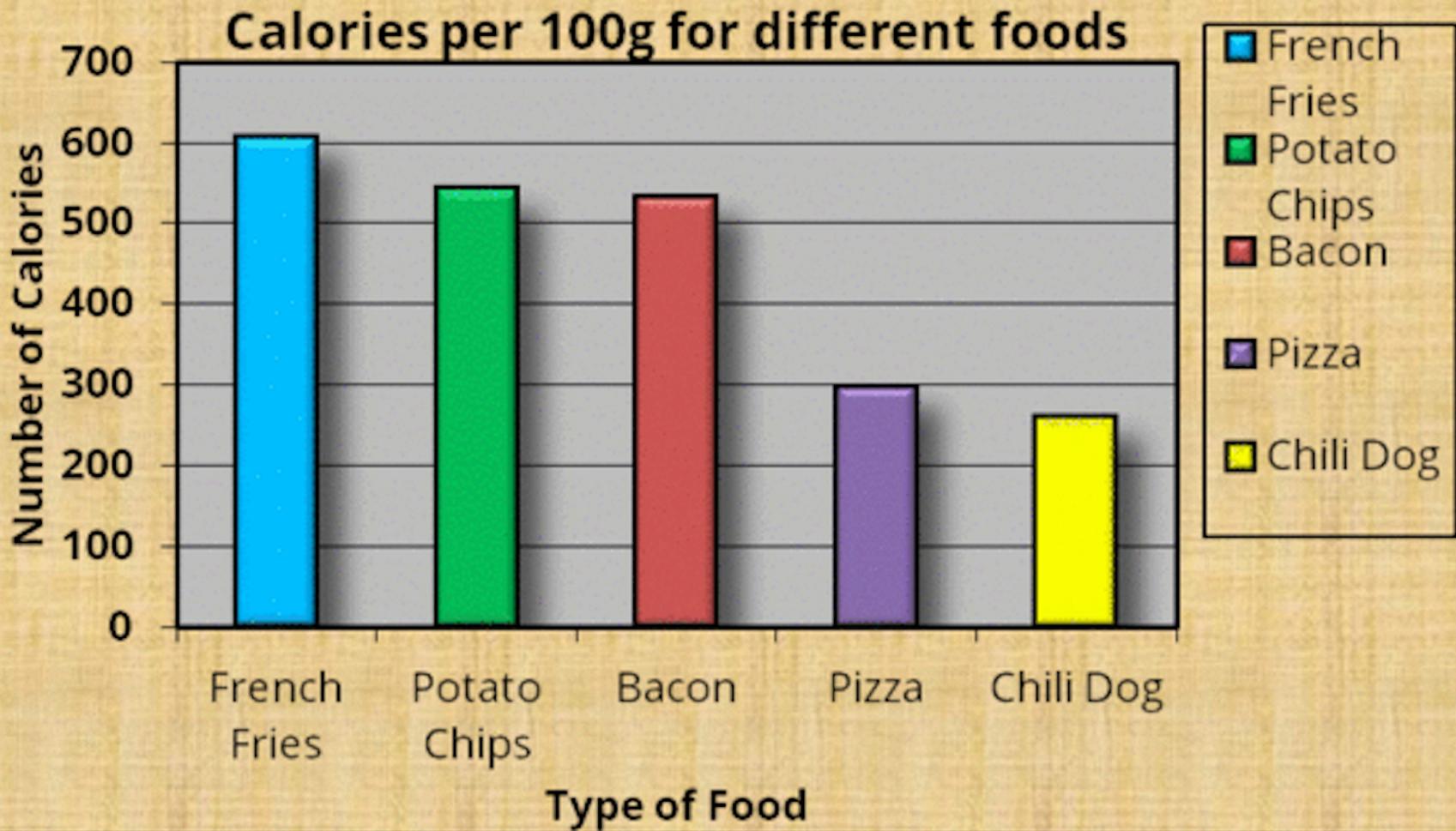
Total House and Senate campaign expenditures,  
in millions



Nigel Holmes

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

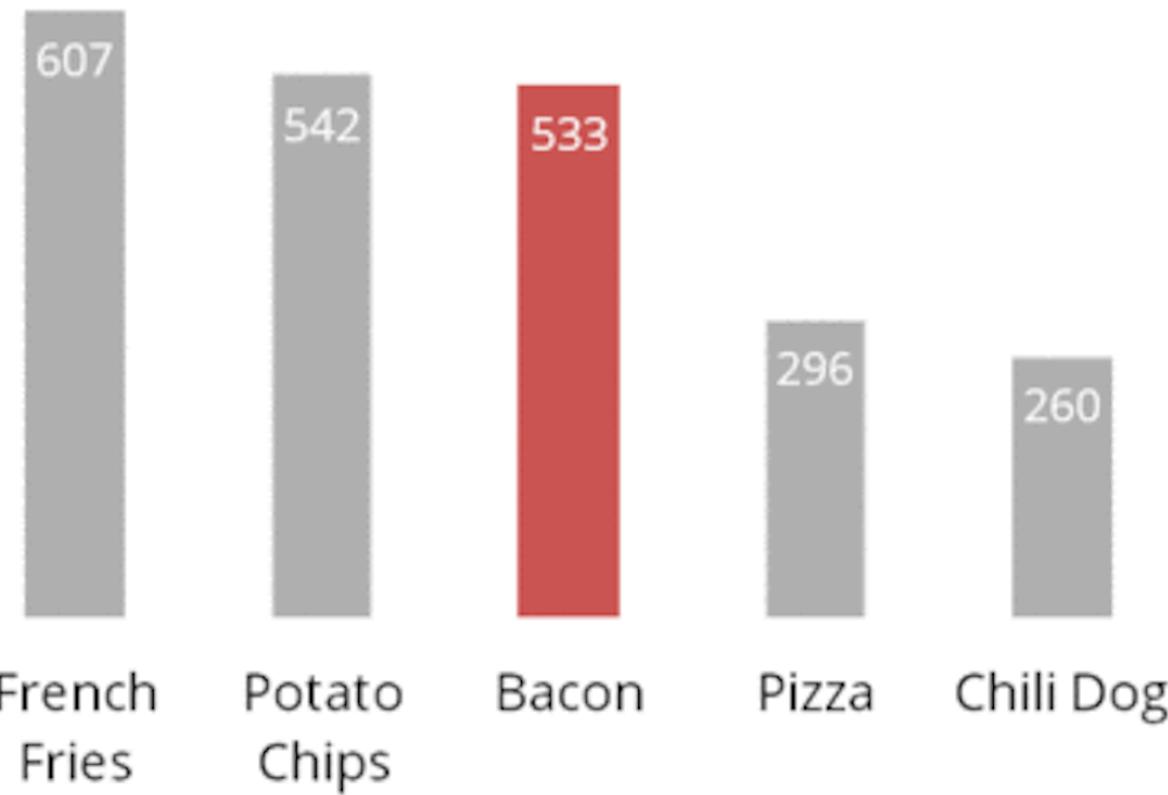




Created by Darkhorse Analytics

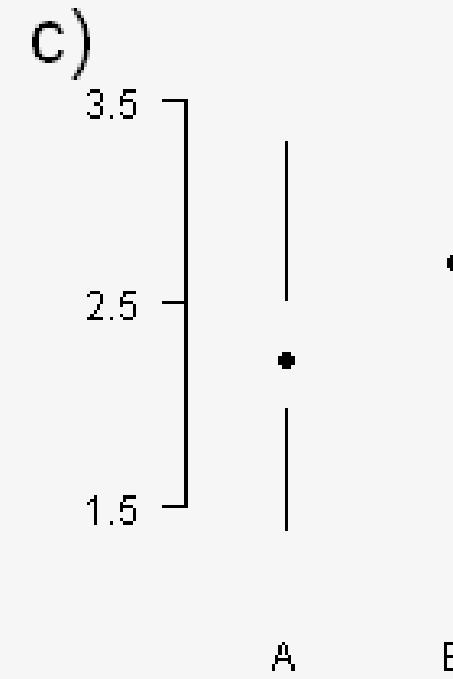
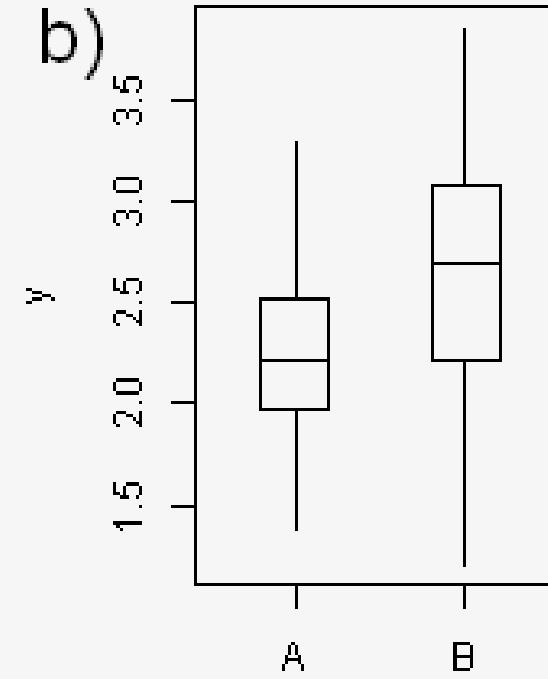
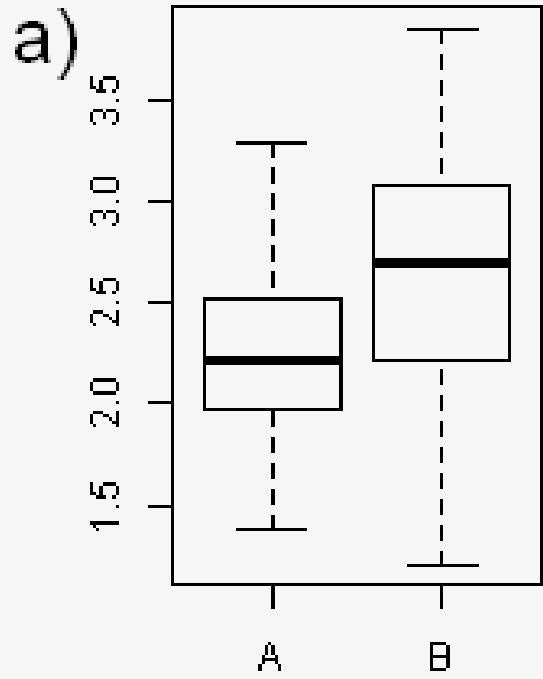
[www.darkhorseanalytics.com](http://www.darkhorseanalytics.com)

## Calories per 100g



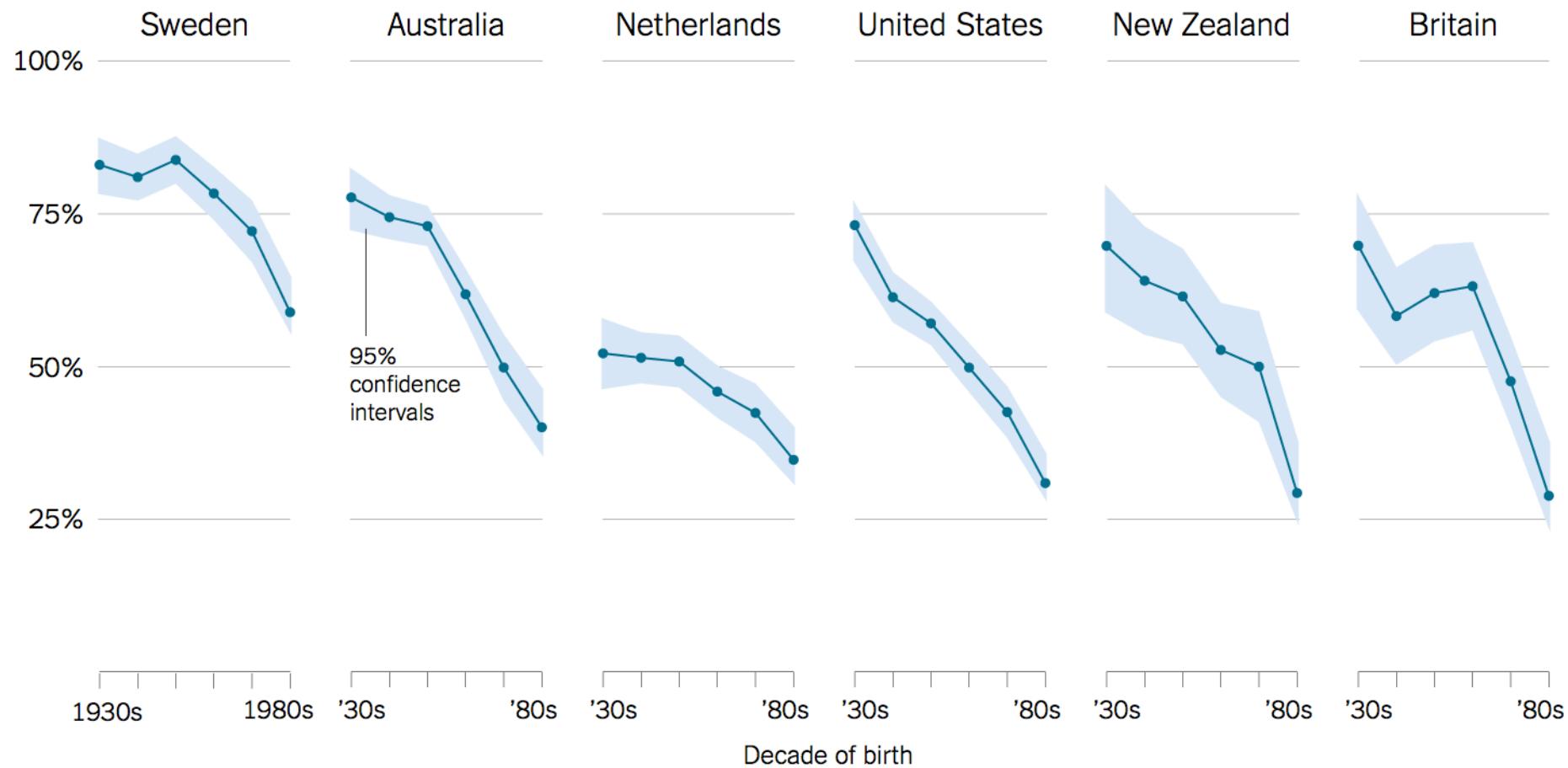
Created by Darkhorse Analytics

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

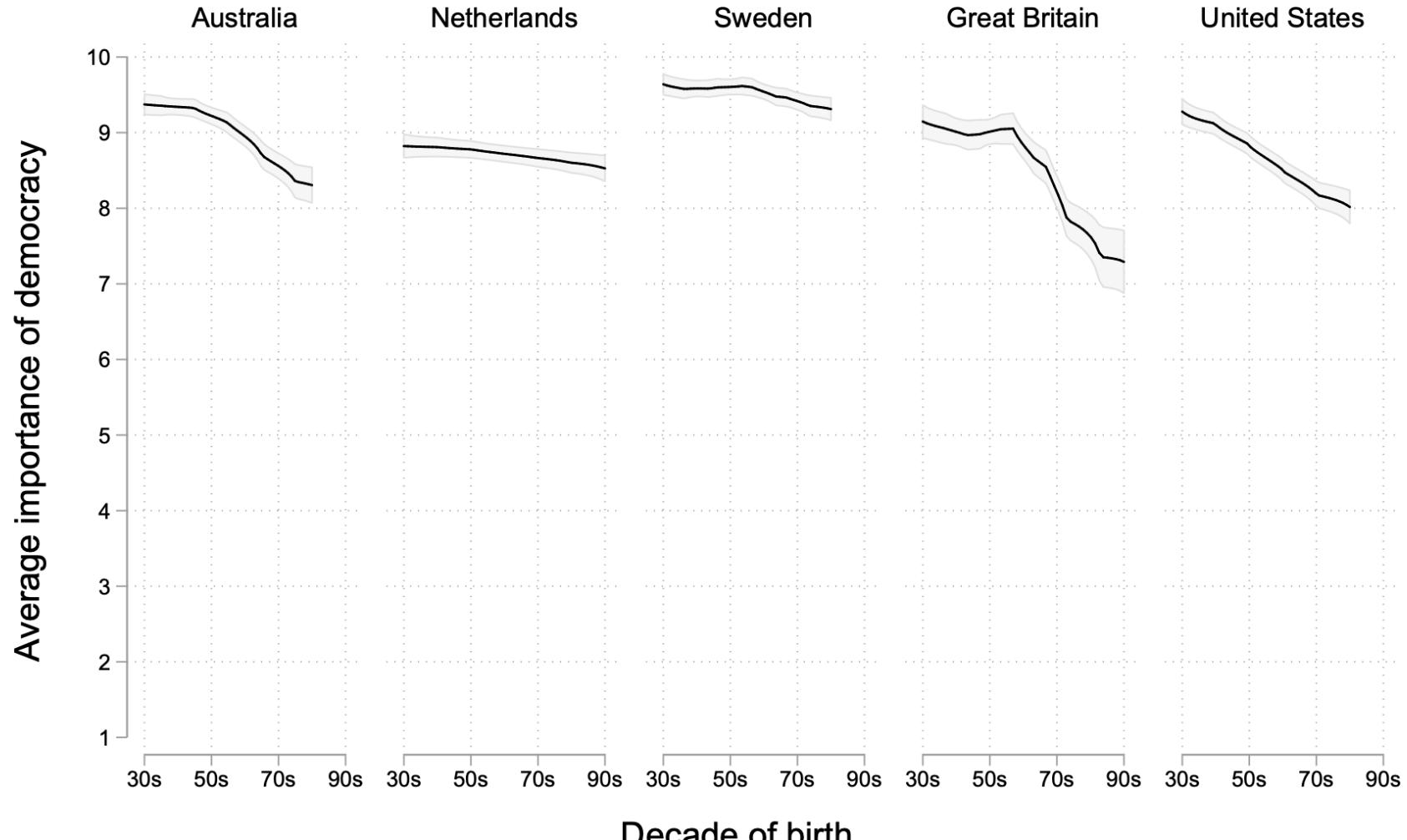


# Bad Data: Junk-Free Junk Charts

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



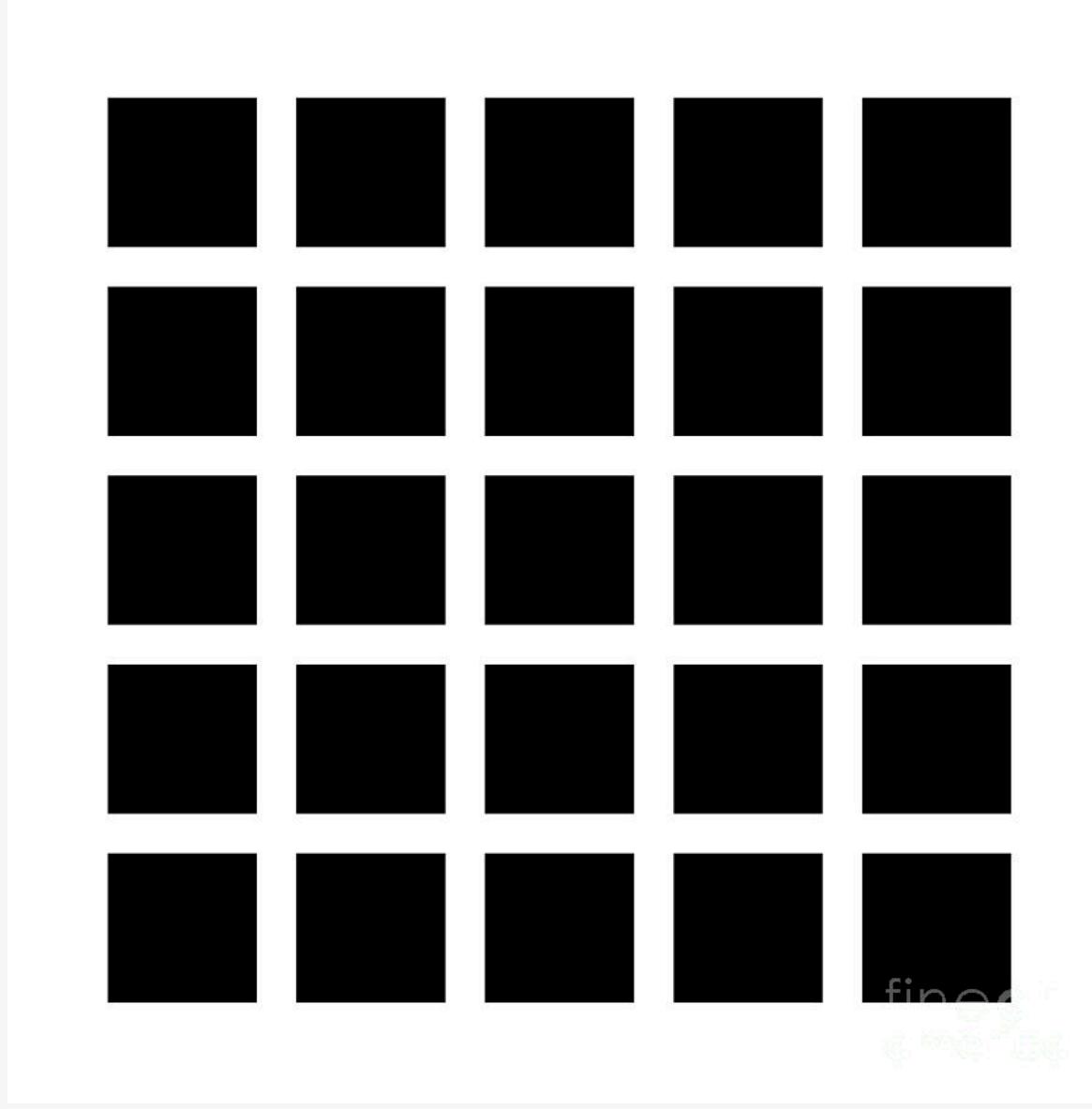
Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times



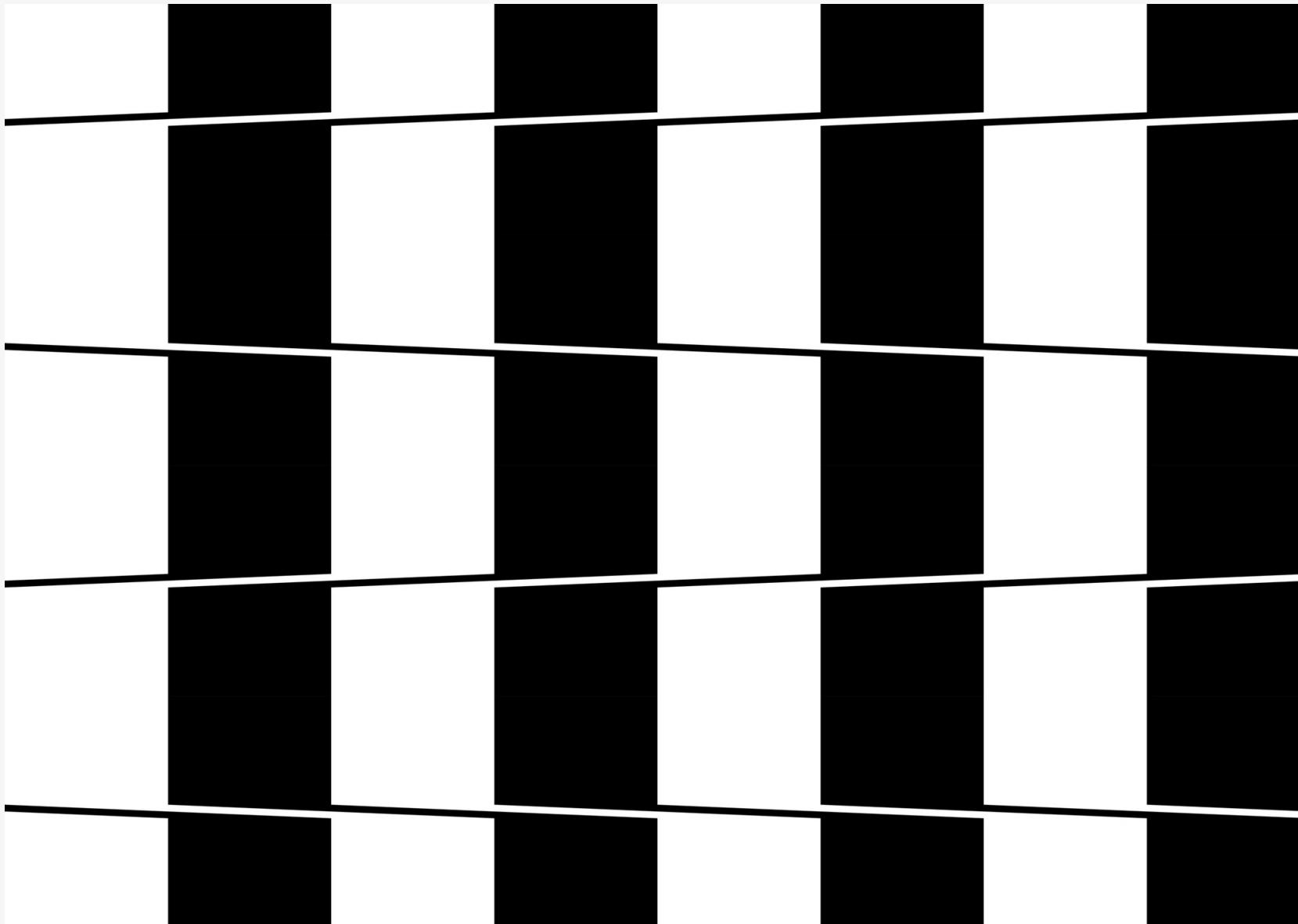
Graph by Erik Voeten, based on WVS 5

# **Bad Perception: Seeing and Not Seeing**

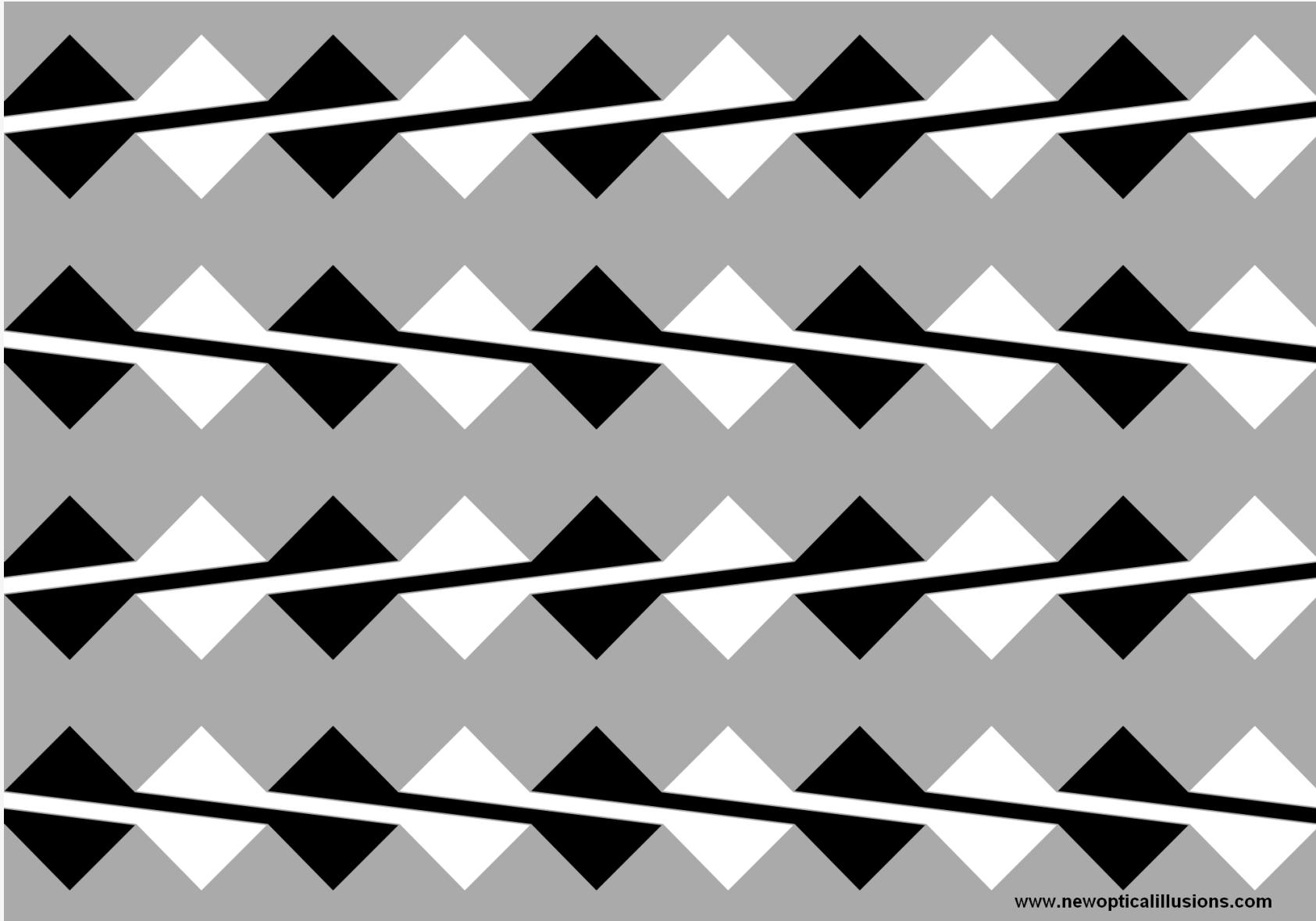
# Edges & Contrasts



Hermann Grid Effect



Fraser Columns



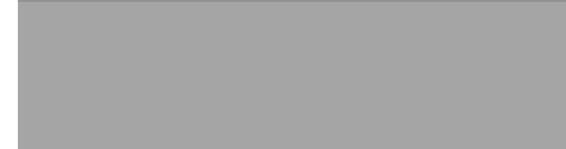
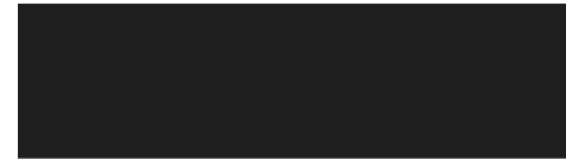
Fraser Diamonds



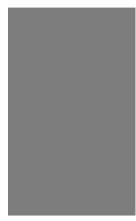
These are two perfectly  
geometrical circles

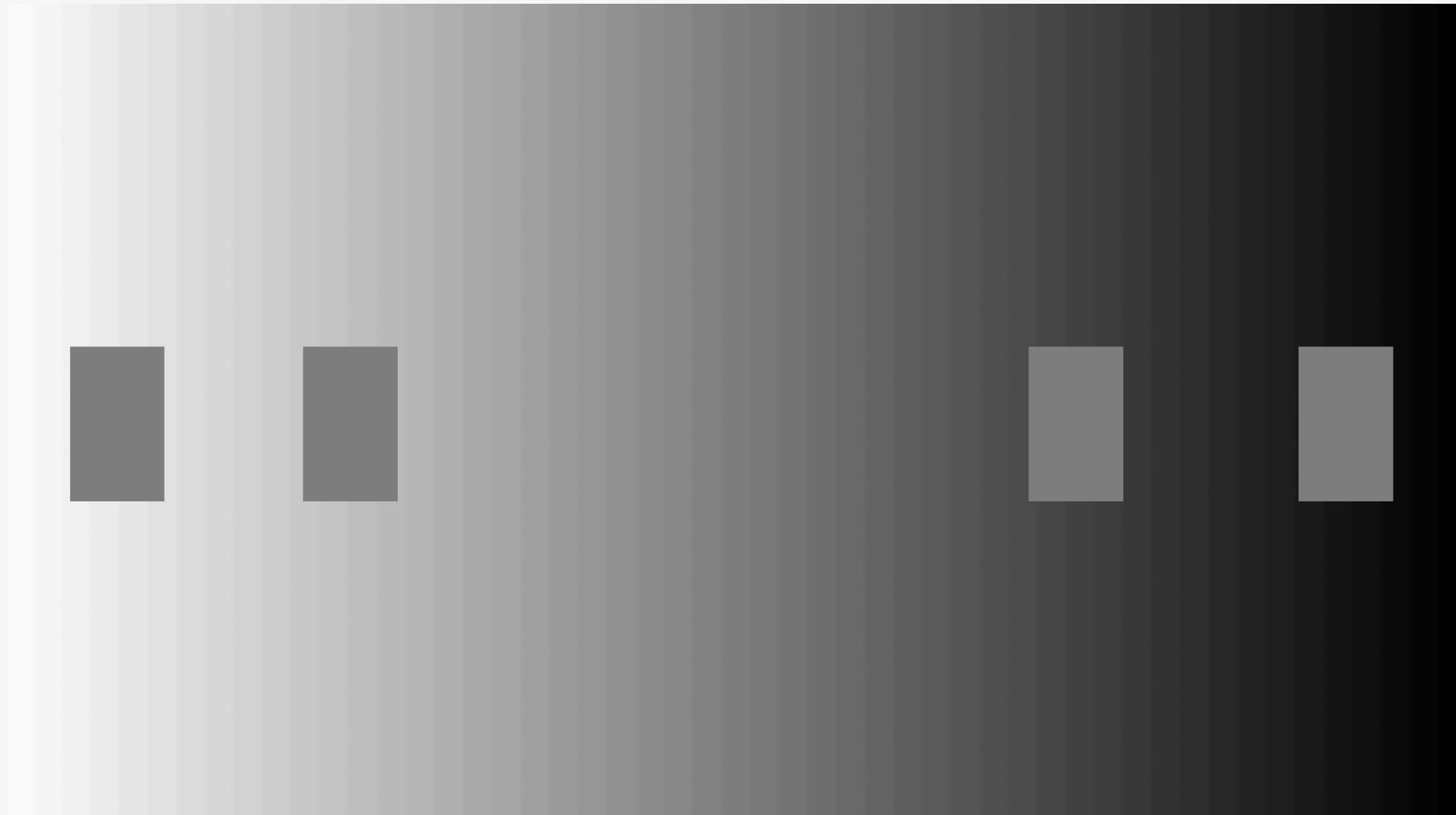


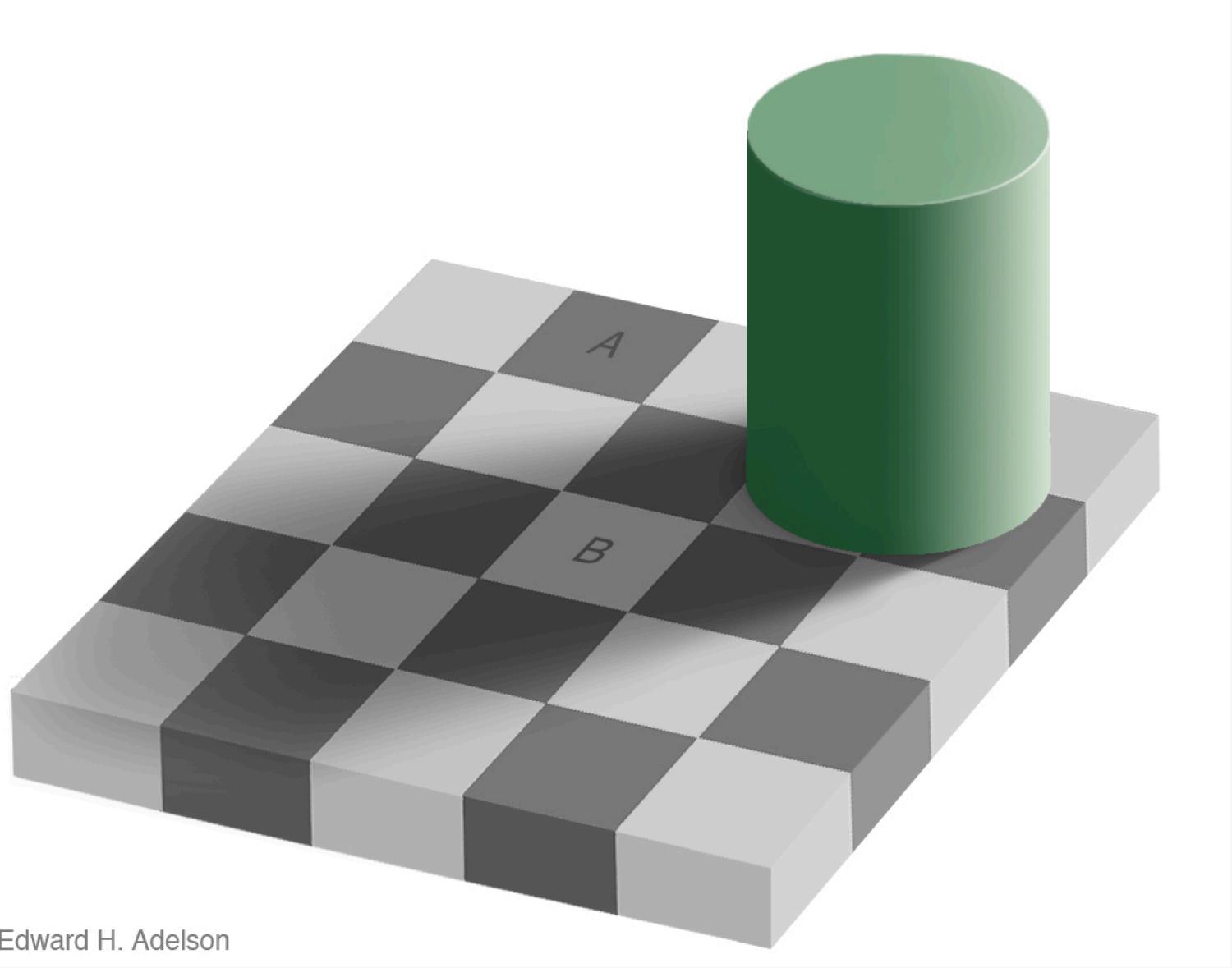
Mach Bands



Mach Bands

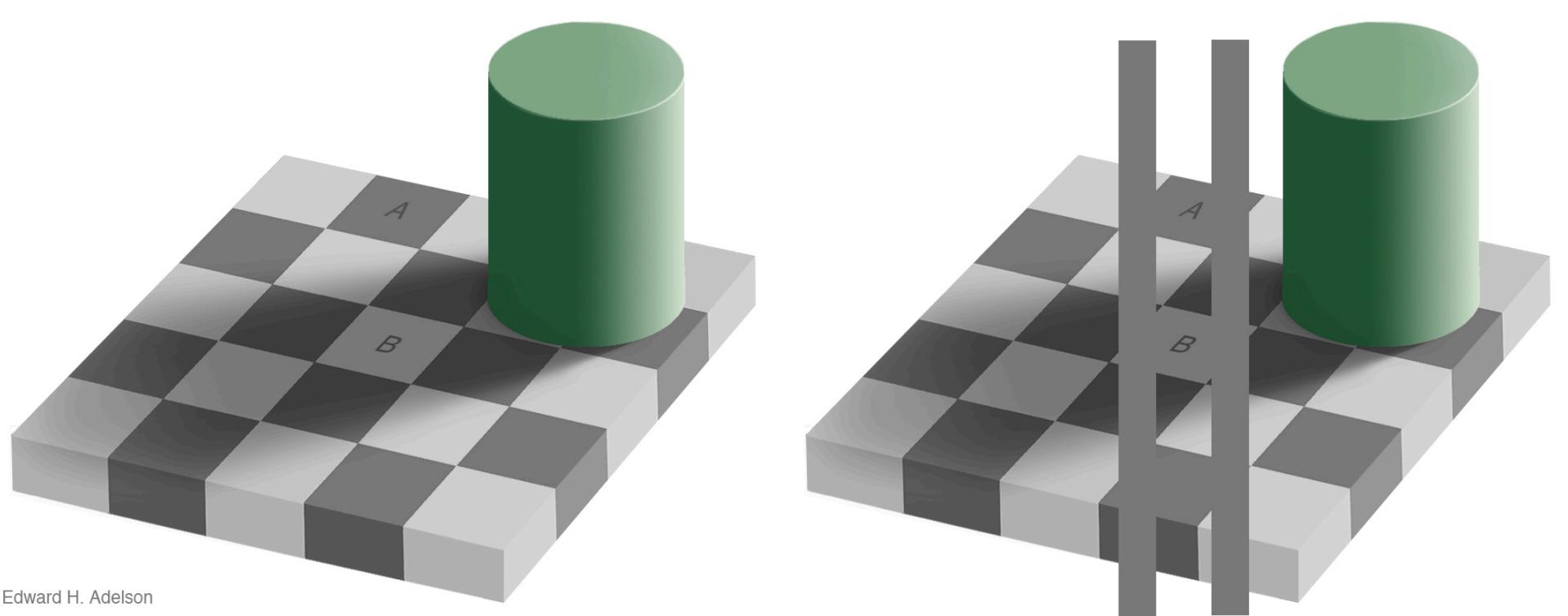






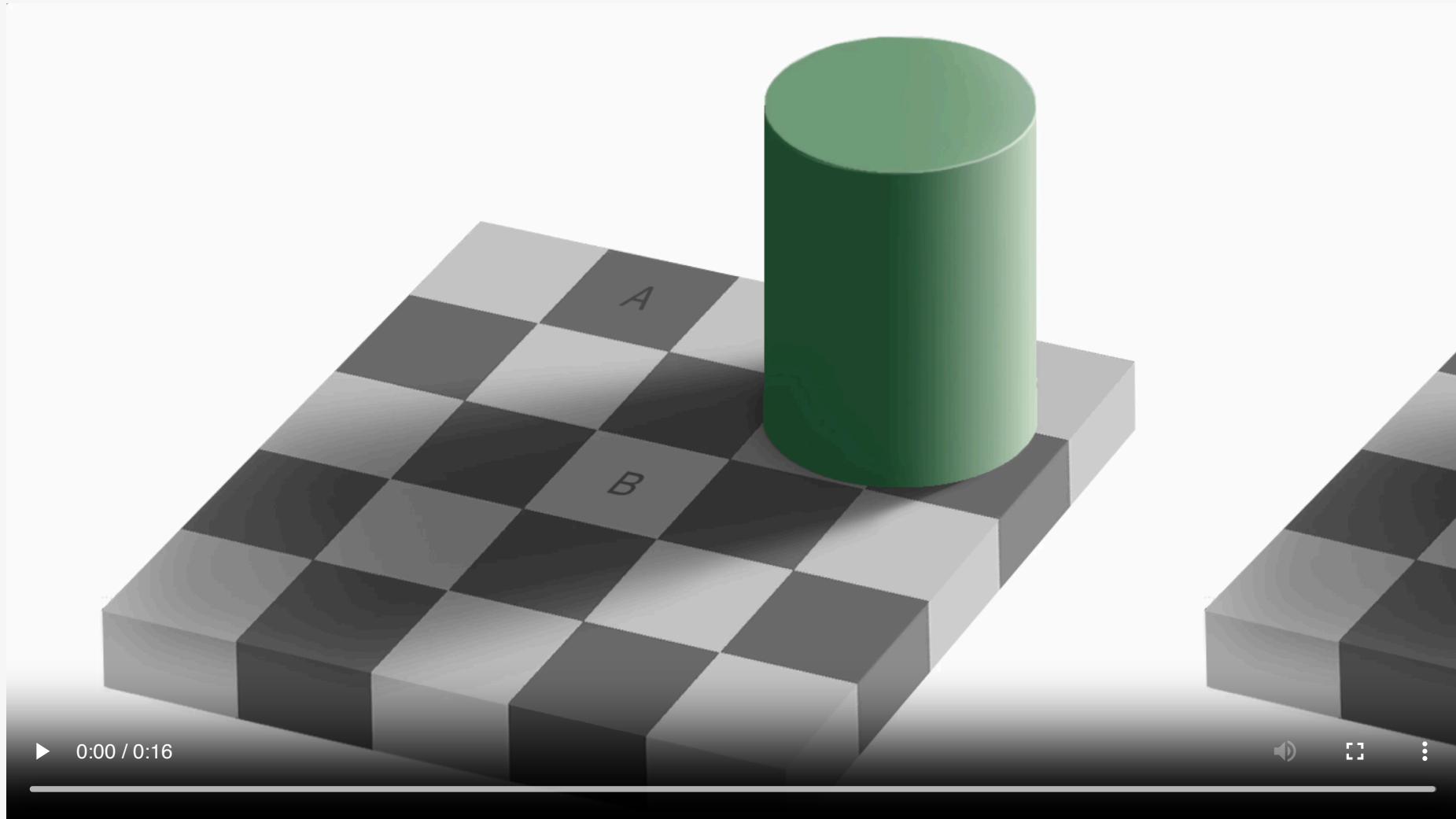
Edward H. Adelson

Edward Adelson



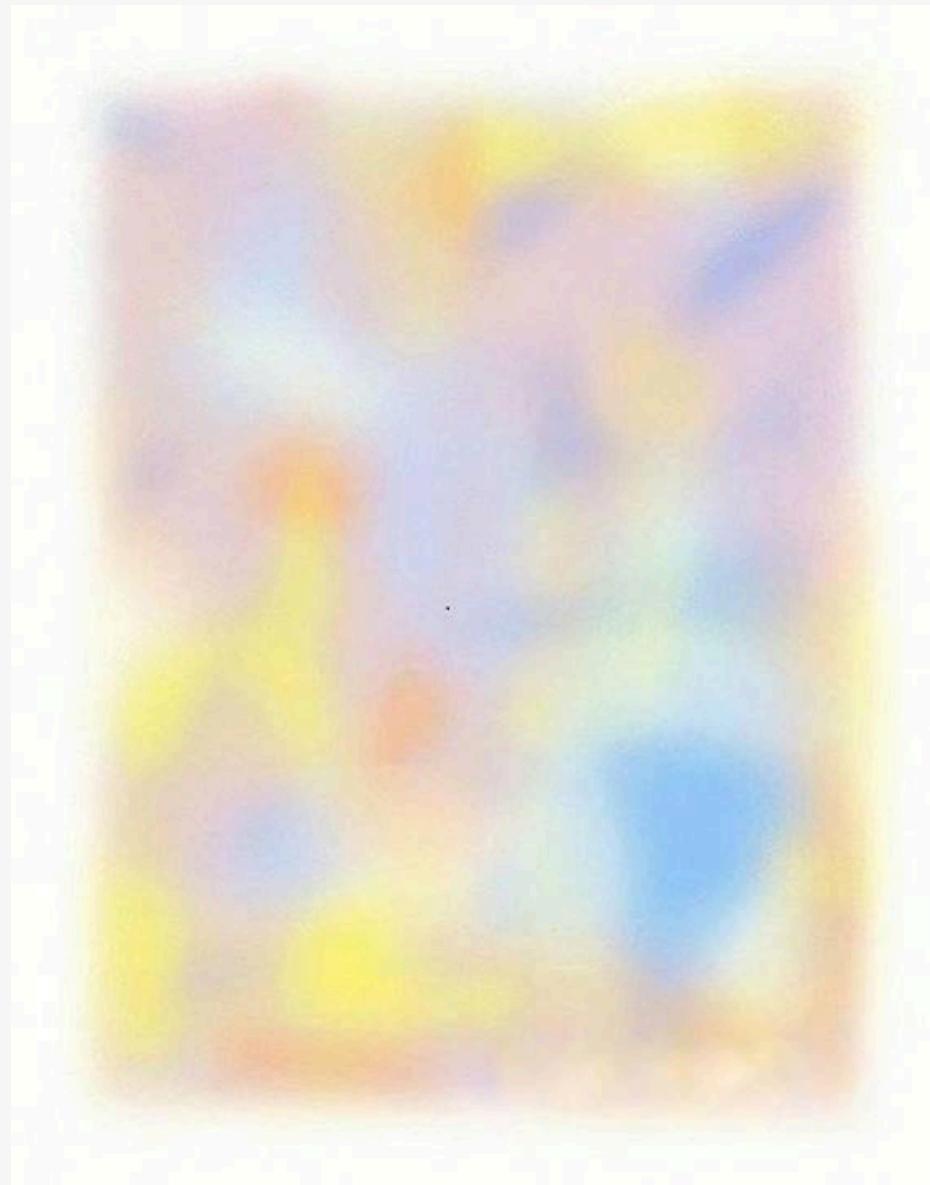
Edward H. Adelson

Edward Adelson

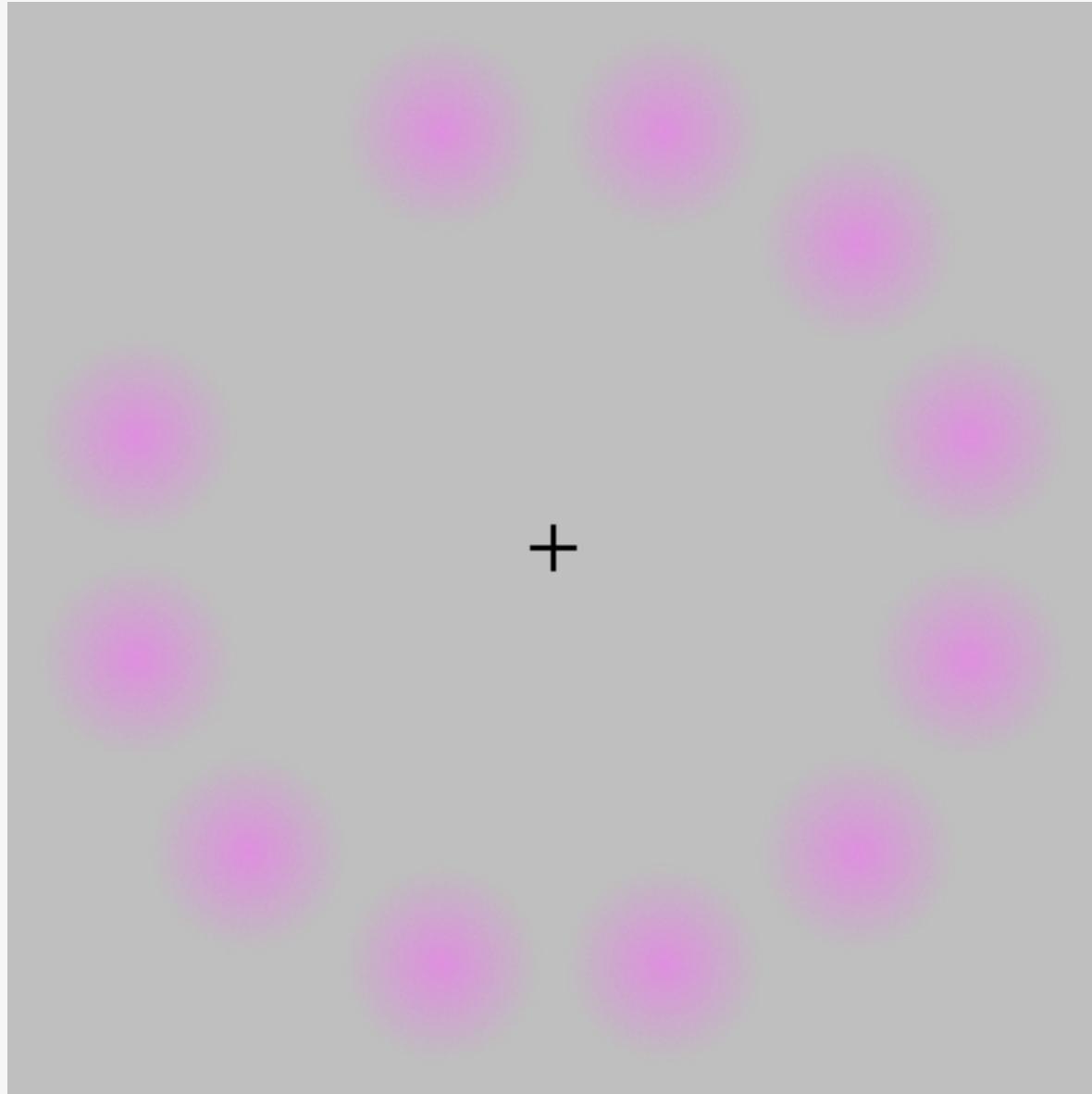


Edward Adelson

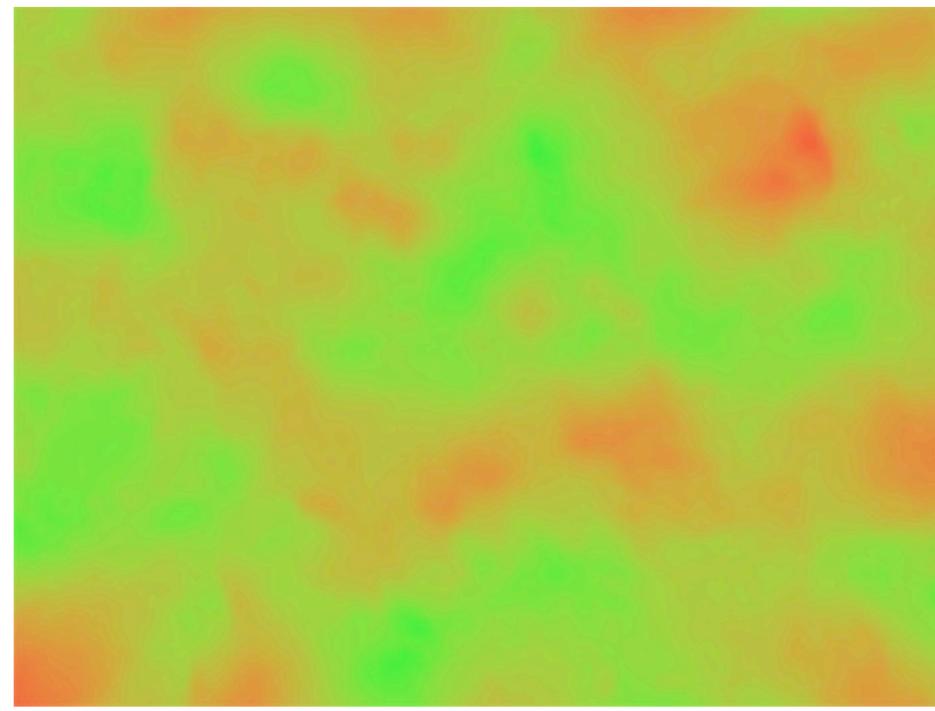
# Luminance and Color



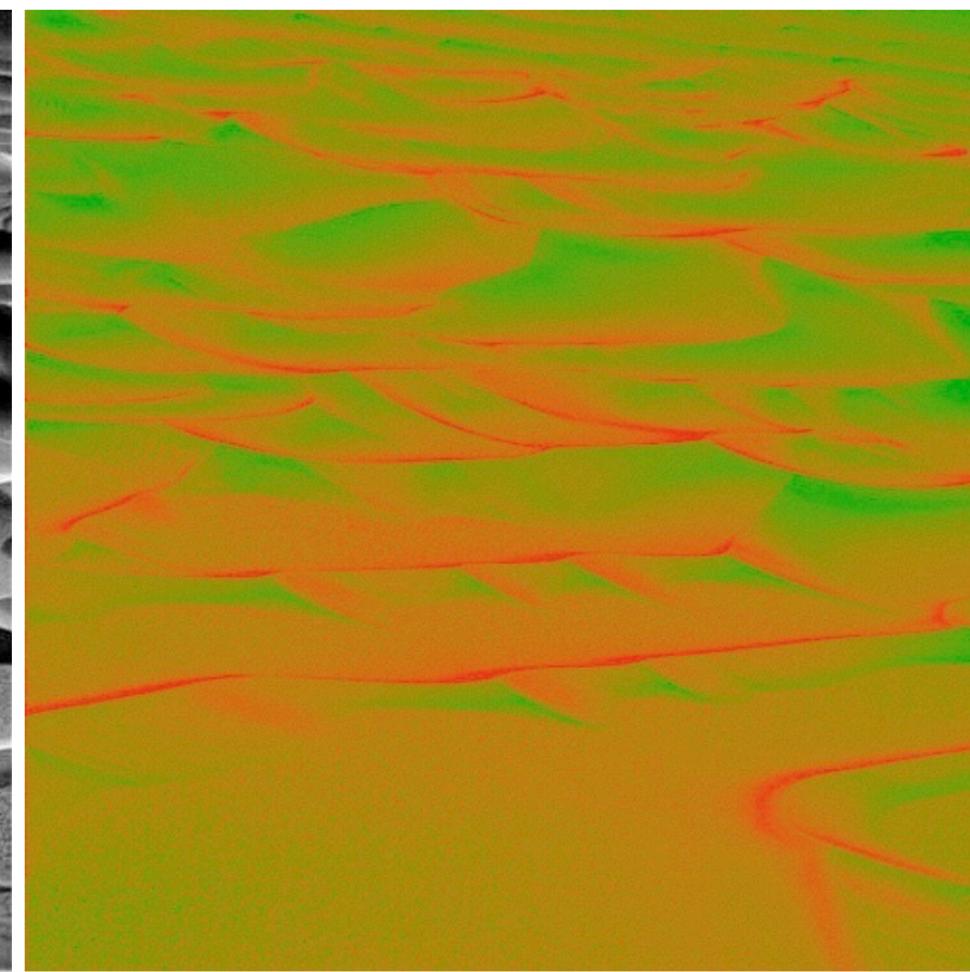
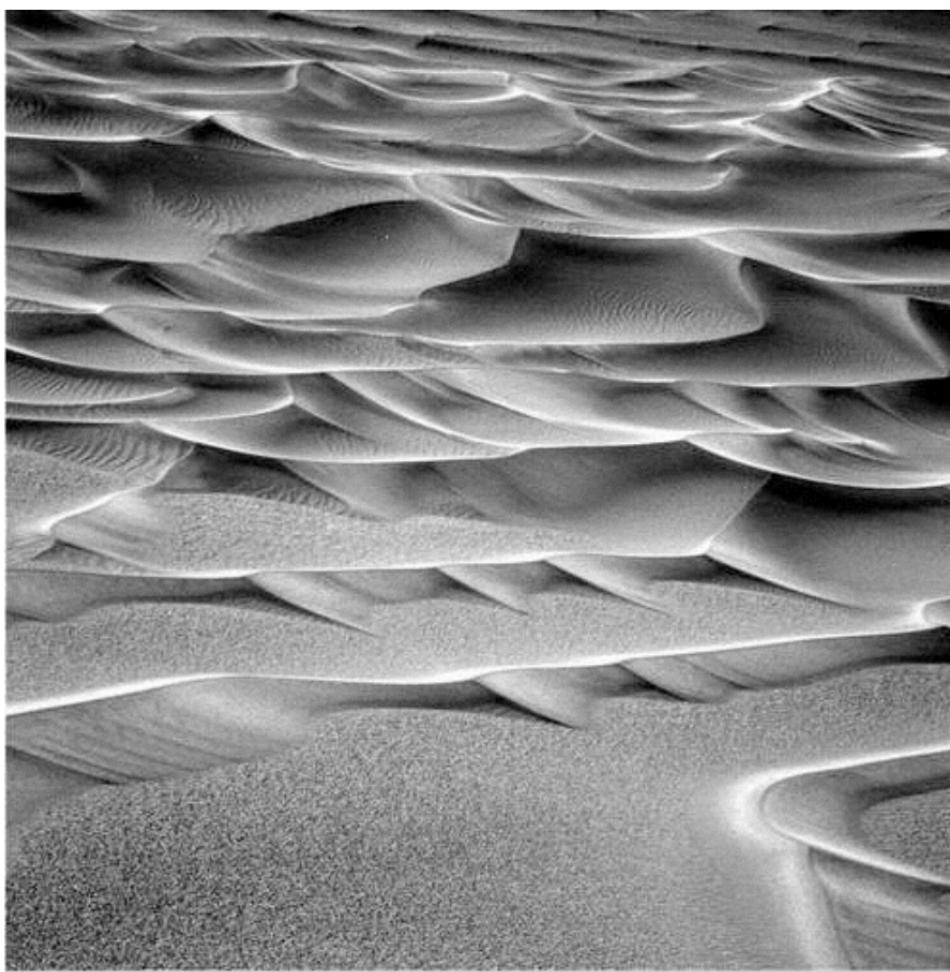
Troxler effect



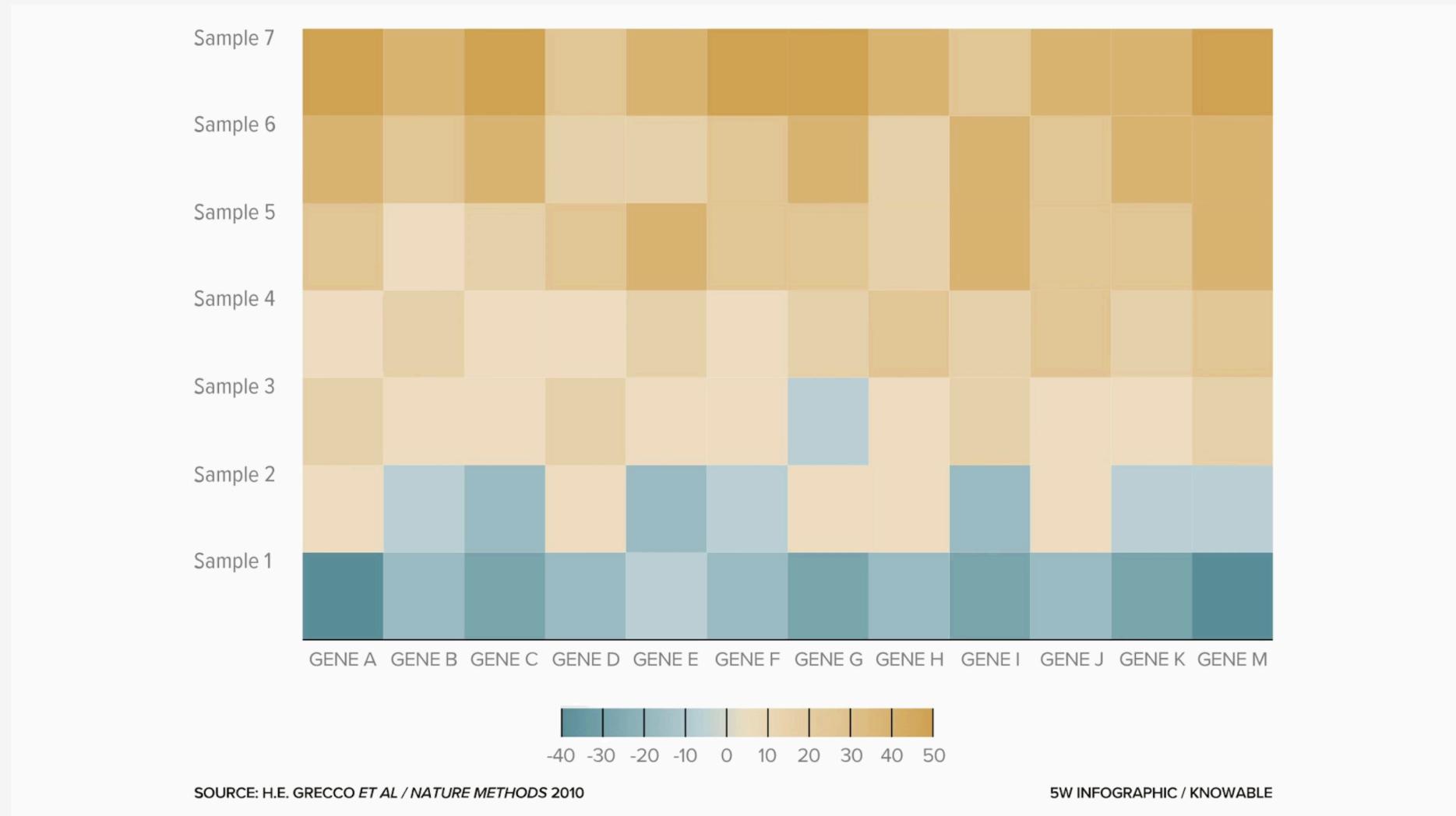
Lilac Chaser

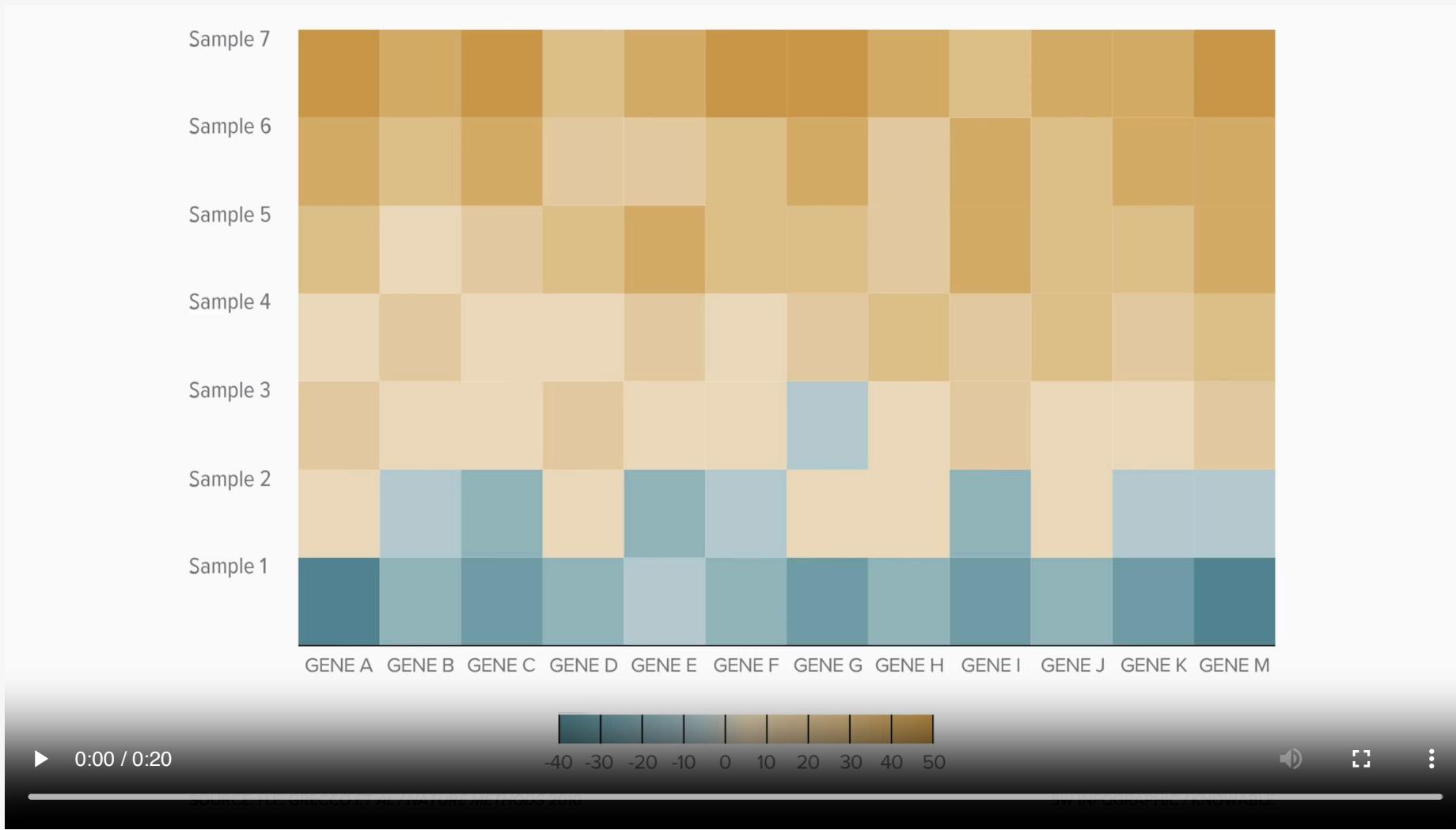


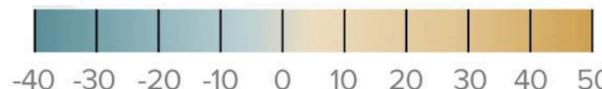
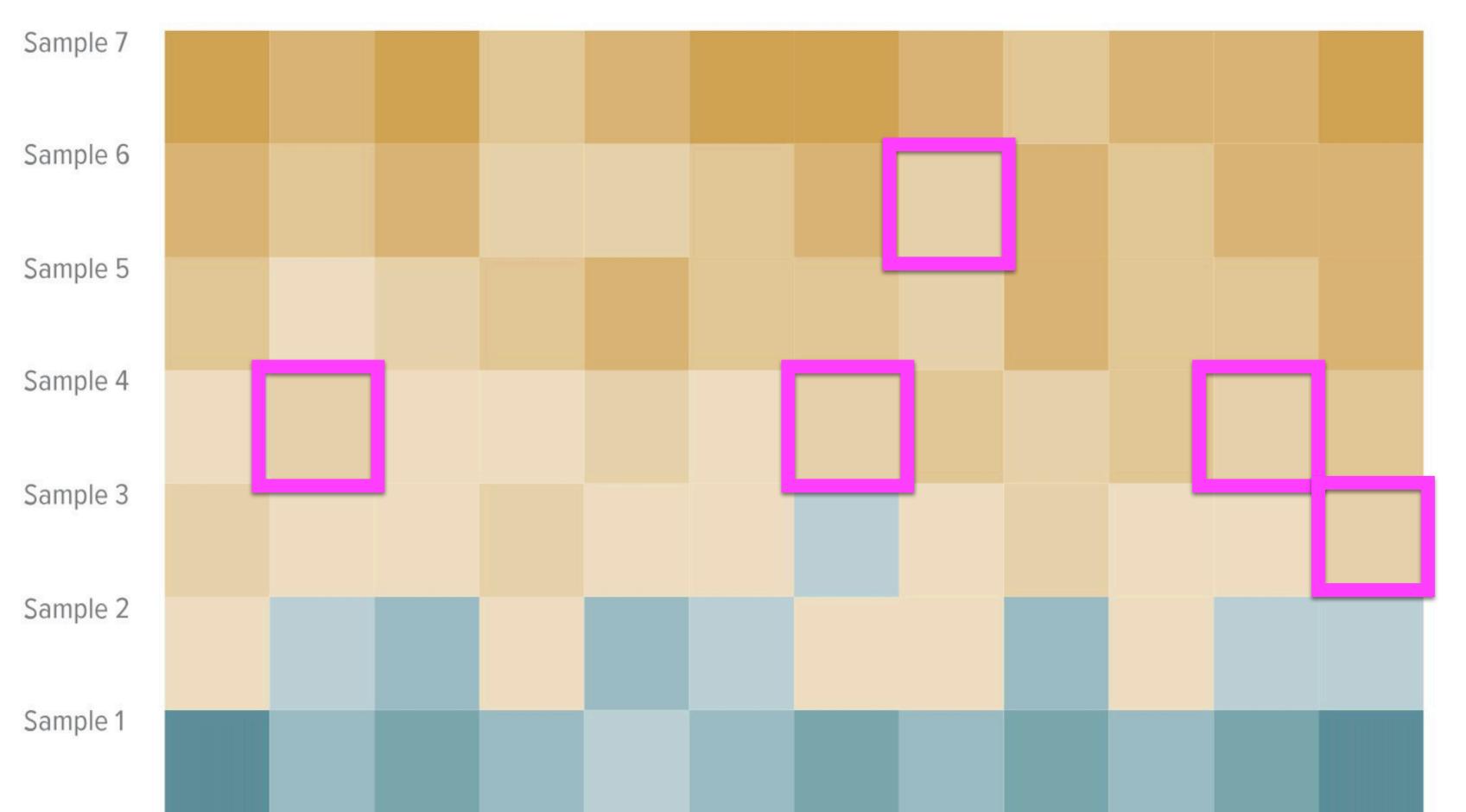
Colin Ware



Colin Ware

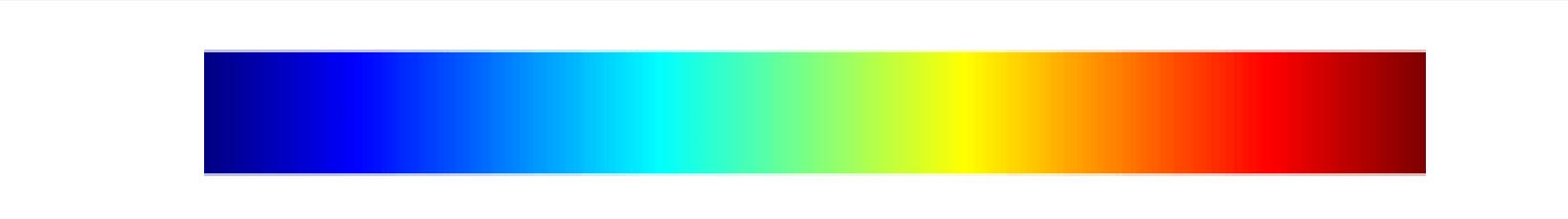


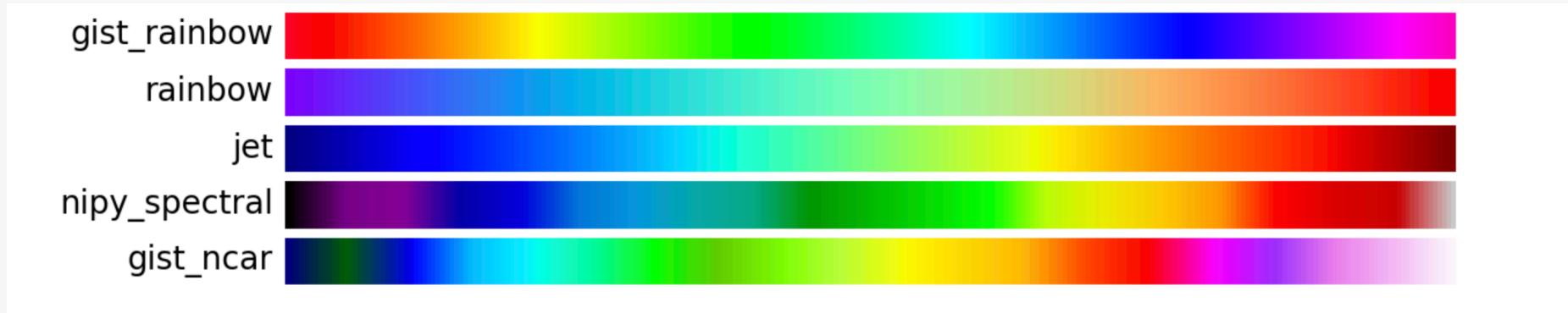


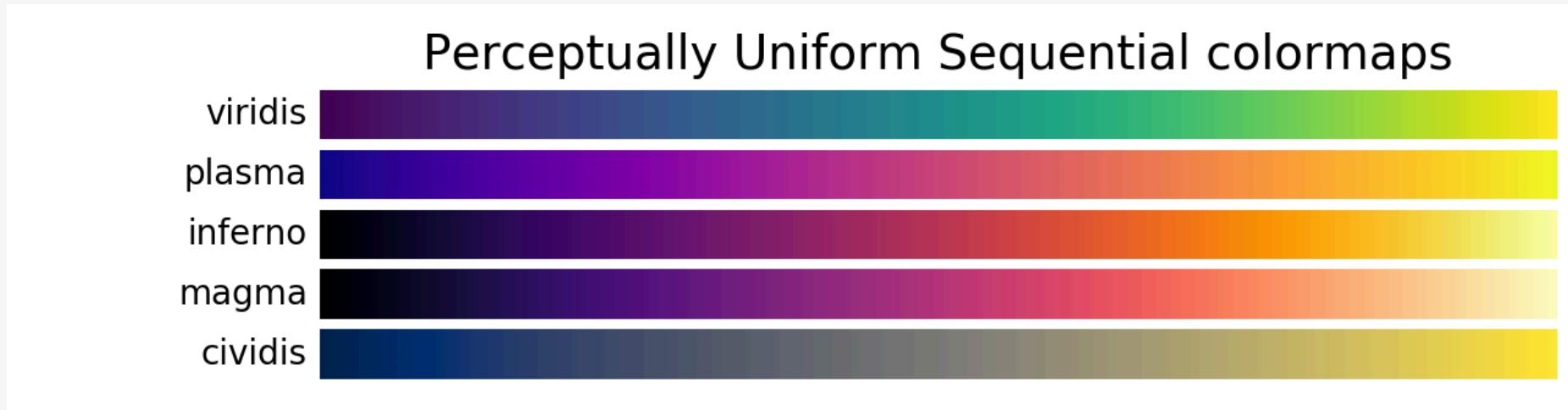
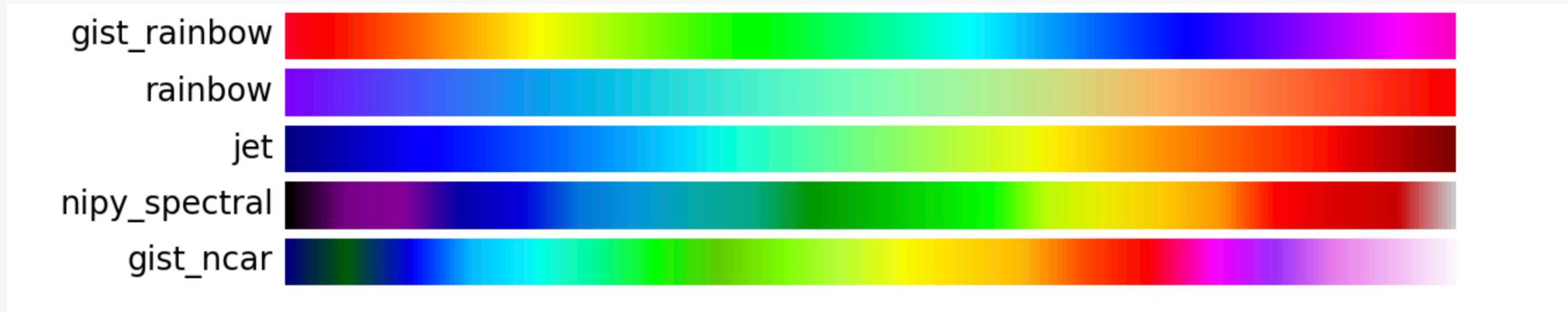


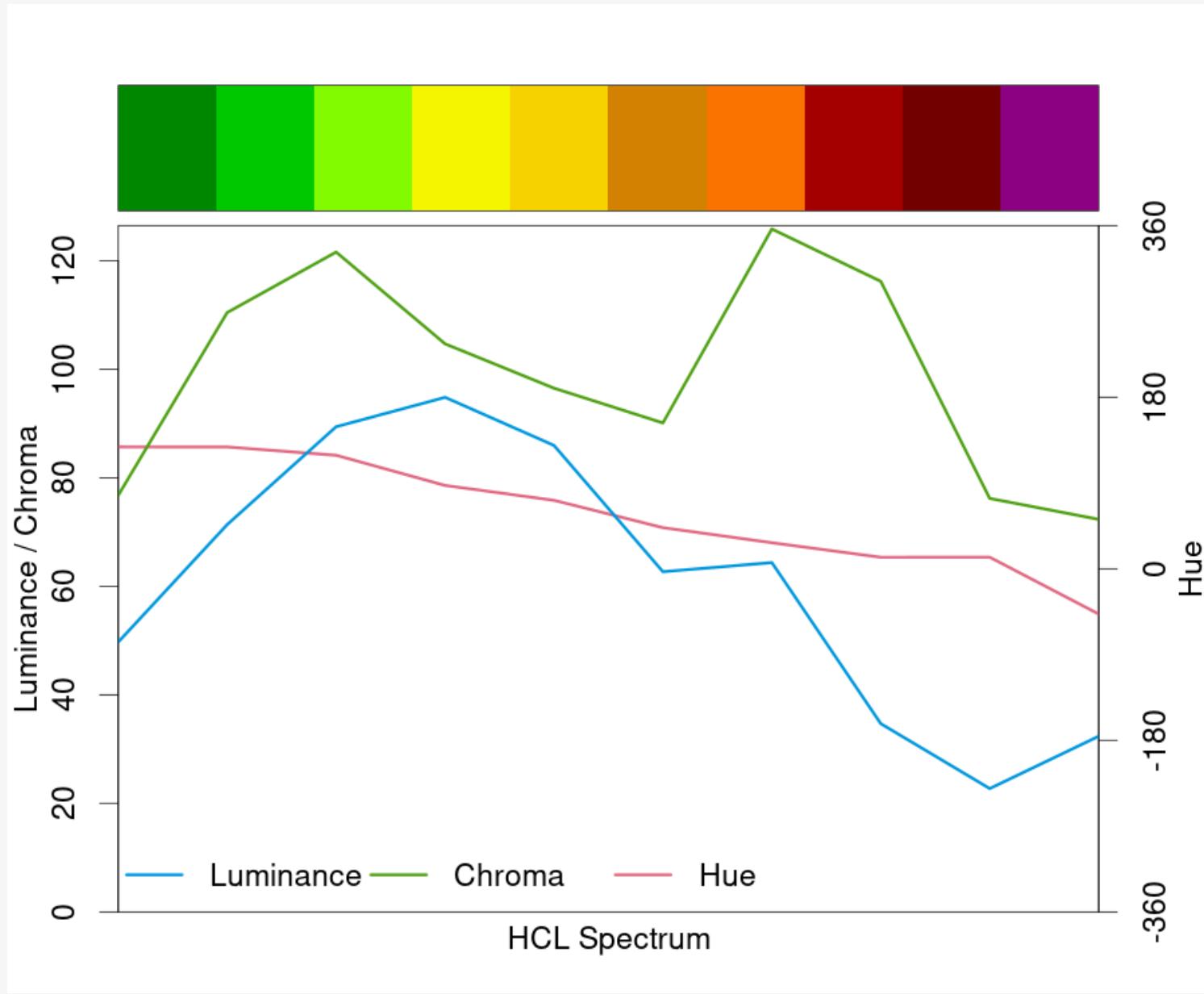
SOURCE: H.E. GRECCO ET AL / NATURE METHODS 2010

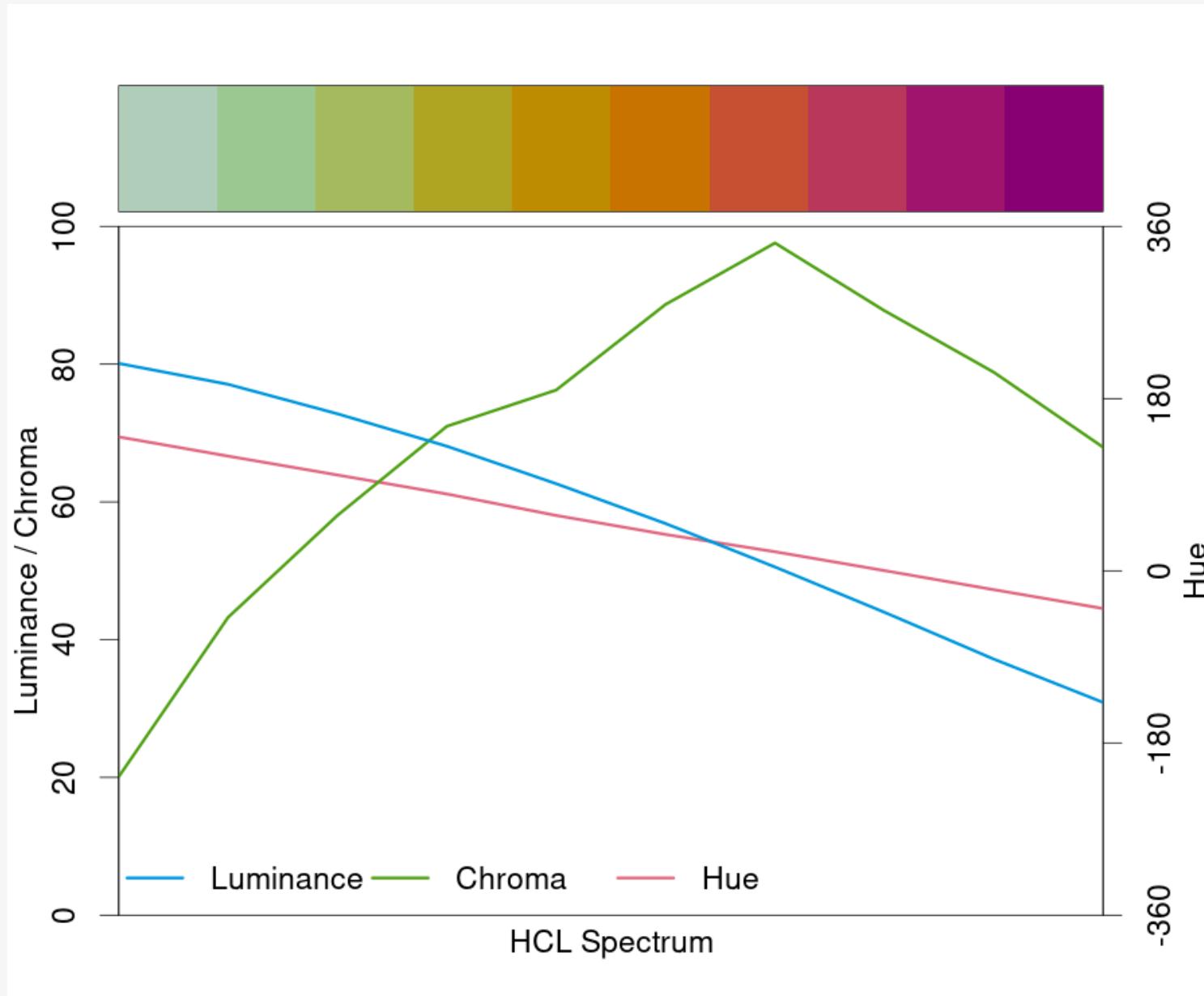
5W INFOGRAPHIC / KNOWABLE





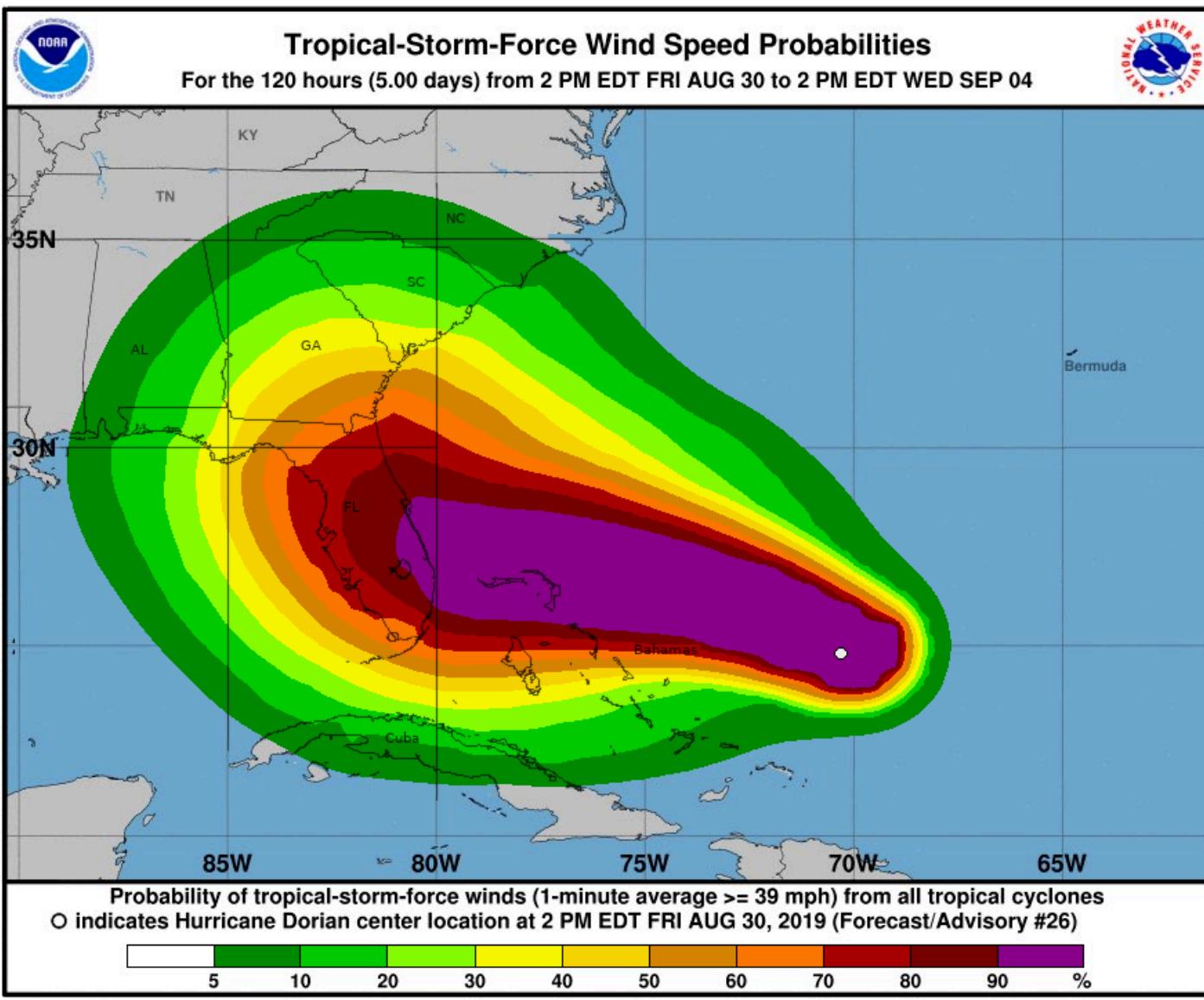


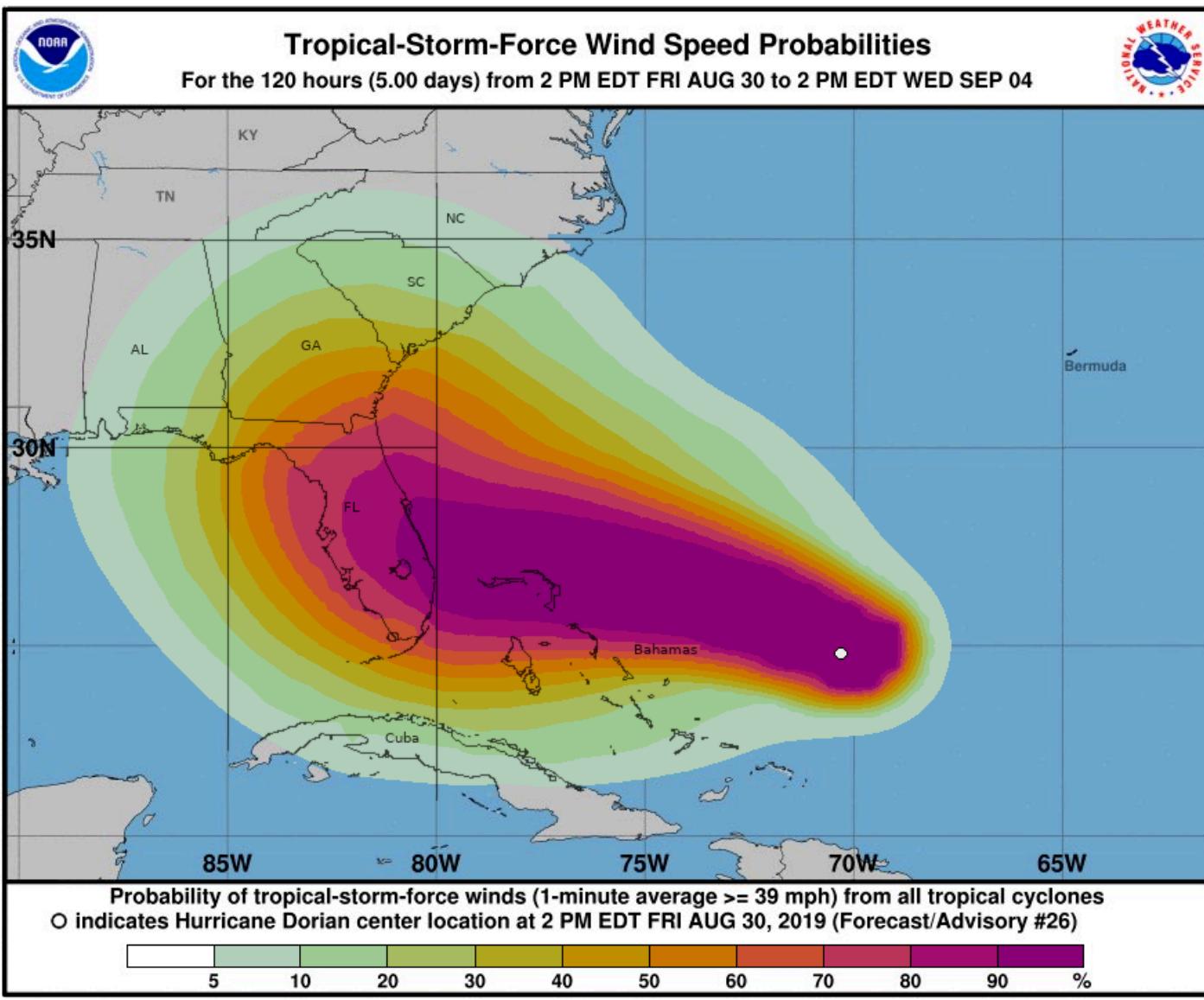




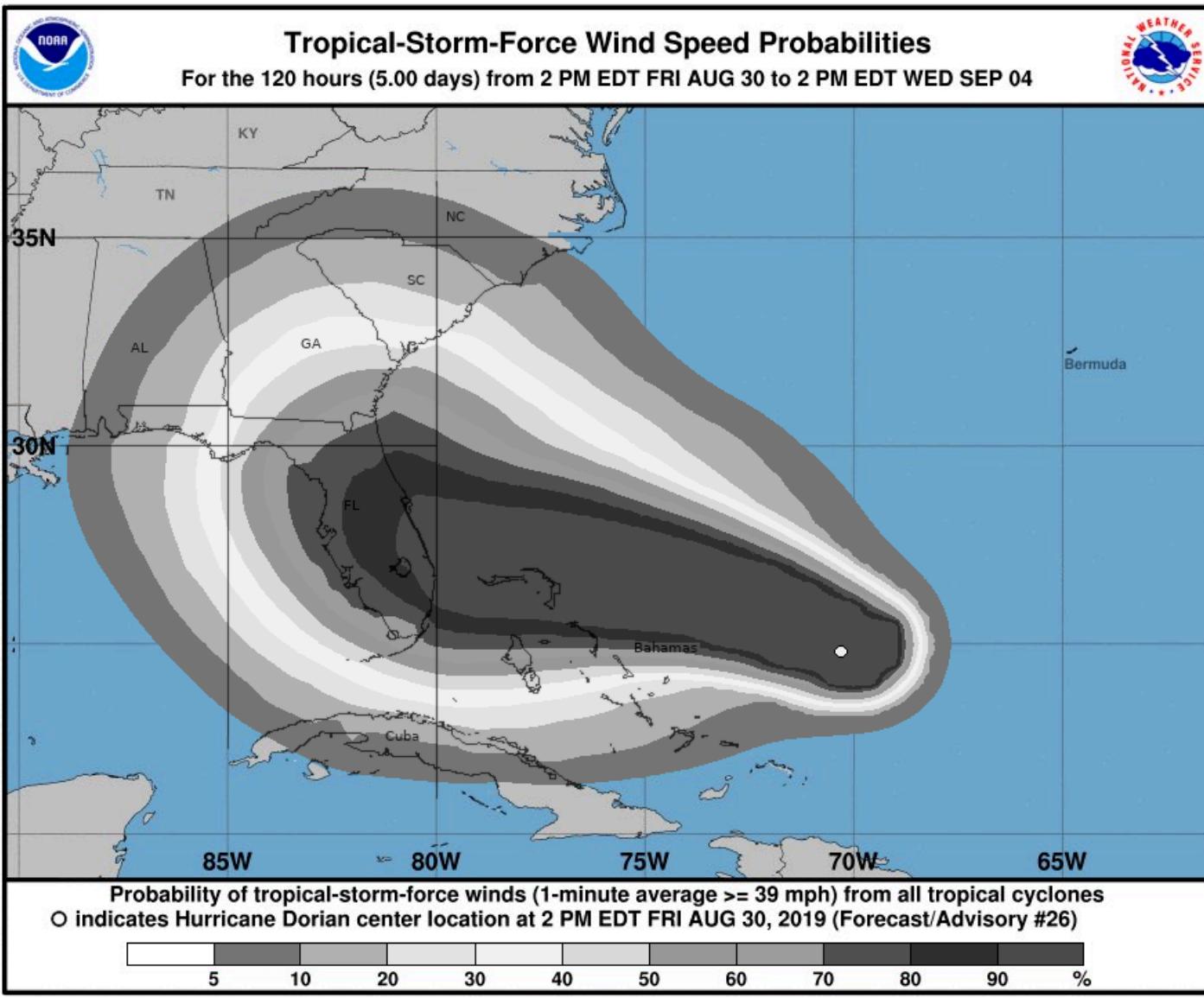




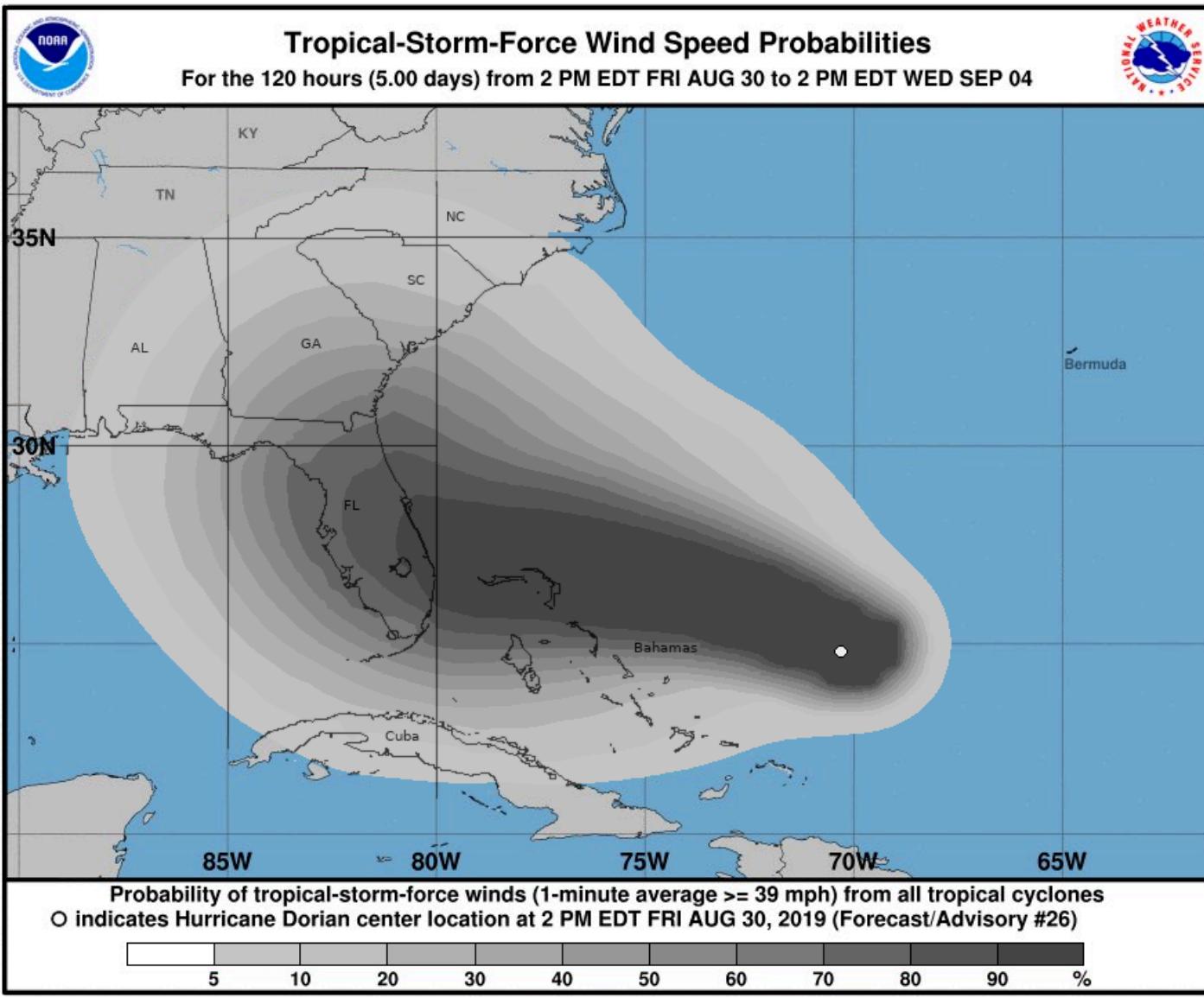




Achim Zeileis



Achim Zeileis



Achim Zeileis