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 12:30:pm to 1:30pm

1:30pm to 3:30pm

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

In between class sessions

slack

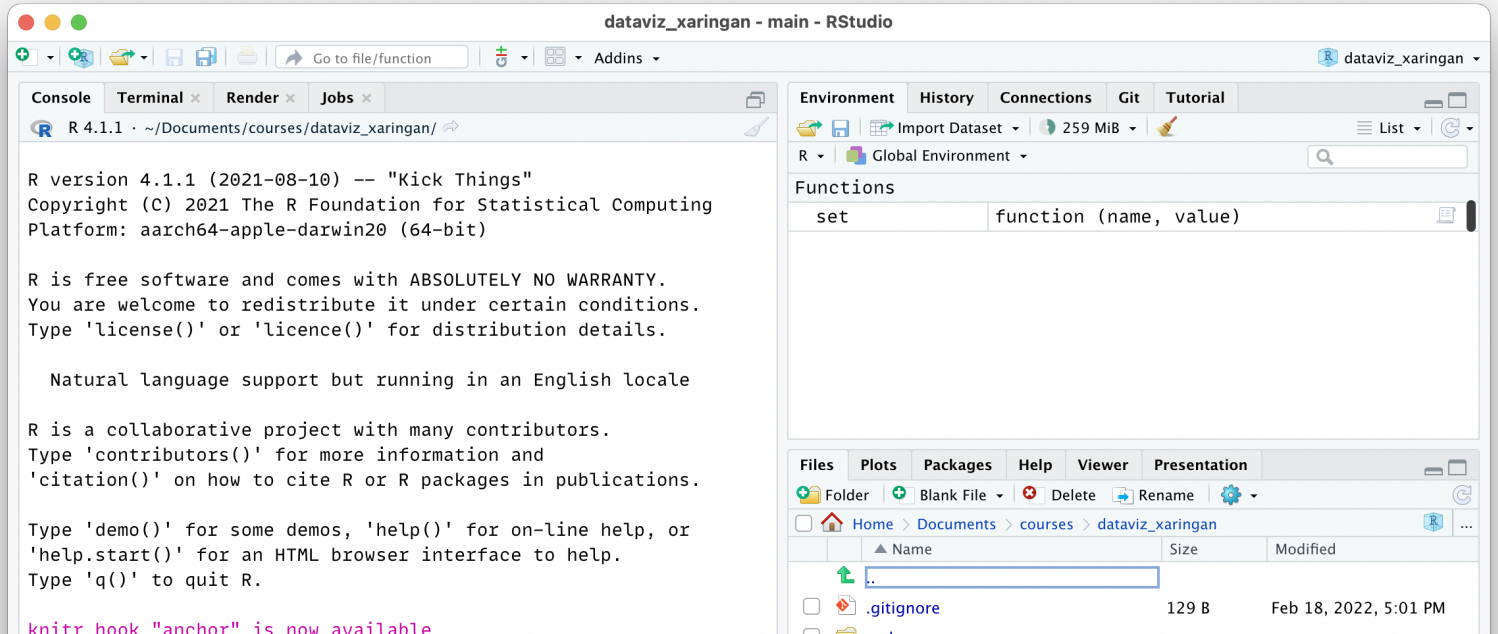
My Setup and Yours

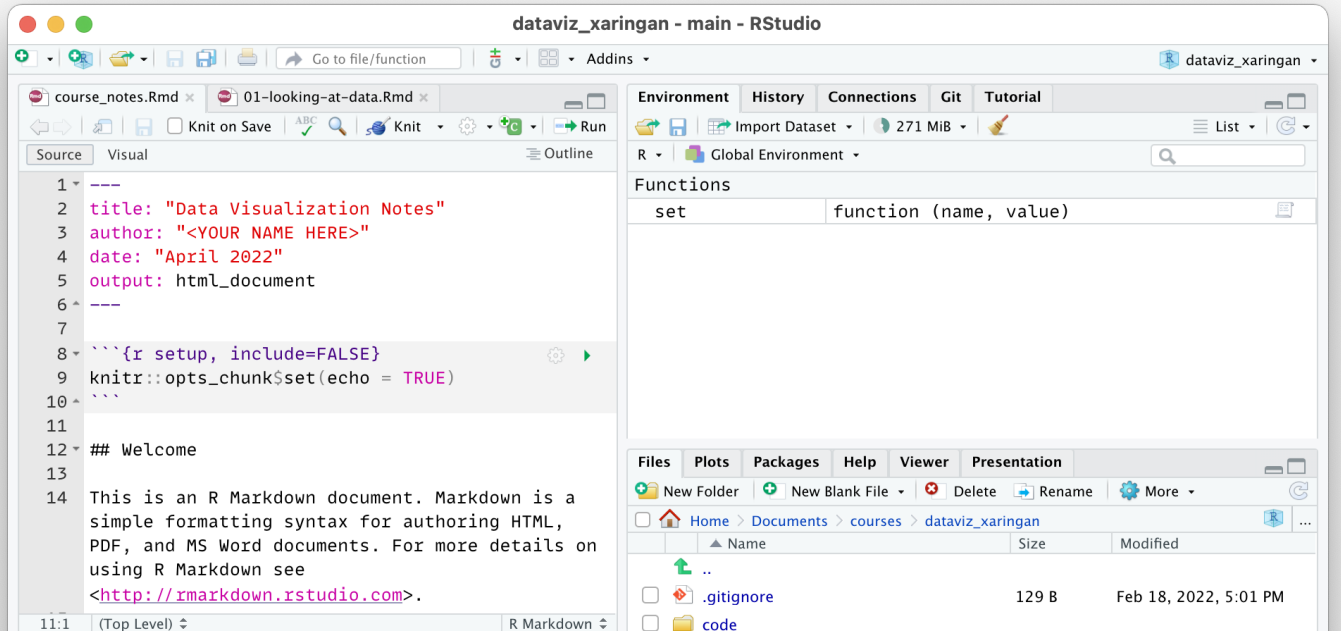
Talking, Slides, and Live-Coding in RStudio

Follow along and take notes in RStudio yourself

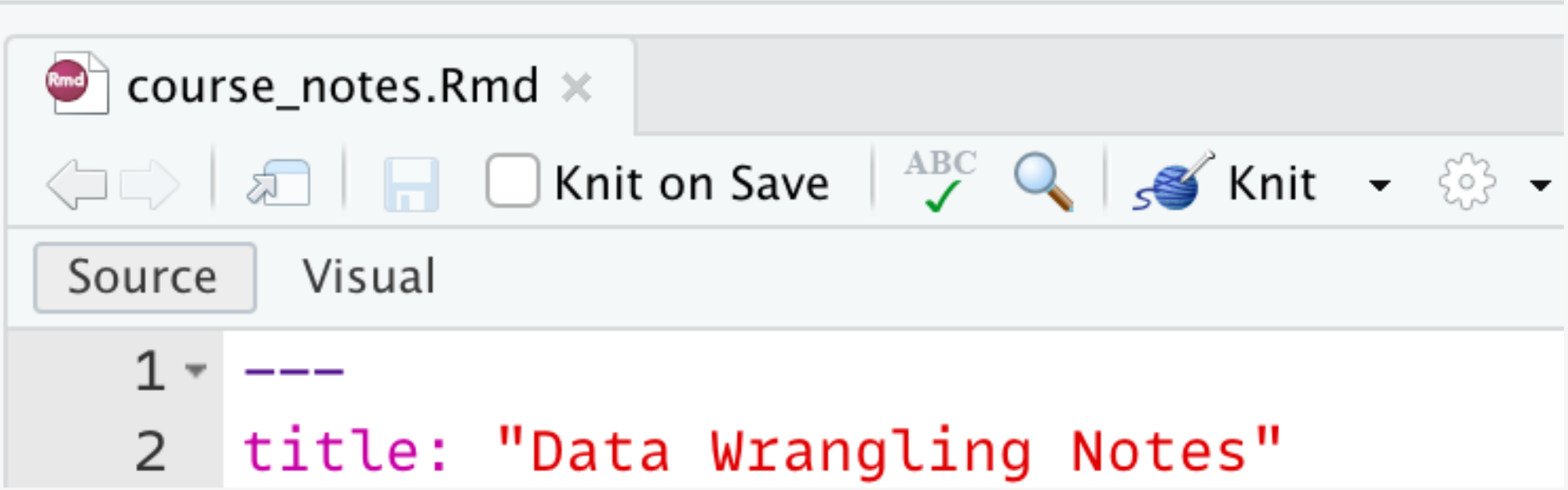
The course packet is also an RStudio project

Get up and Running


dataviz_xaringan - main - RStudio
Go to file/function
Addins
dataviz_xaringan
Console
Terminal
Render
Jobs
R 4.1.1 · ~/Documents/courses/dataviz_xaringan/
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
Environment
History
Connections
Git
Tutorial
Import Dataset
259 MiB
List
R
Global Environment
Functions
set
function (name, value)
Files
Plots
Packages
Help
Viewer
Presentation
Folder
Blank File
Delete
Rename
Home
Documents
courses
dataviz_xaringan
Name
Size
Modified
..
.gitignore
129 B
Feb 18, 2022, 5:01 PM


dataviz_xaringan - main - RStudio
Go to file/function
Addins
dataviz_xaringan
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
Environment
History
Connections
Git
Tutorial
Import Dataset
271 MiB
List
R
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

Try rendering your notes



Now write the following code

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

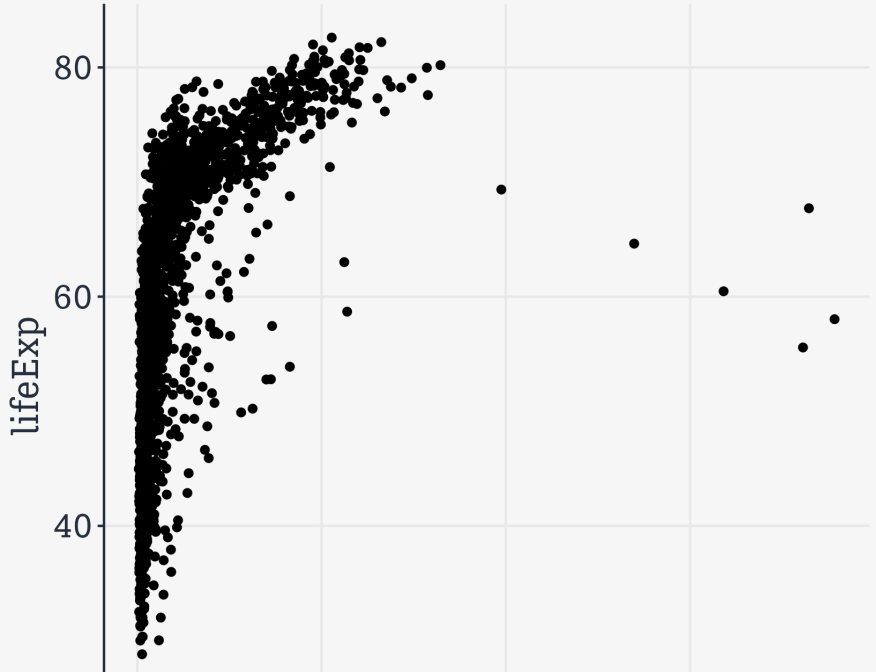
```
library(tidyverse)
library(gapminder)

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.

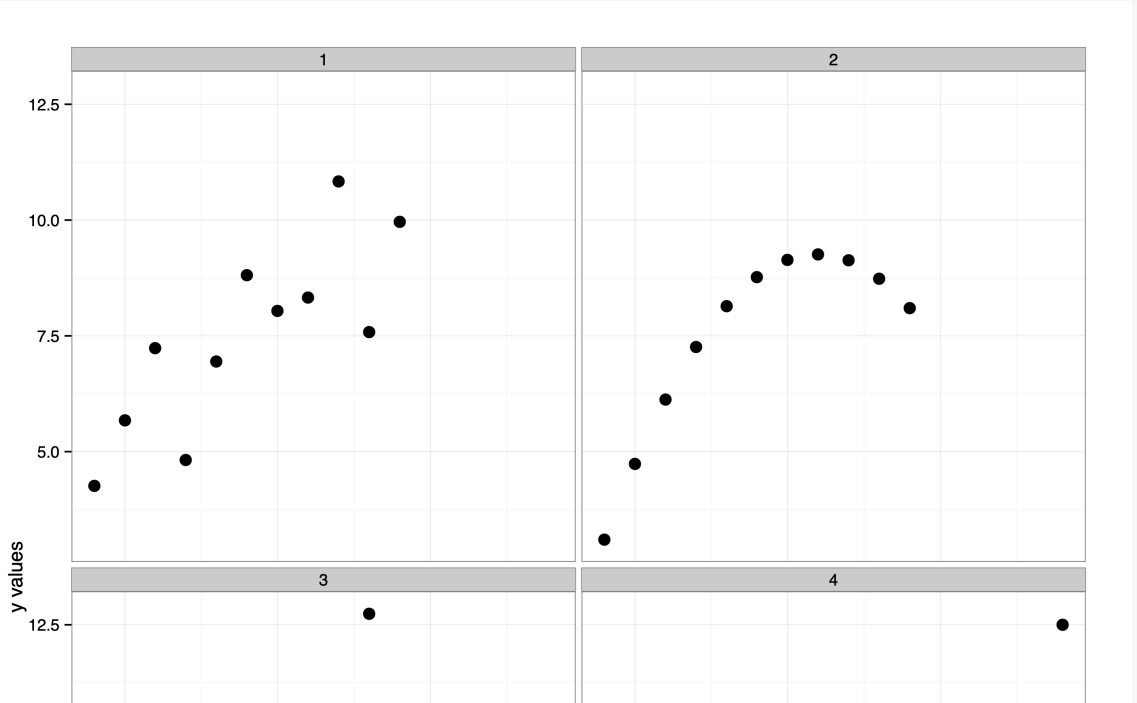
```
library(tidyverse)
library(gapminder)

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



p + geom_point()
```

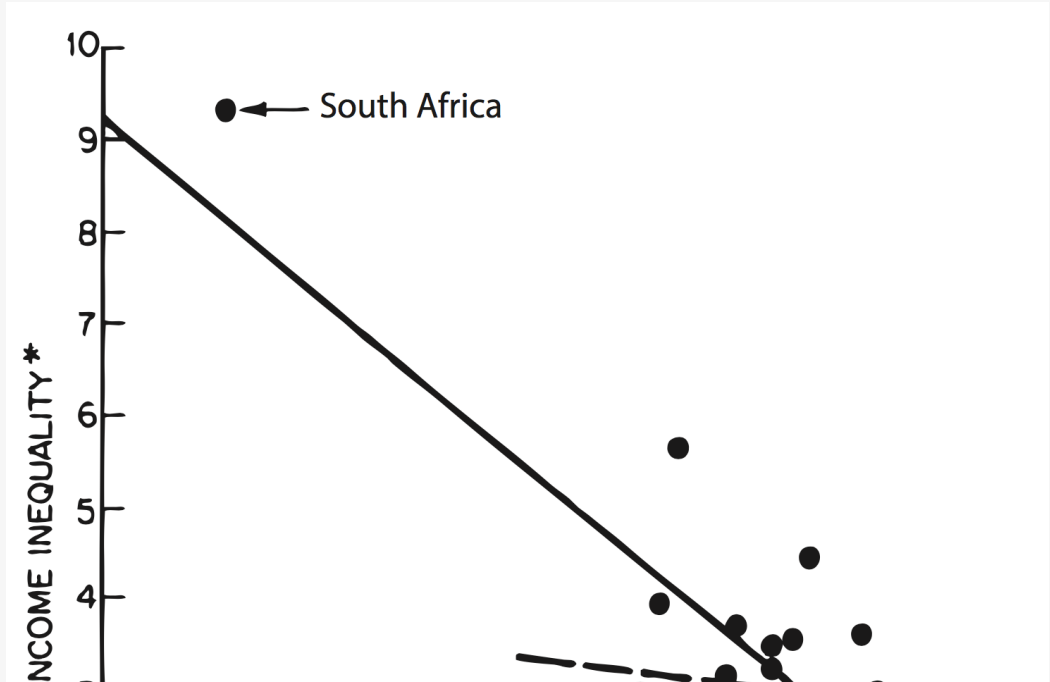


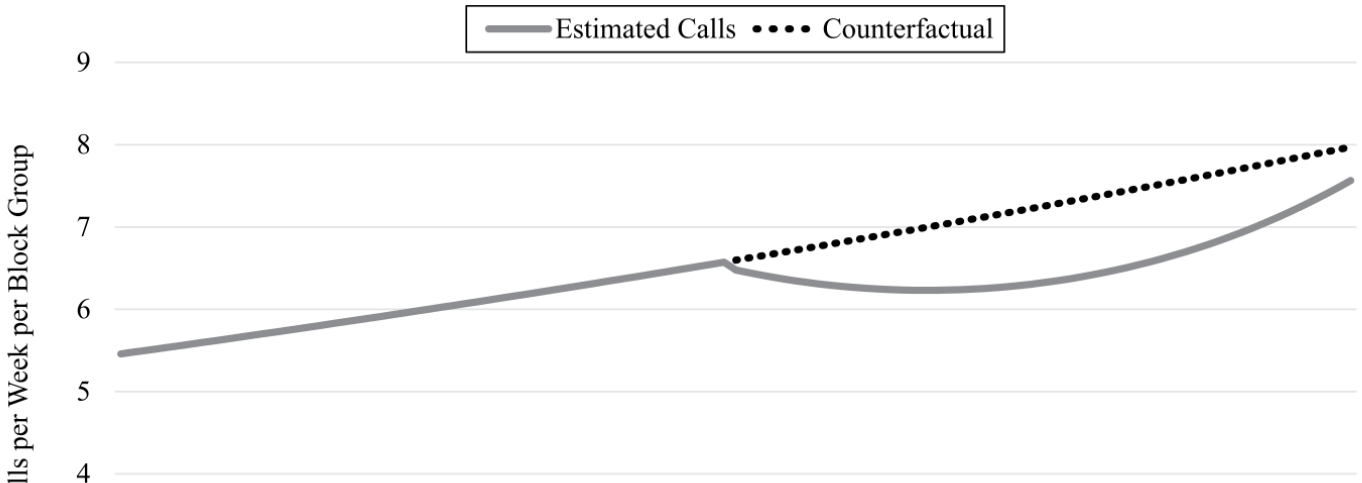
You should

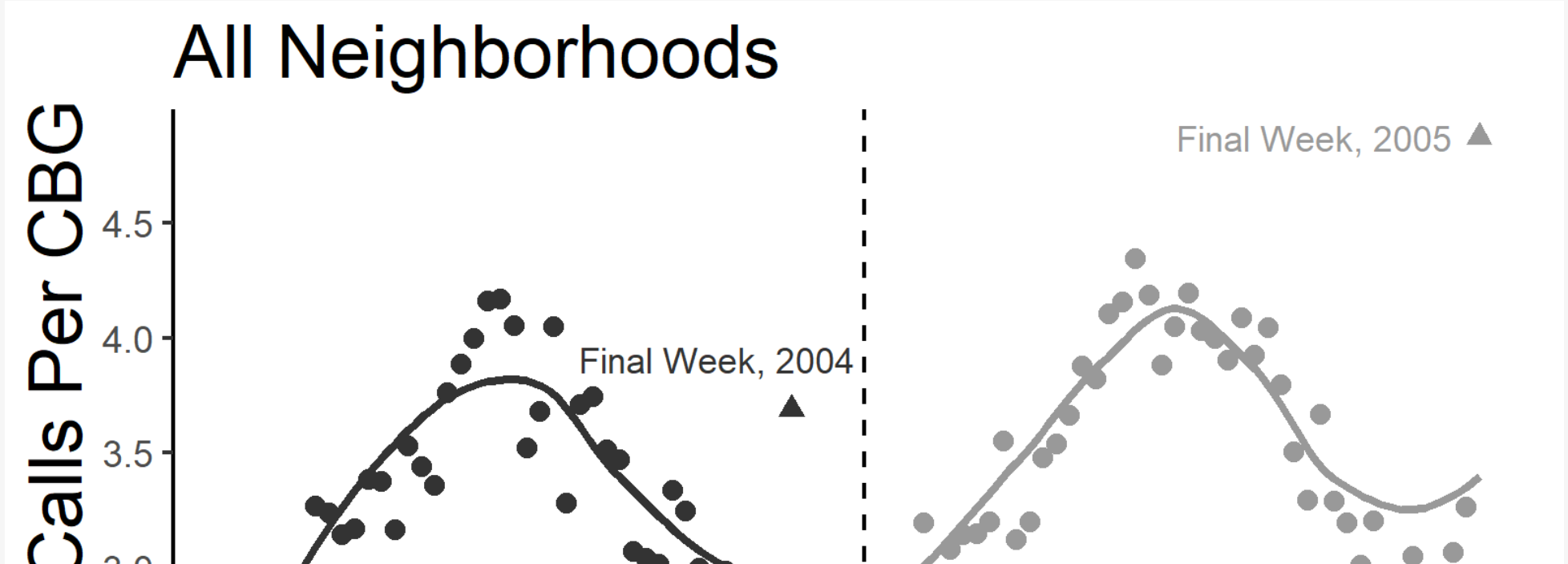
.kjh-yellowlook at
your data

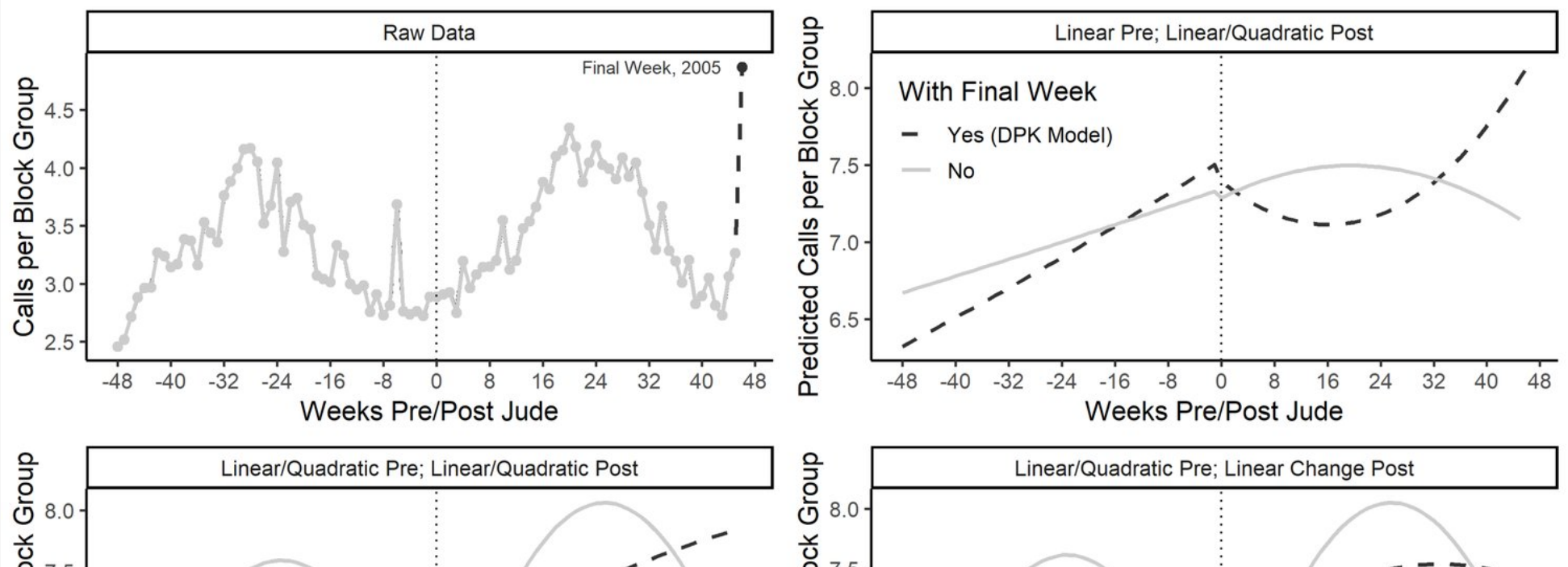
Seeing things

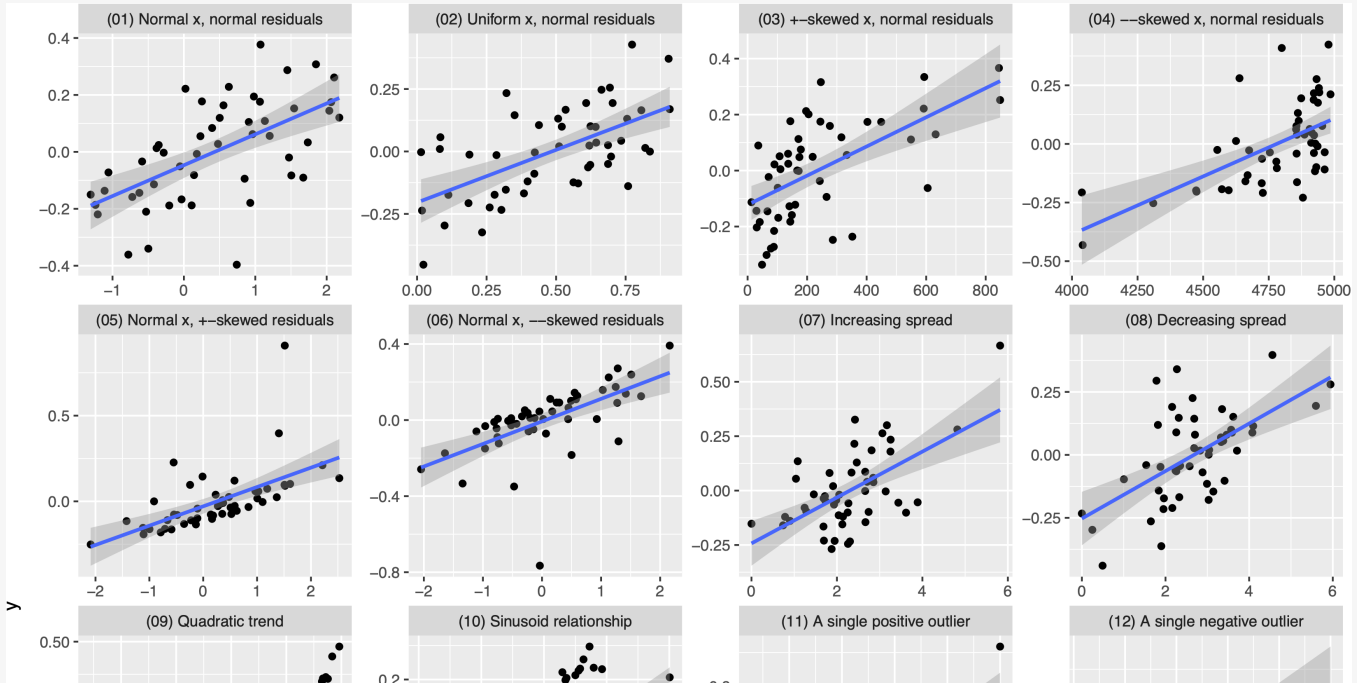

1
2
3
4
12.5
10.0
7.5
5.0
y values

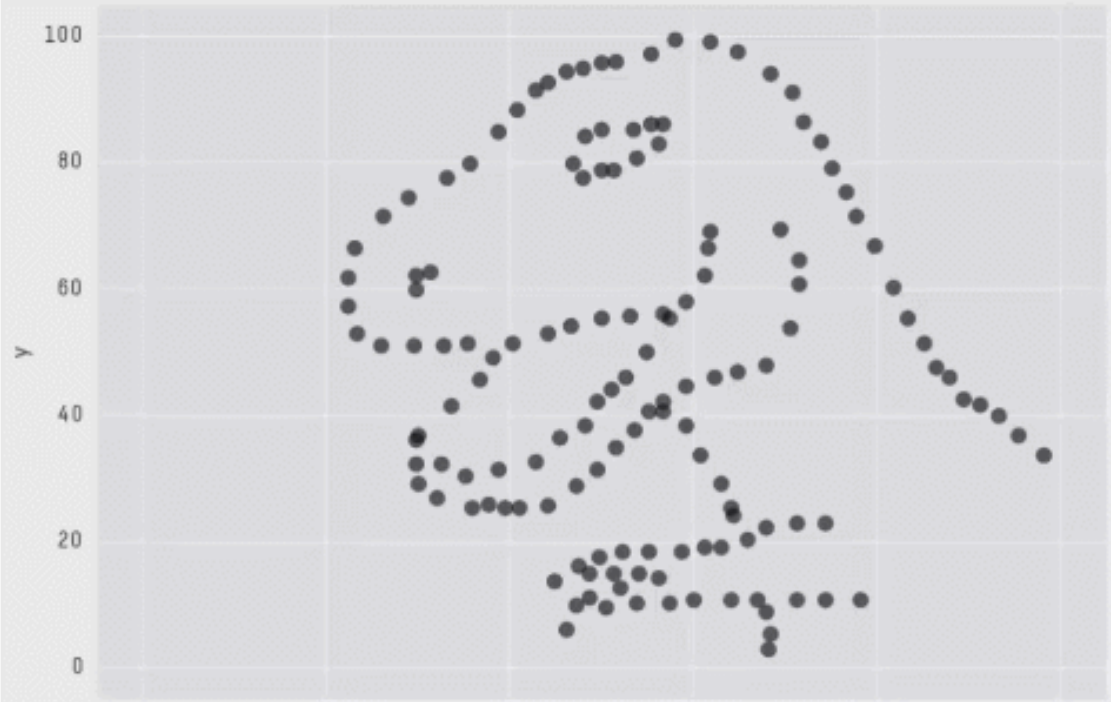

10
9
8
7
6
5
4
South Africa
INCOME INEQUALITY*


Estimated Calls
Counterfactual
9
8
7
6
5
4
lls per Week per Block Group


All Neighborhoods
Calls Per CBG
4.5
4.0
3.5
Final Week, 2004
Final Week, 2005

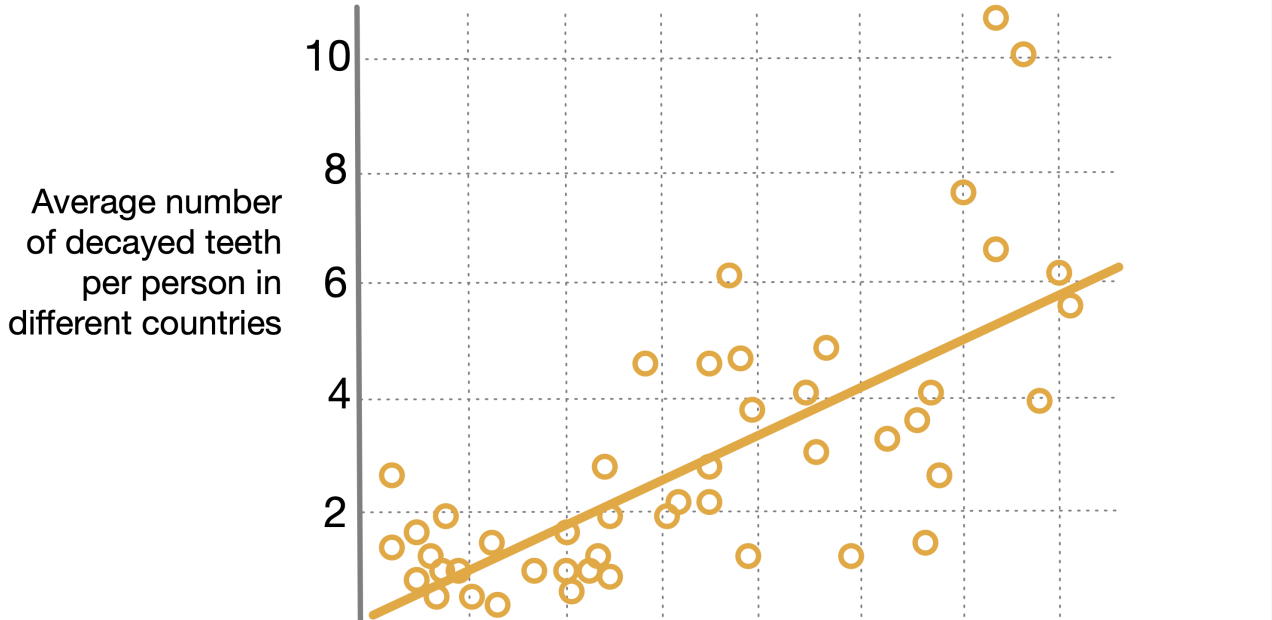

Raw Data
Final Week, 2005
Calls per Block Group
Weeks Pre/Post Jude
Linear Pre; Linear/Quadratic Post
With Final Week
Yes (DPK Model)
No
Predicted Calls per Block Group
Weeks Pre/Post Jude
Linear/Quadratic Pre; Linear/Quadratic Post
Linear/Quadratic Pre; Linear Change Post


(01) Normal x, normal residuals
(02) Uniform x, normal residuals
(03) +–skewed x, normal residuals
(04) ––skewed x, normal residuals
(05) Normal x, +–skewed residuals
(06) Normal x, ––skewed residuals
(07) Increasing spread
(08) Decreasing spread
(09) Quadratic trend
(10) Sinusoid relationship
(11) A single positive outlier
(12) A single negative outlier
y

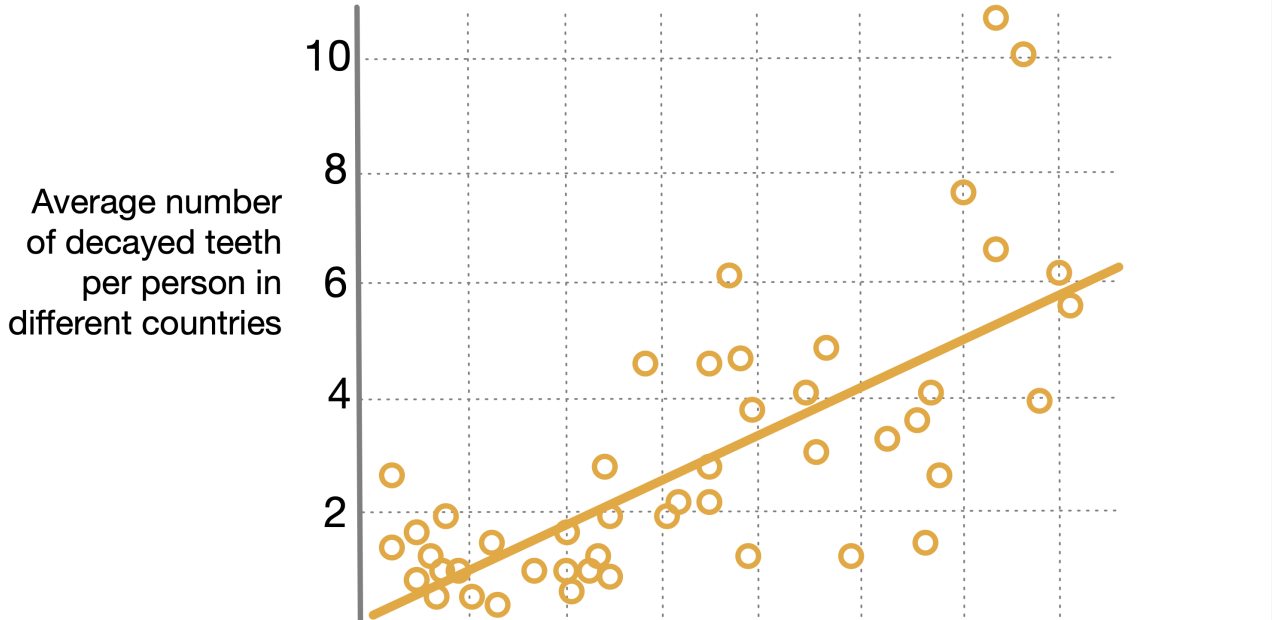


X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

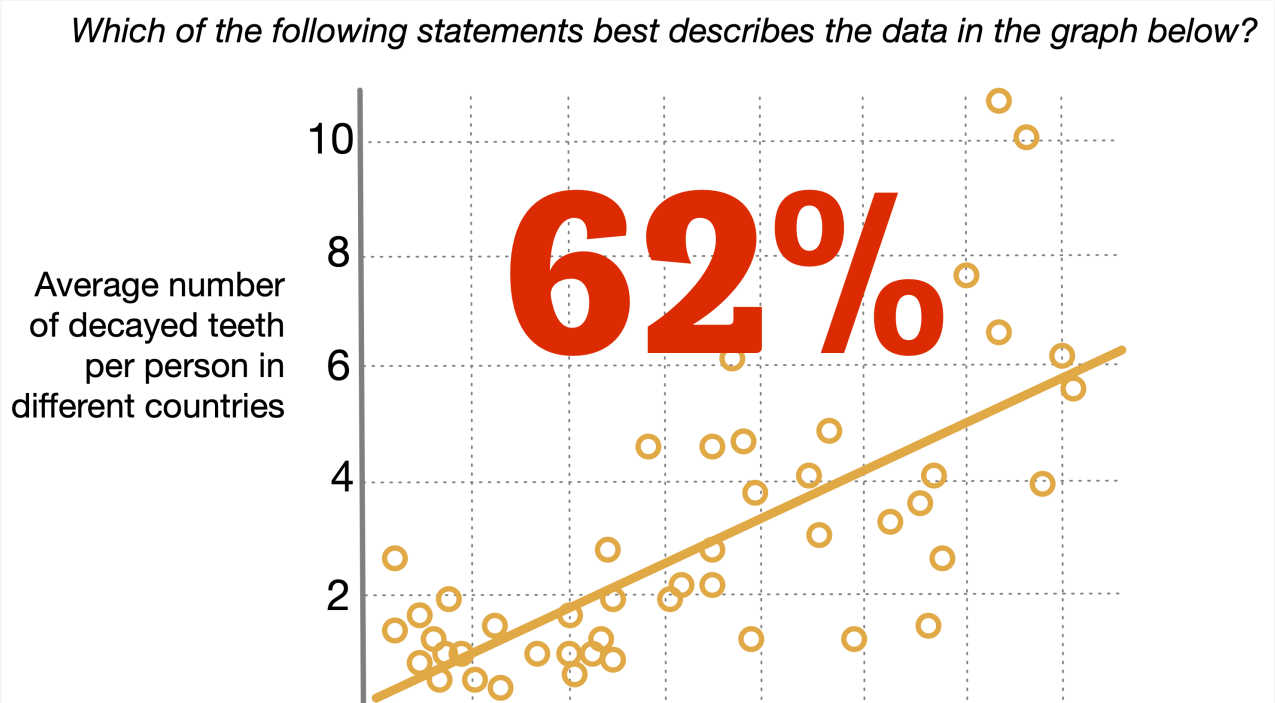
Which of the following statements best describes the data in the graph below?



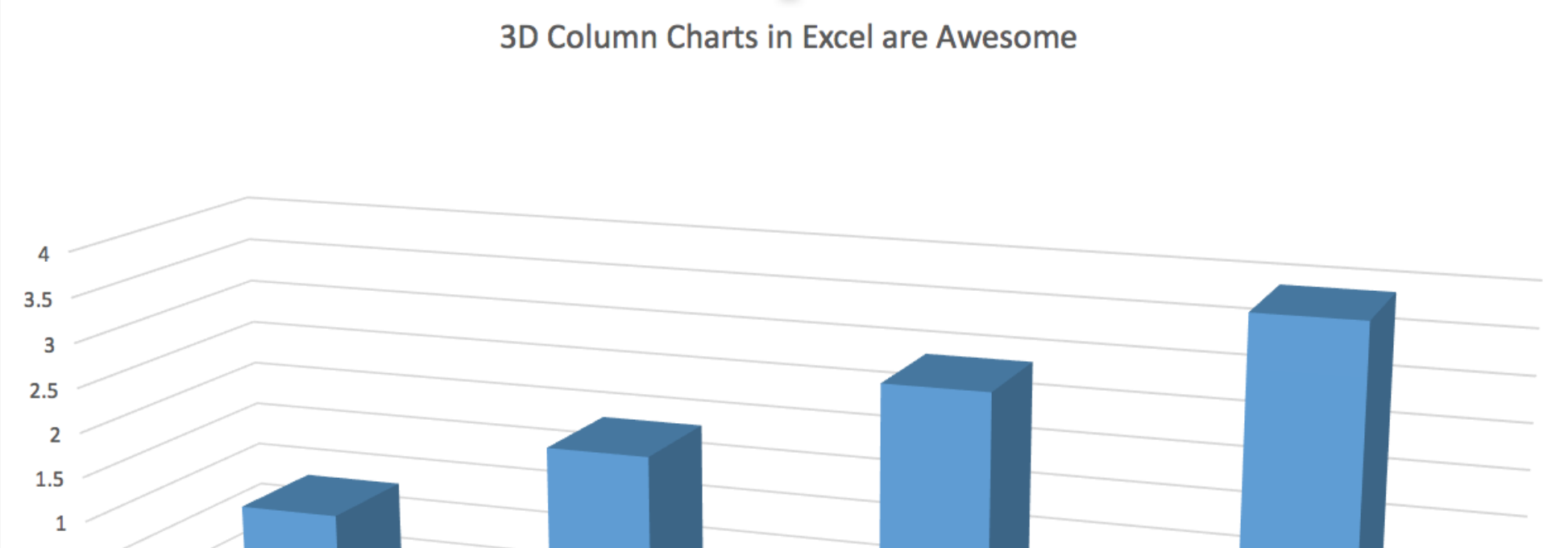
Which of the following statements best describes the data in the graph below?



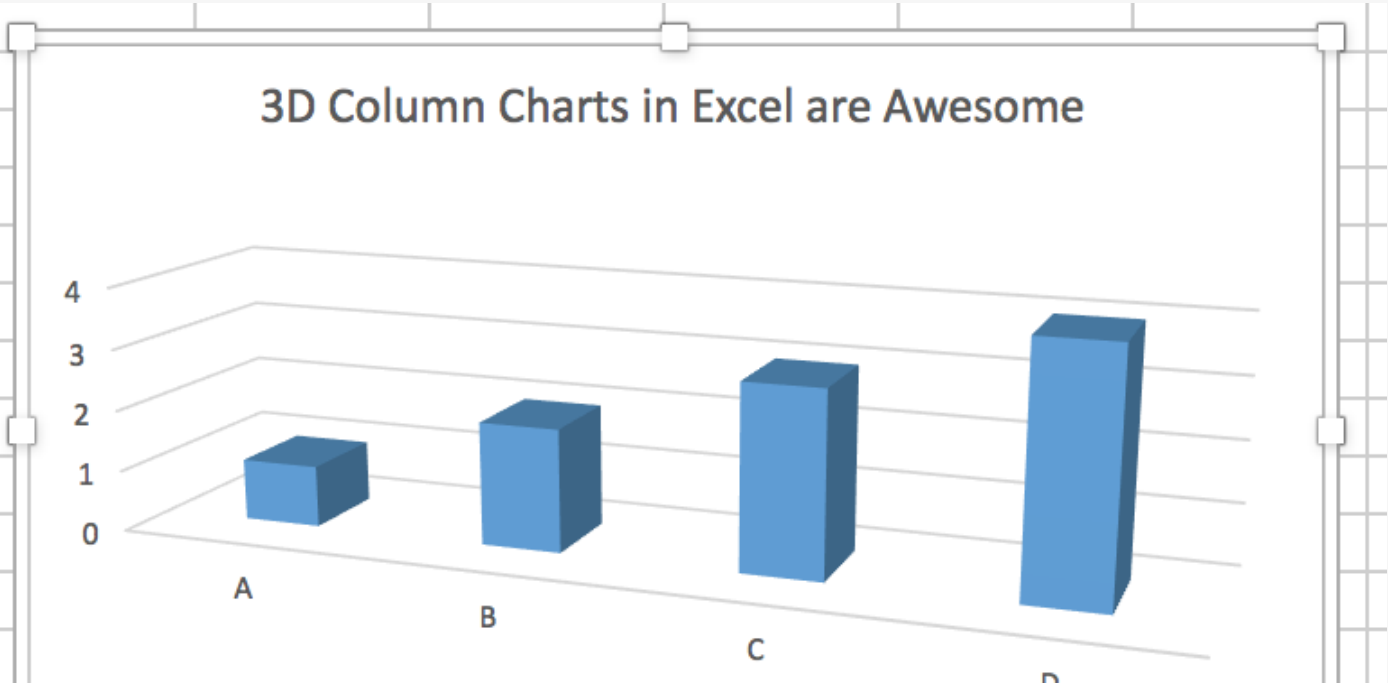
Which of the following statements best describes the data in the graph below?



Not Seeing Things


3D Column Charts in Excel are Awesome
4
3.5
3
2.5
2
1.5
1

A	1
B	2
C	3
D	4



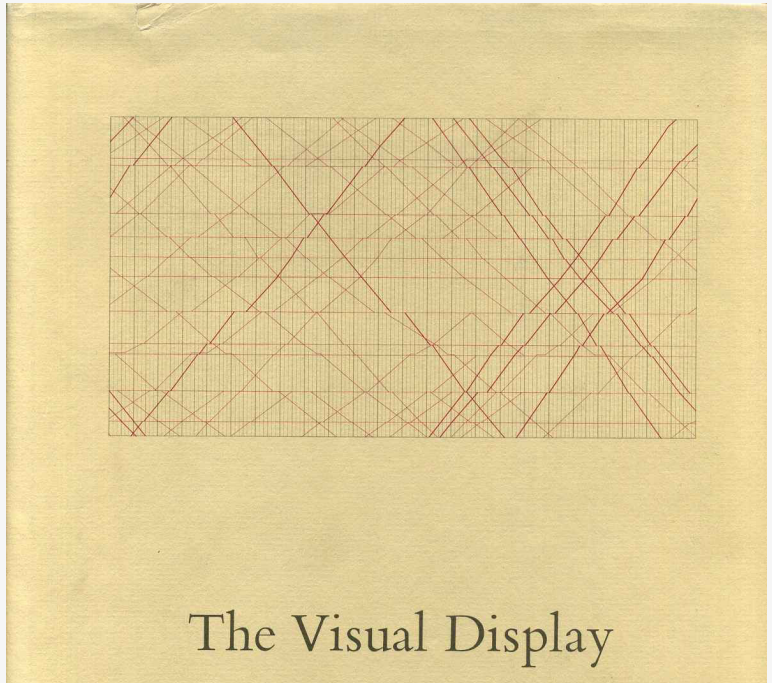
Bad Taste

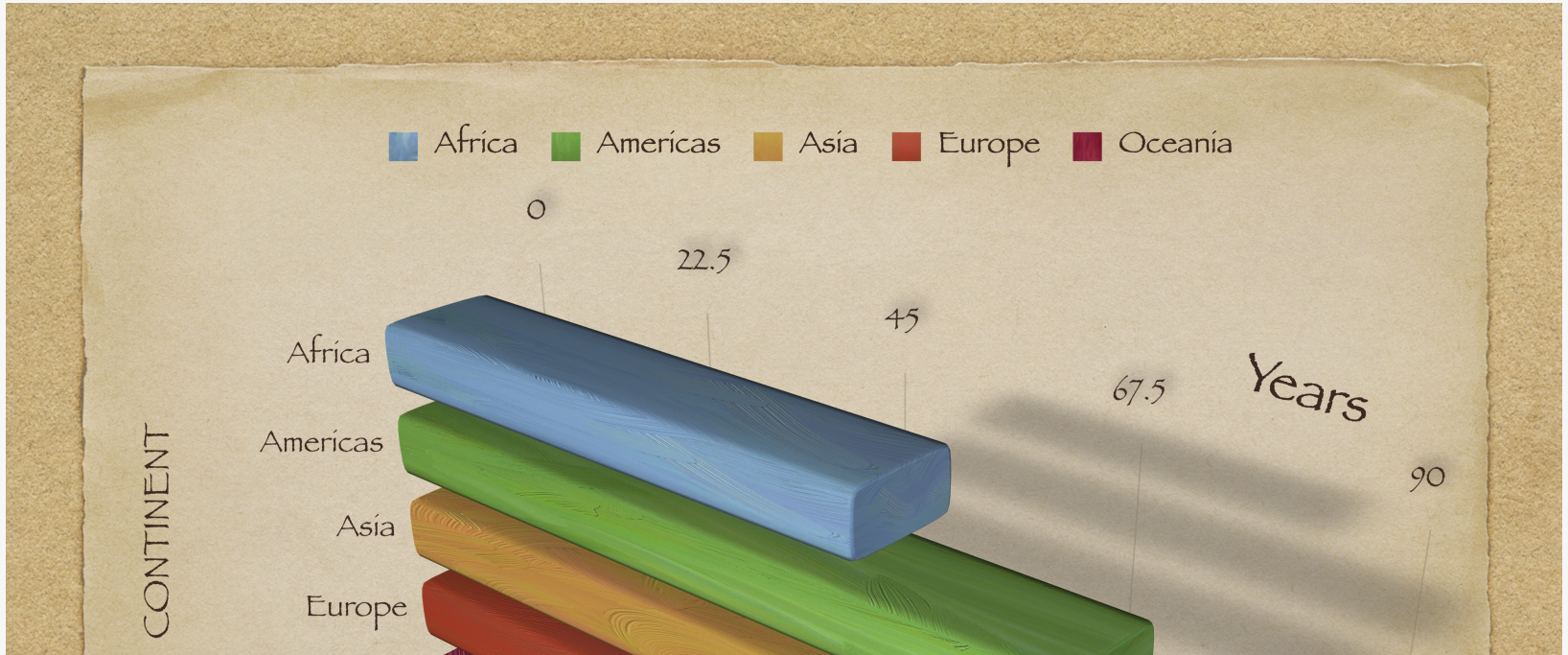
Bad Data

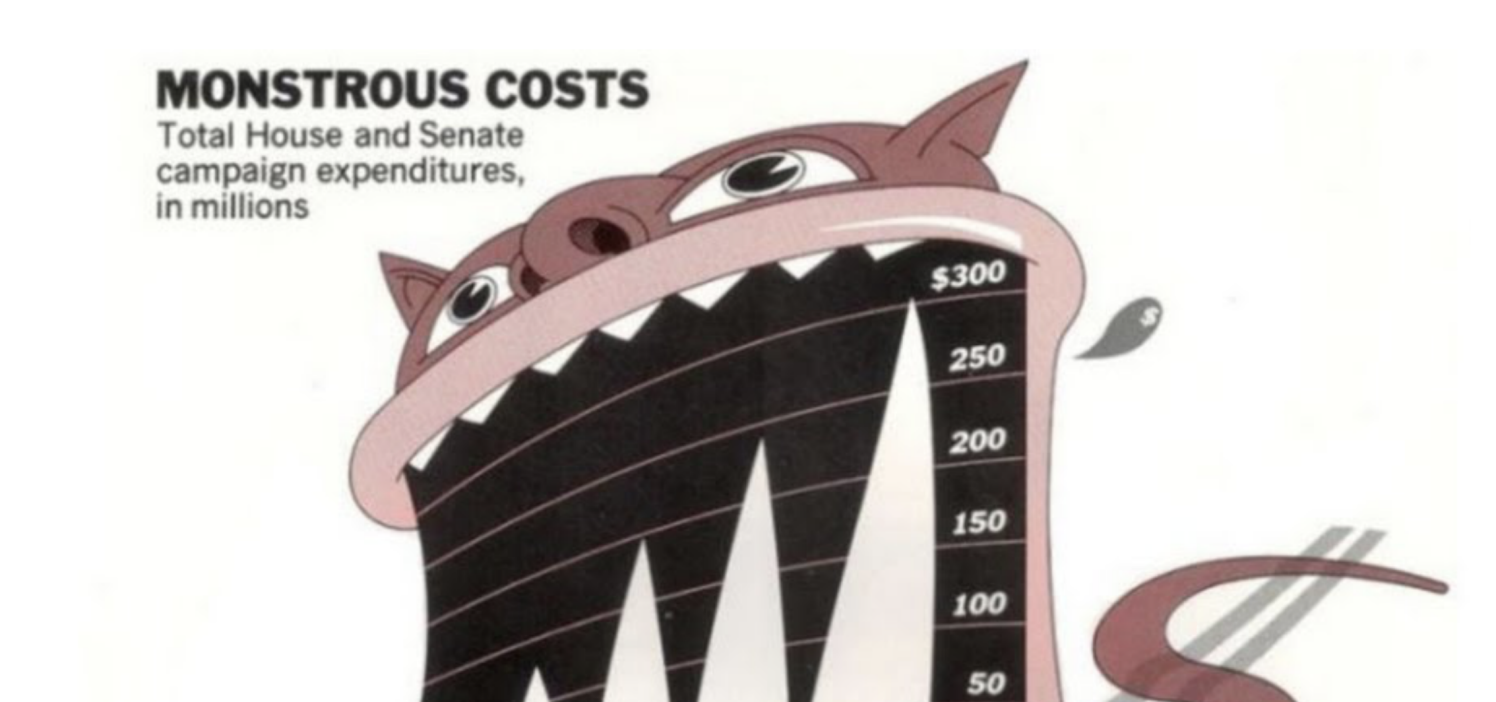
Bad Perception

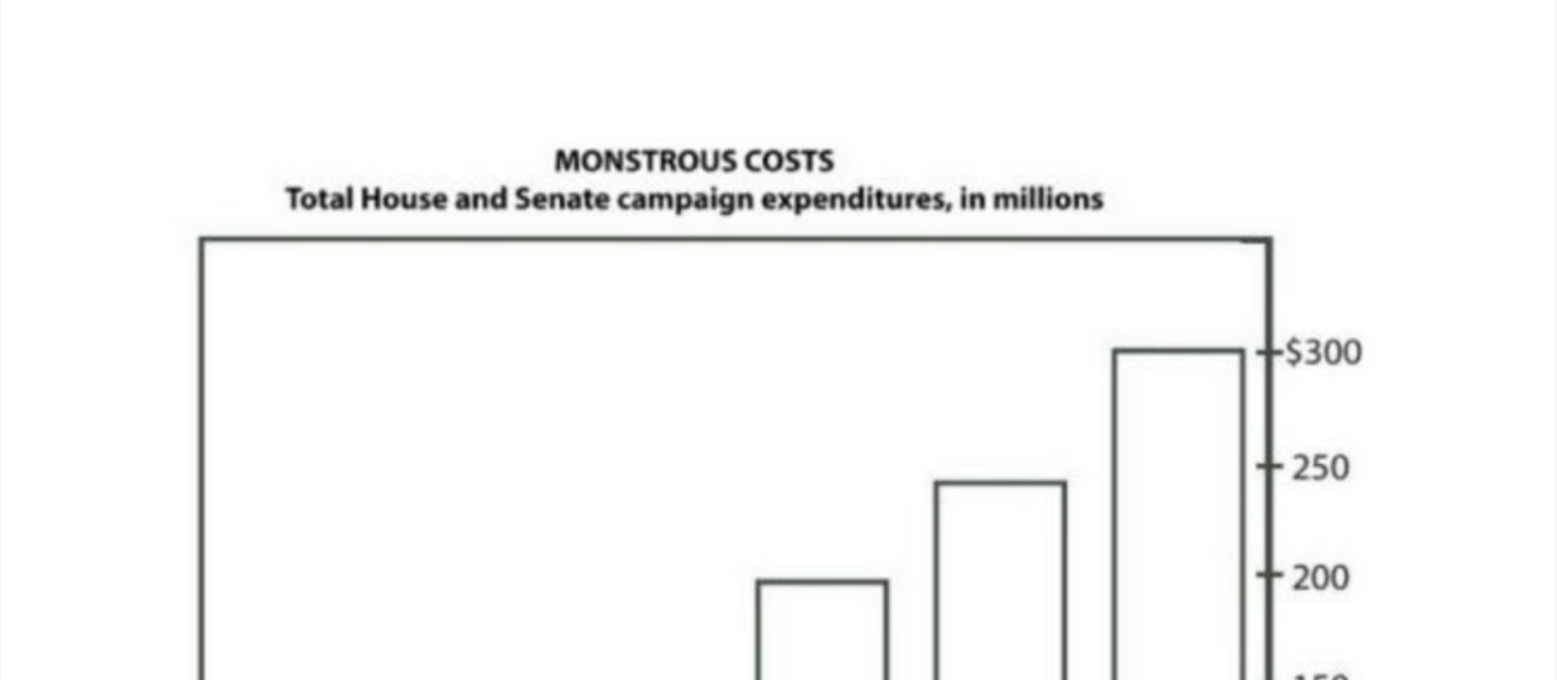
Bad Taste: Simplify, Simplify?

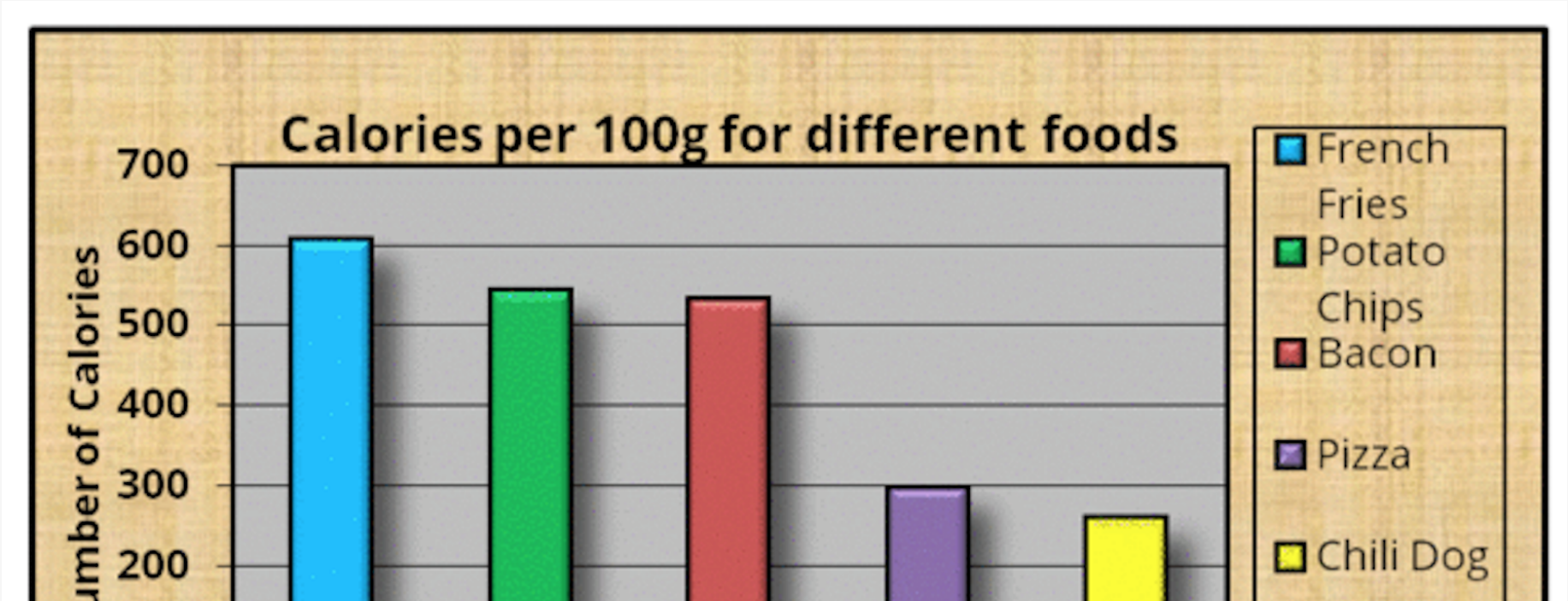
Tufte's "Data to Ink Ratio"

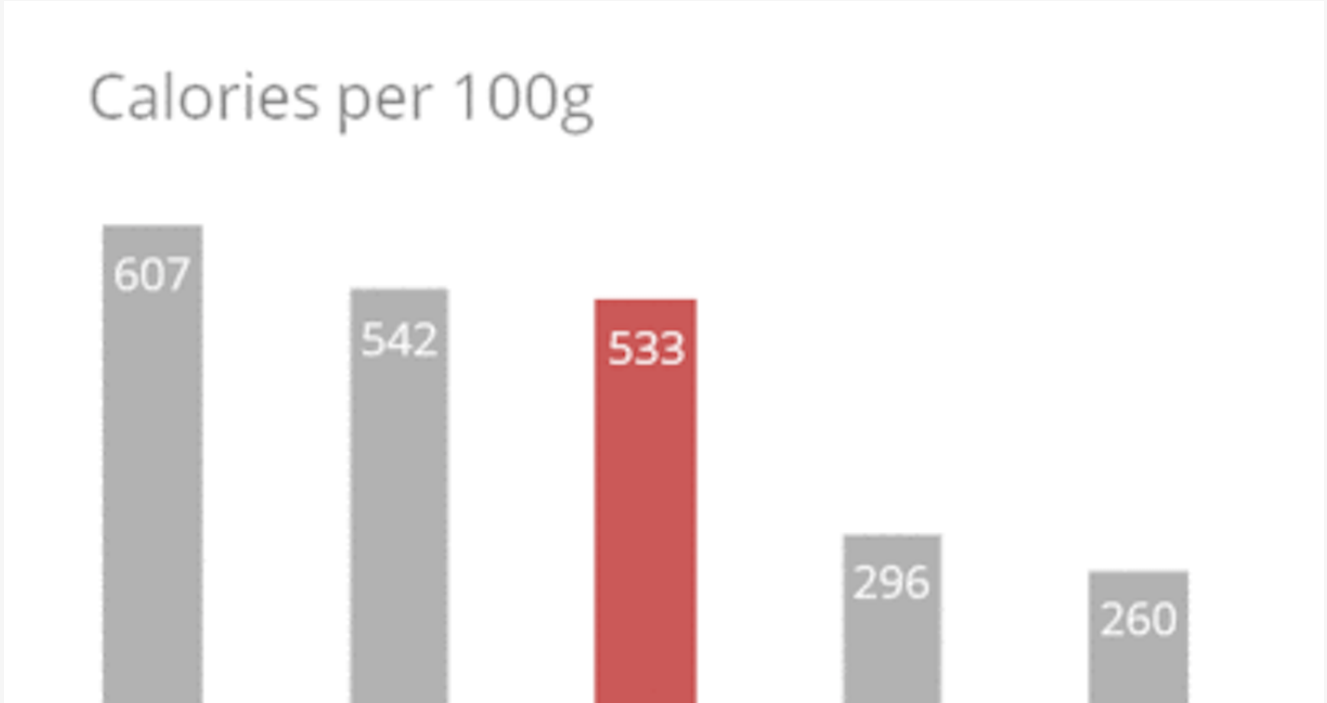



Africa
Americas
Asia
Europe
Oceania
0
22.5
45
67.5
90
Years
CONTINENT
Africa
Americas
Asia
Europe


MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions
$300
250
200
150
100
50
$


MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions
$300
250
200

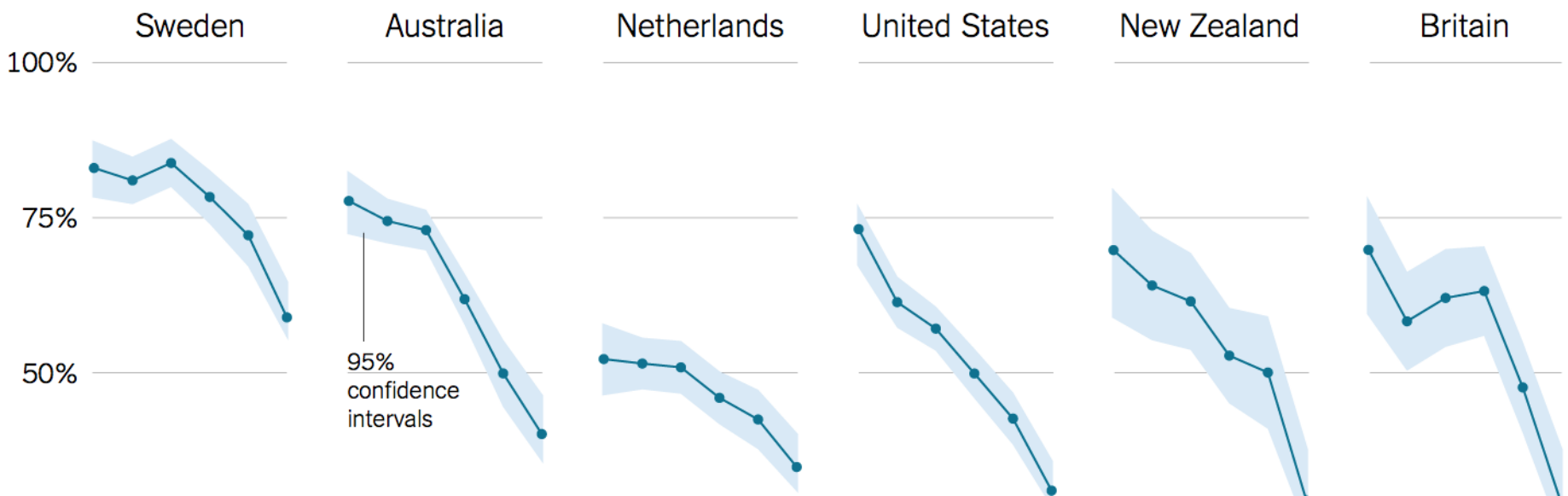

Calories per 100g for different foods
Number of Calories
700
600
500
400
300
200
French Fries
Potato Chips
Bacon
Pizza
Chili Dog

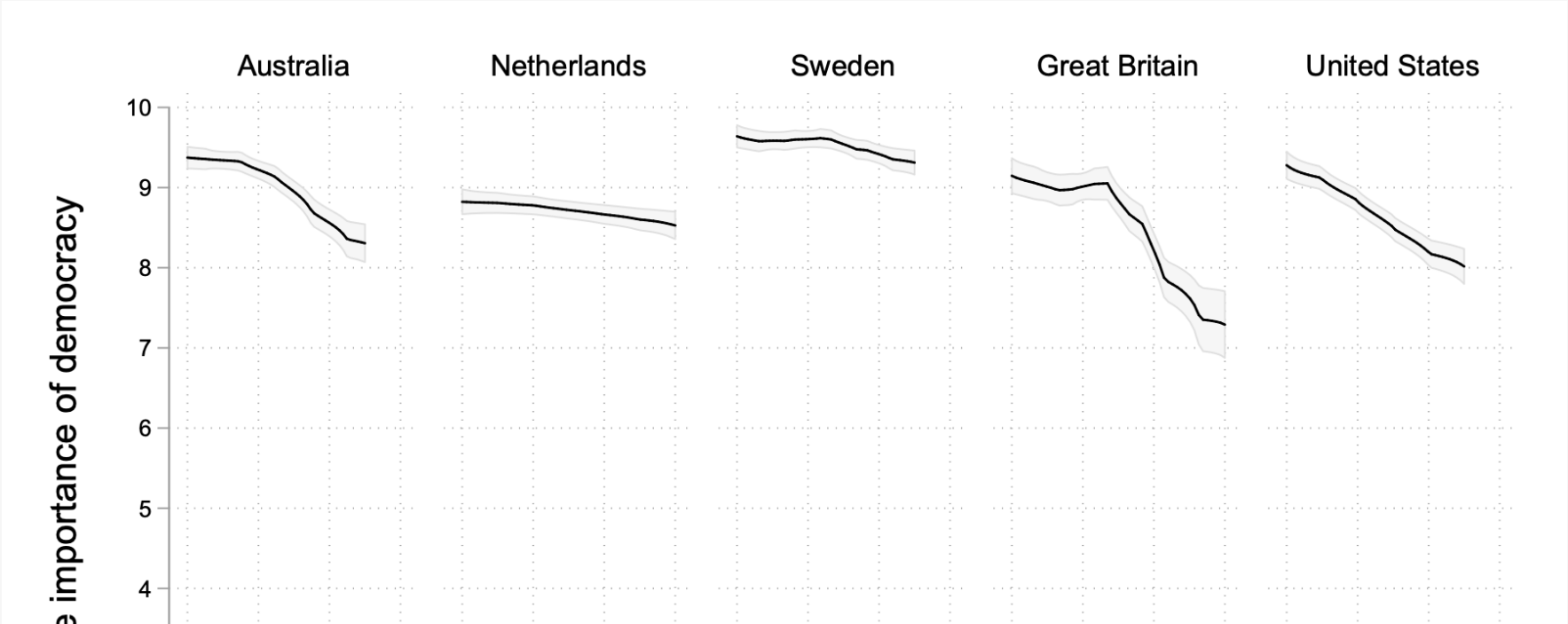

Calories per 100g
607
542
533
296
260

a)
b)
c)
y
1.5
2.0
2.5
3.0
3.5

Bad Data: Junk-Free Junk Charts

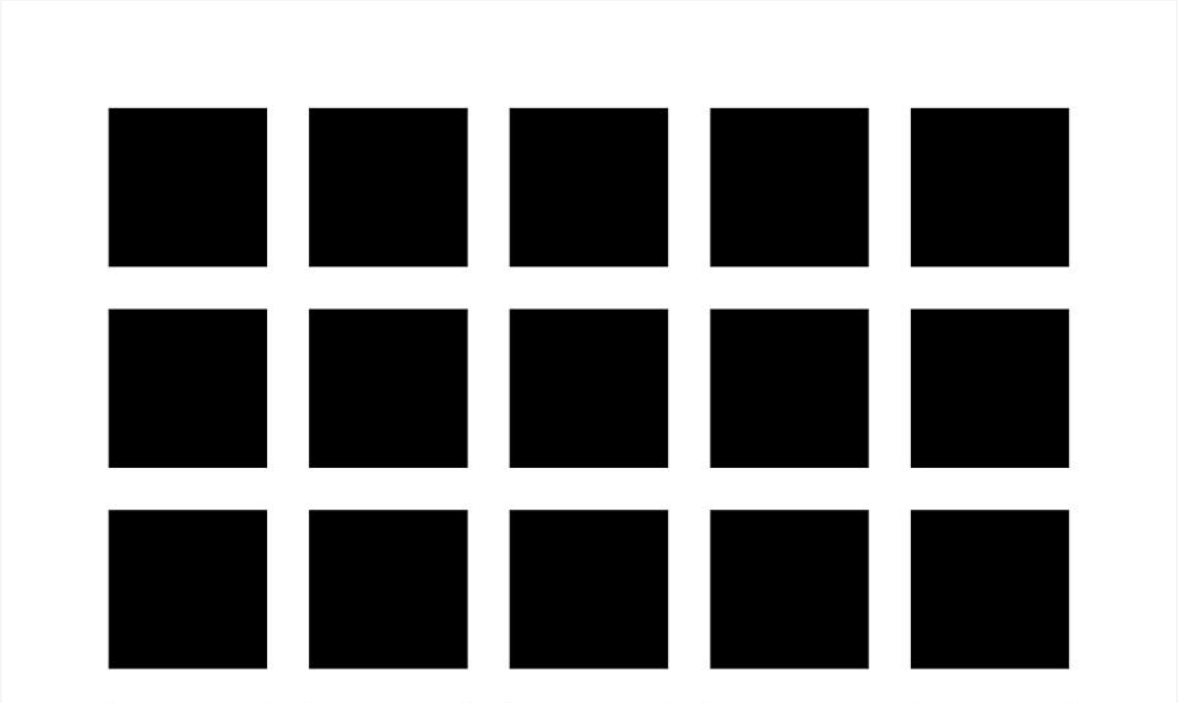
Percentage of people who say it is “essential” to live in a democracy

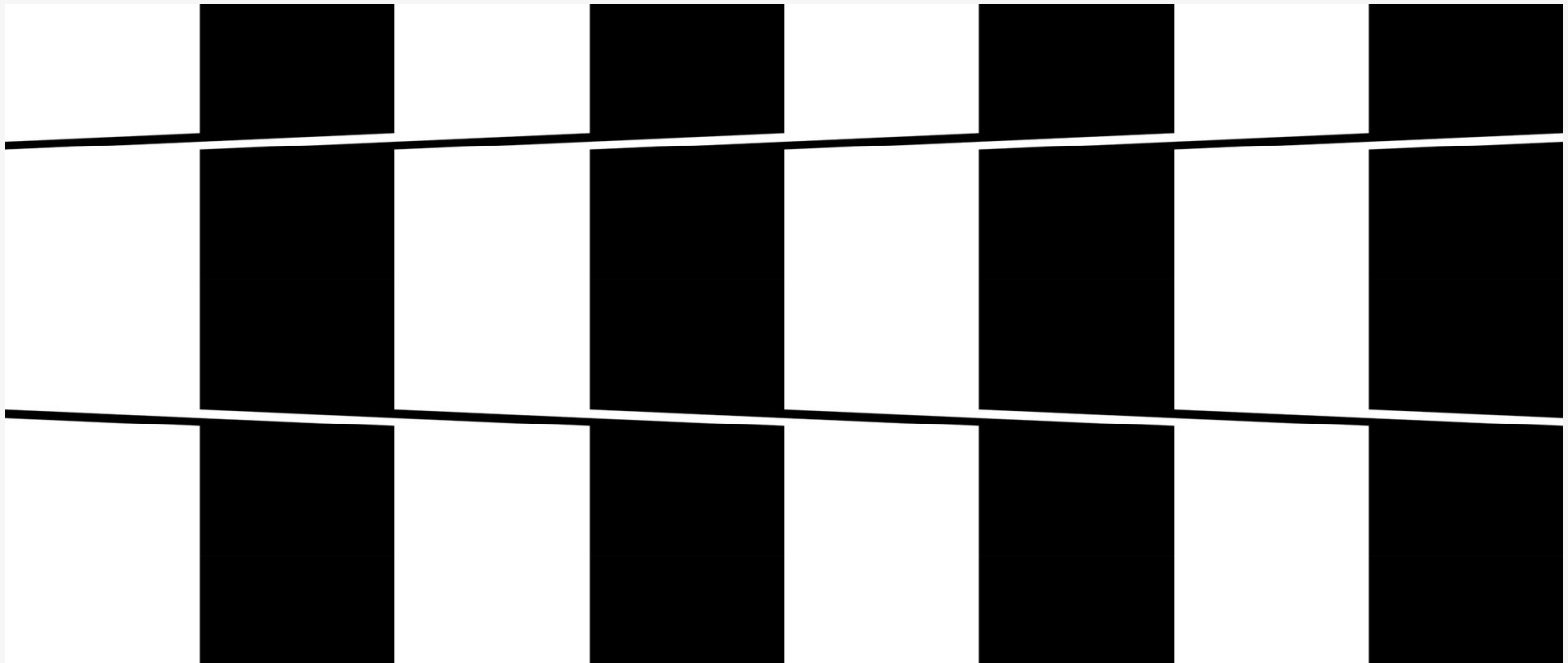


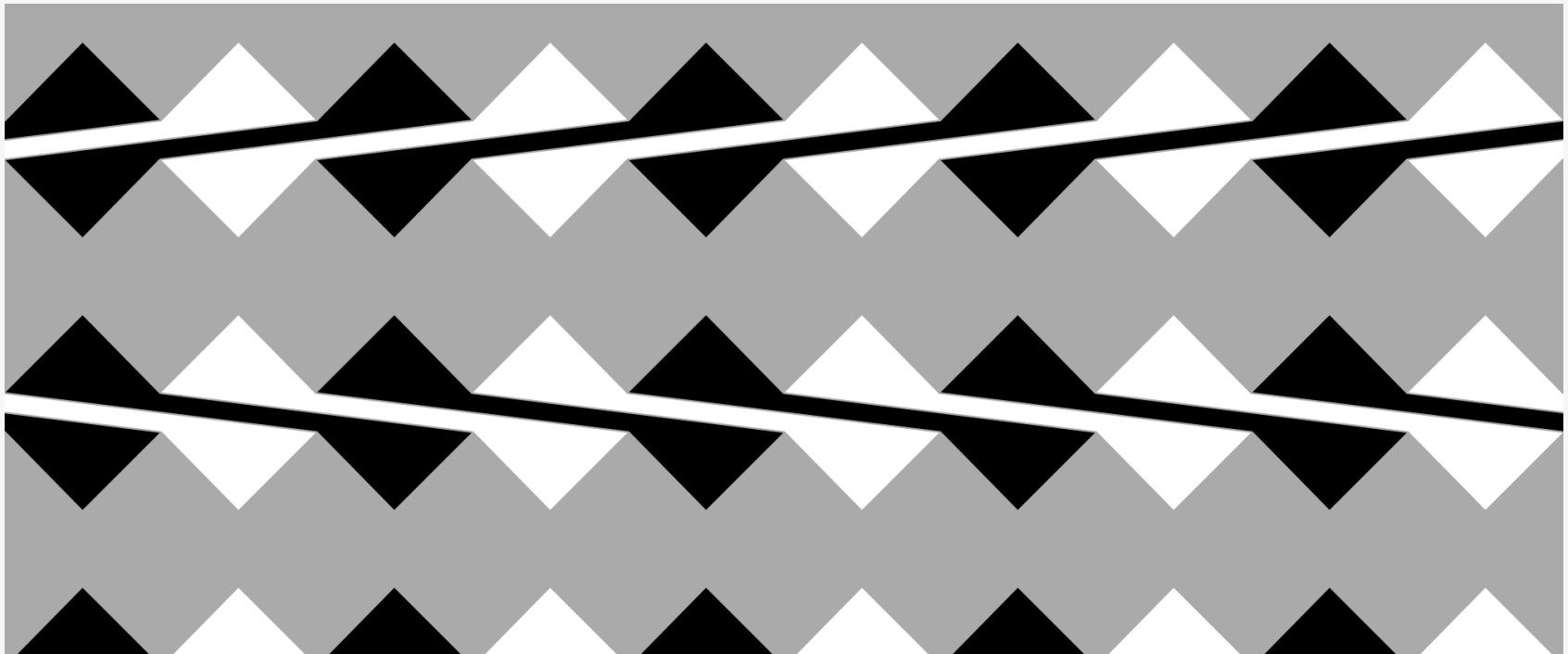

Australia
Netherlands
Sweden
Great Britain
United States
10
9
8
7
6
5
4
e importance of democracy

Bad Perception: Seeing and Not Seeing

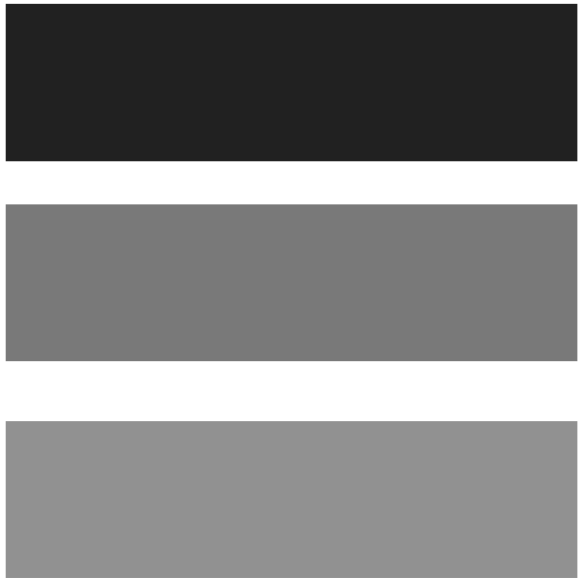
Edges & Contrasts

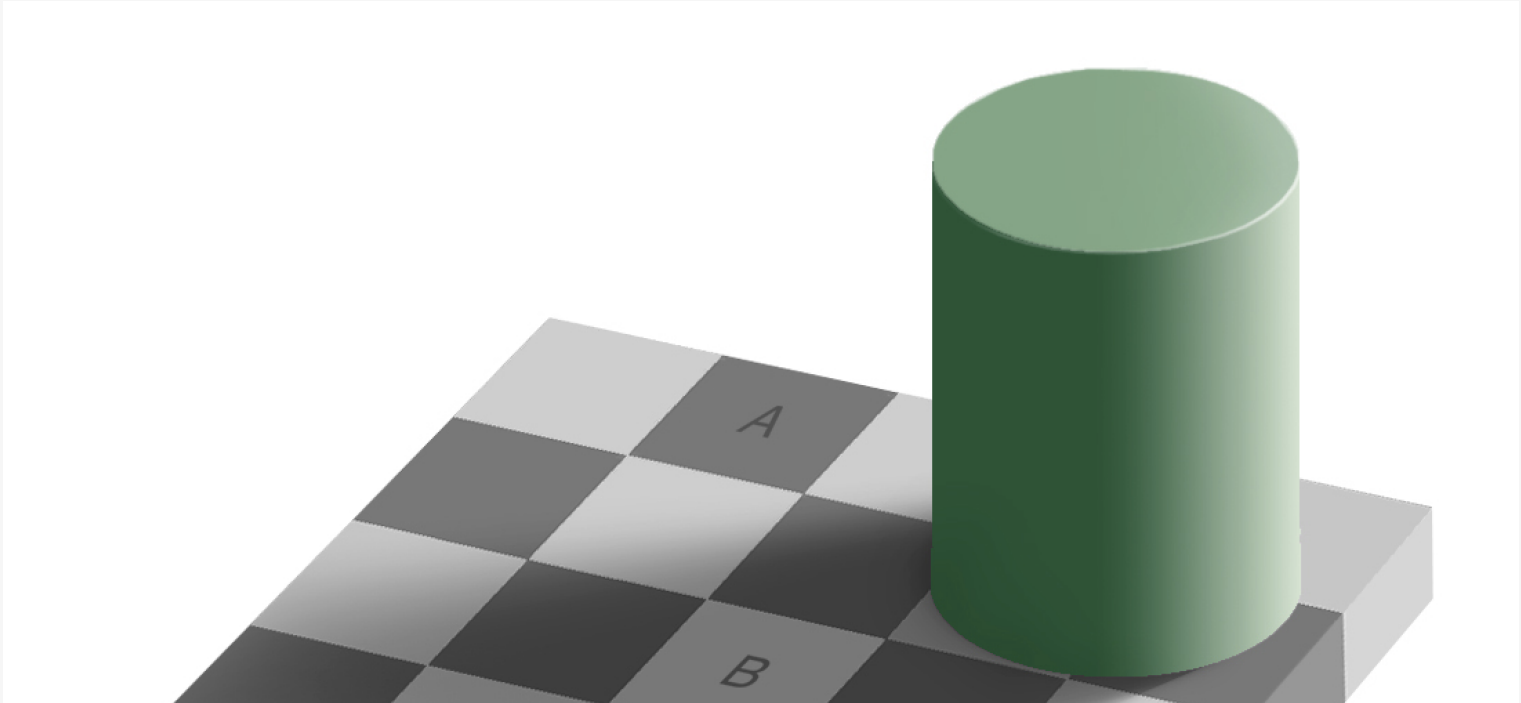


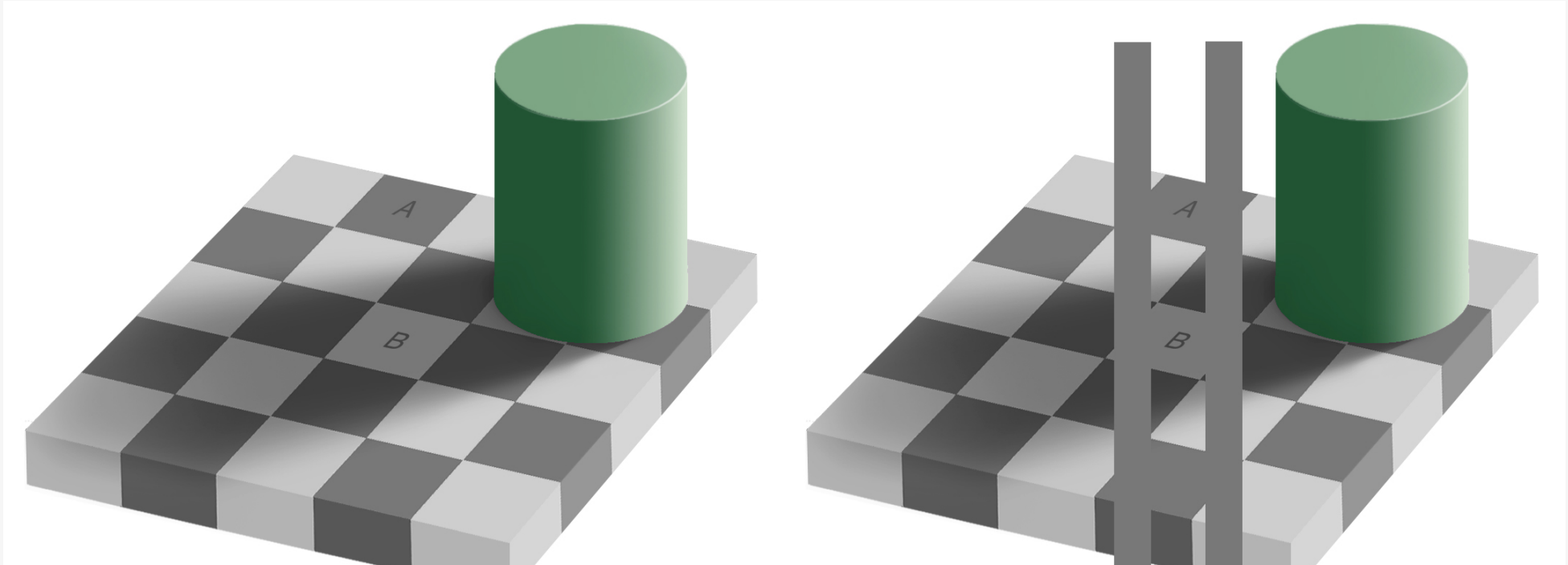


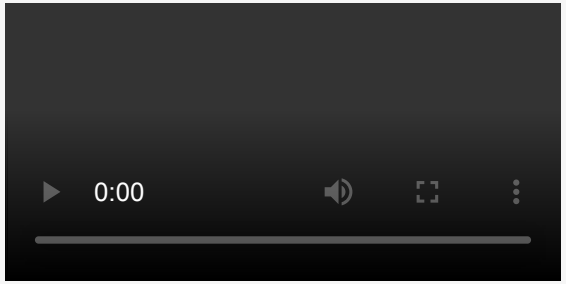



These are two perfectly
geometrical circles



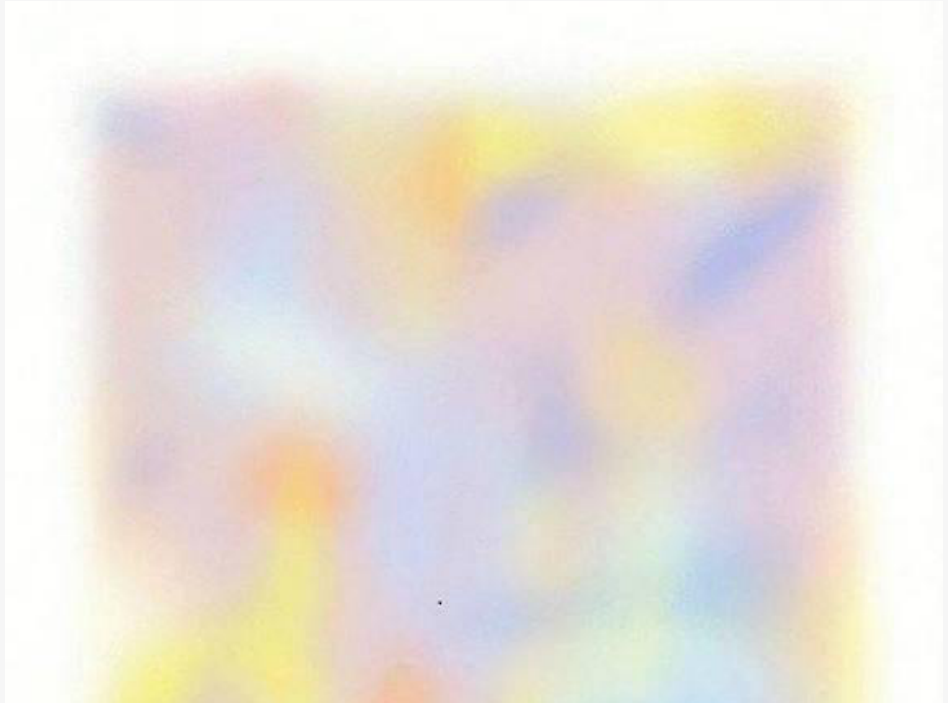

A
B

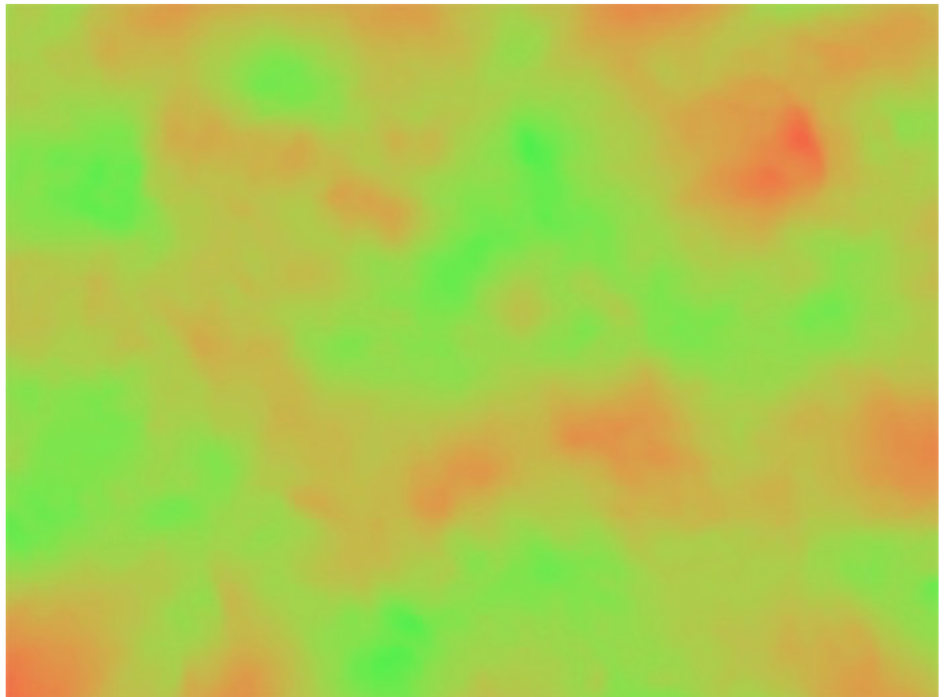

A
B
A
B

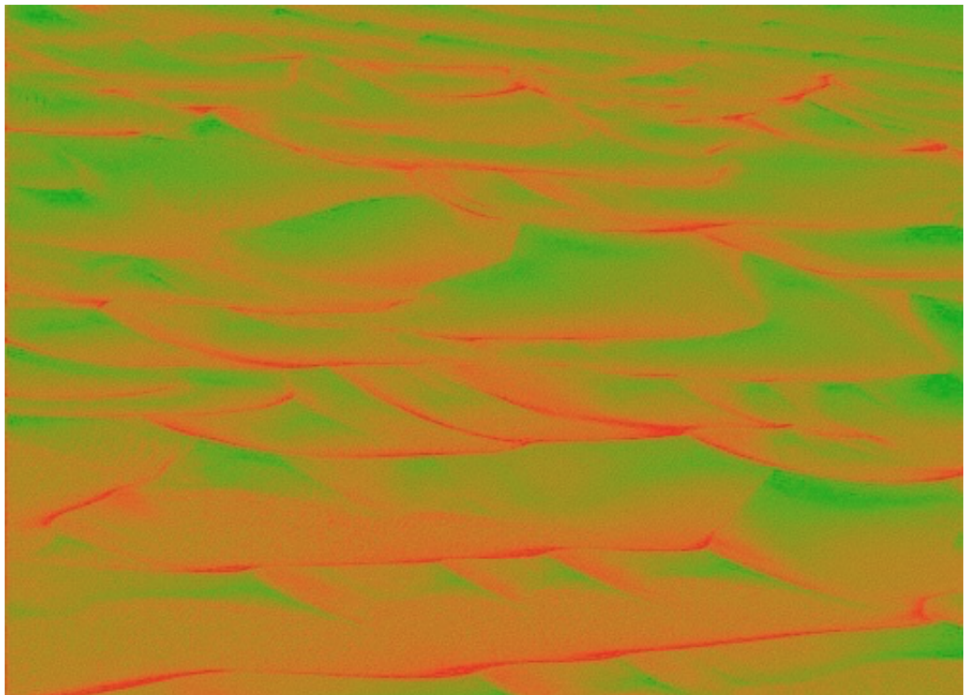


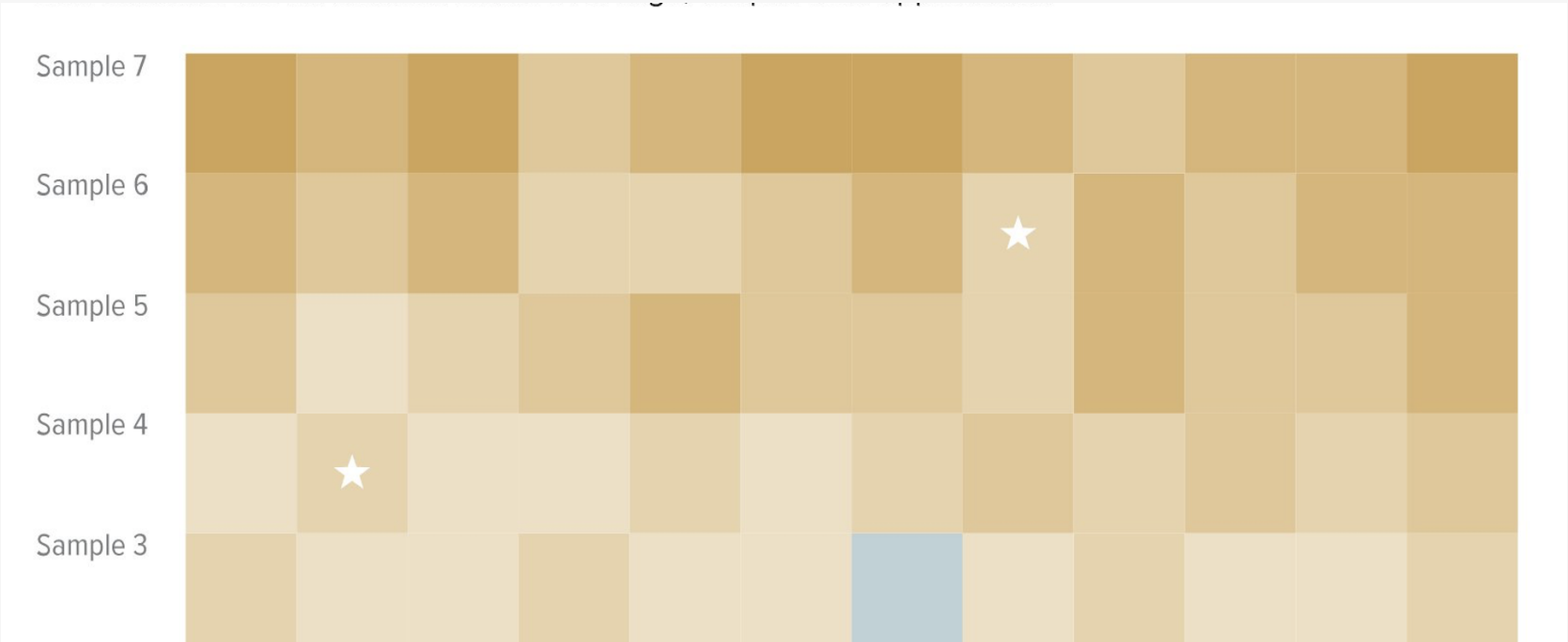
Edward Adelson

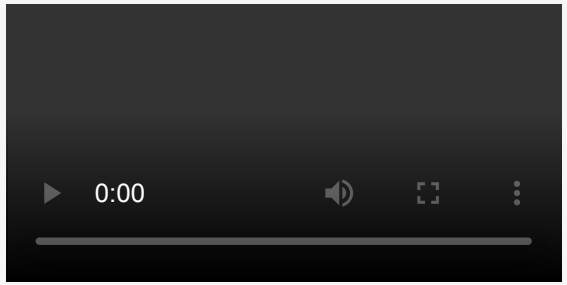
Luminance and Color

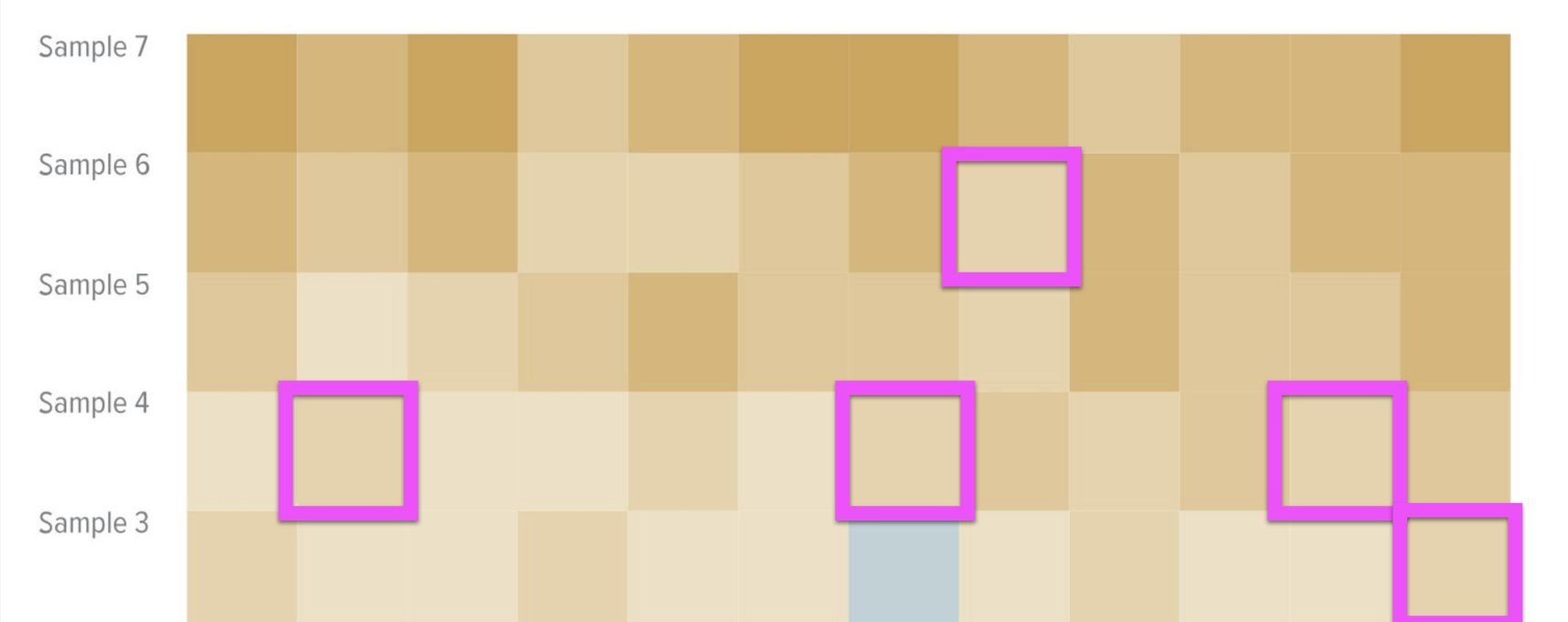


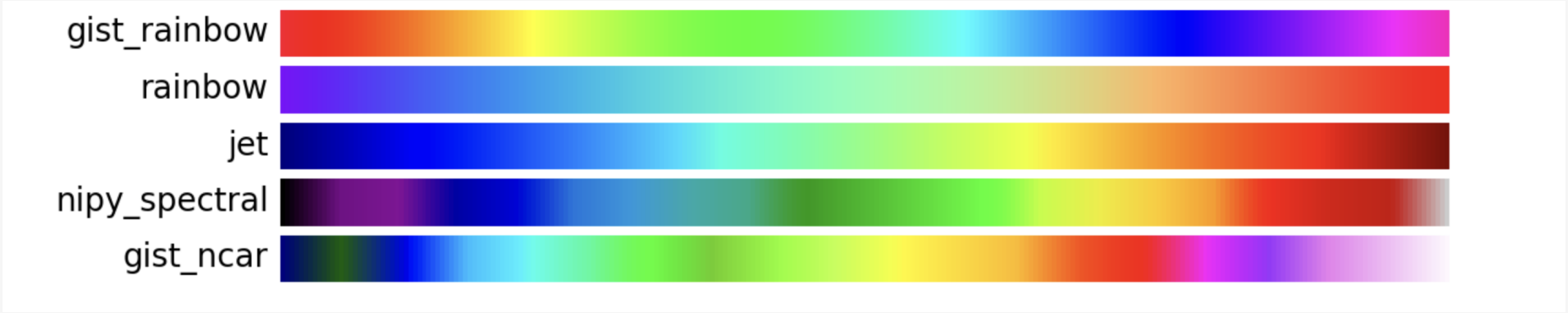


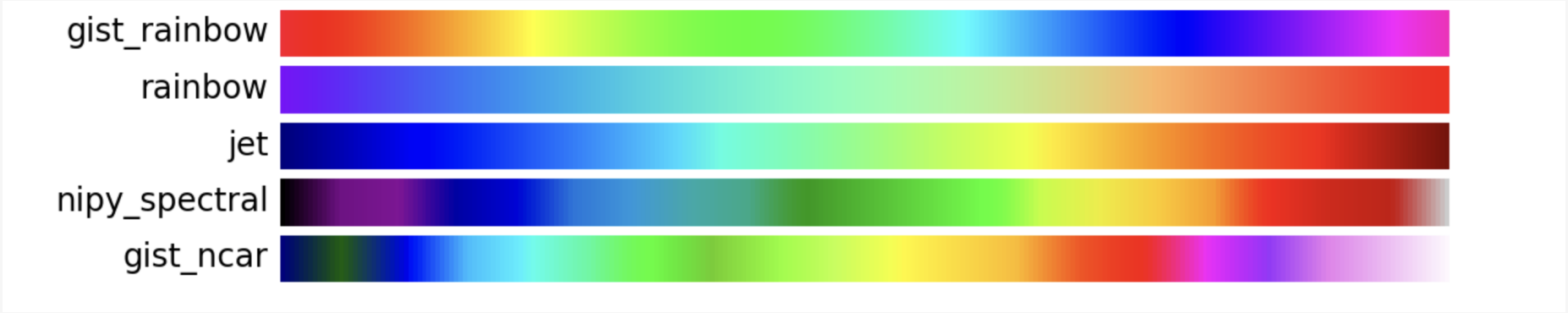



Sample 7
Sample 6
Sample 5
Sample 4
Sample 3

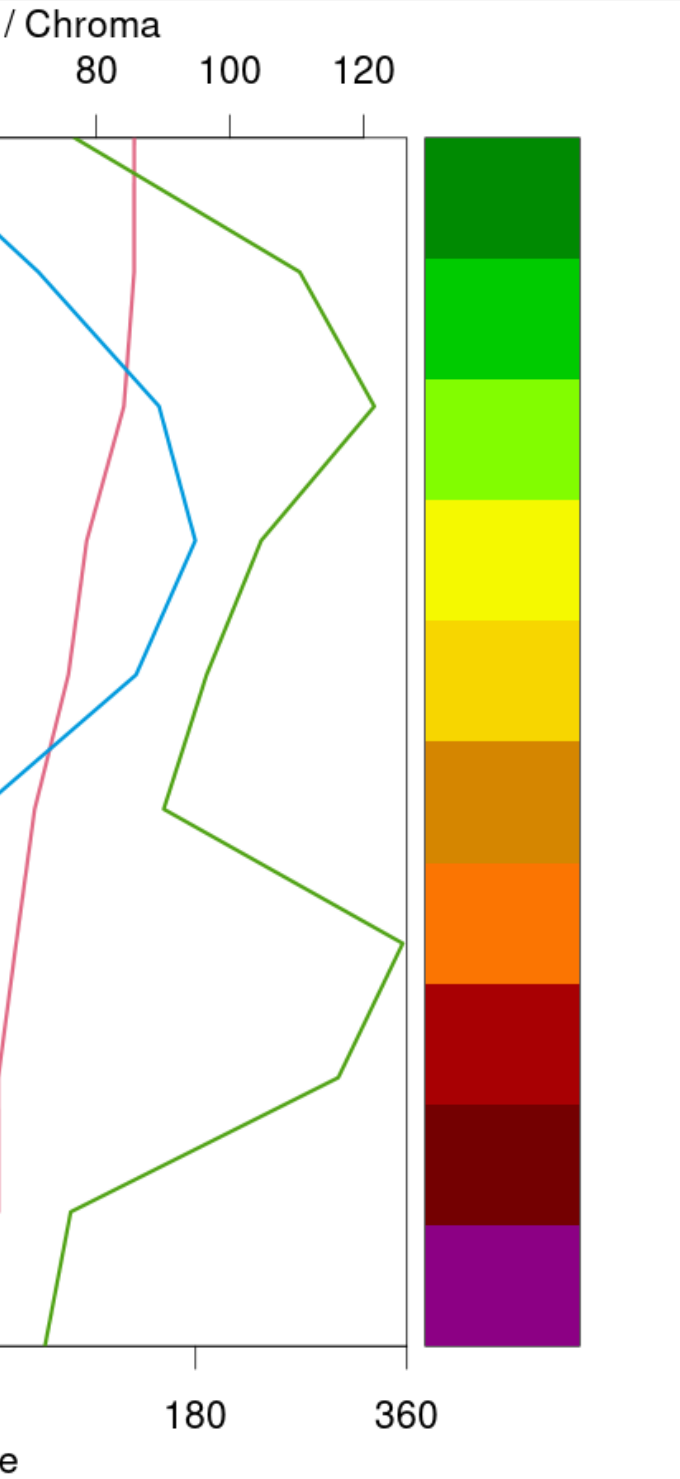

0:00

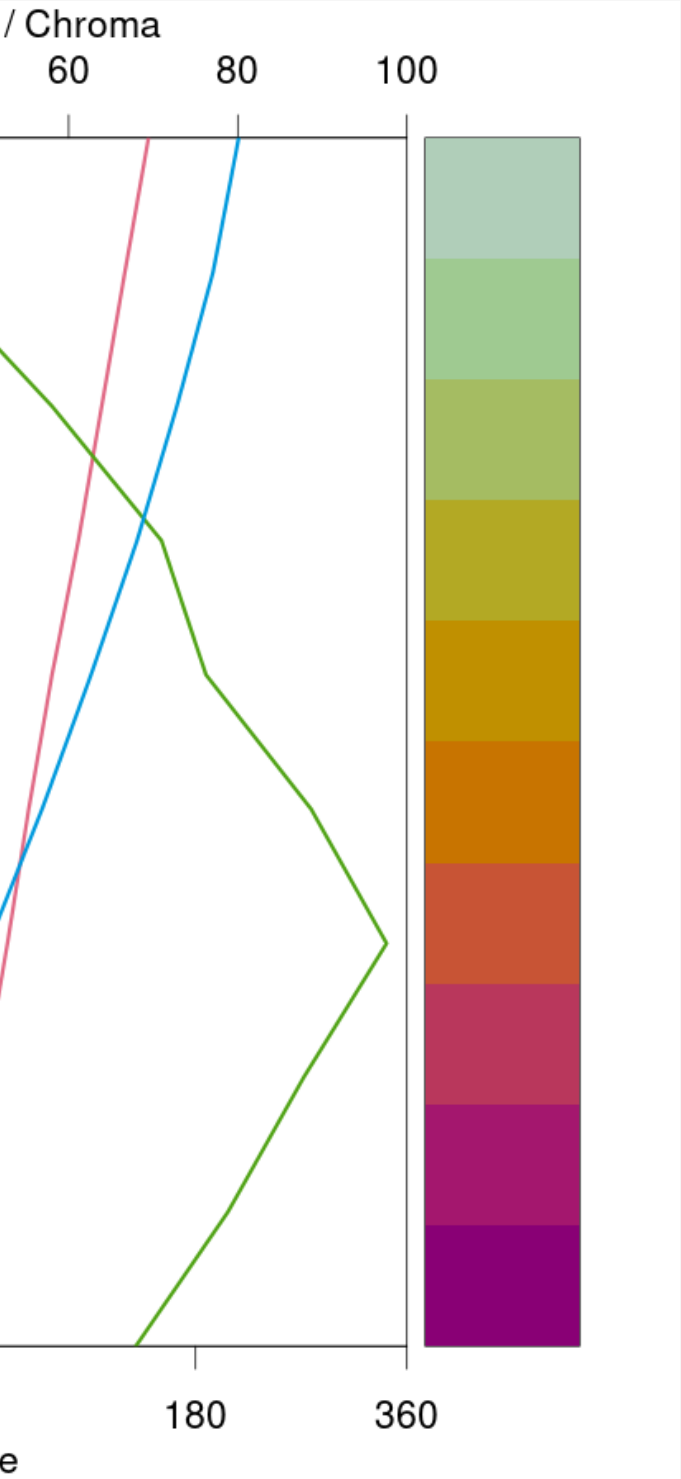

Sample 7
Sample 6
Sample 5
Sample 4
Sample 3

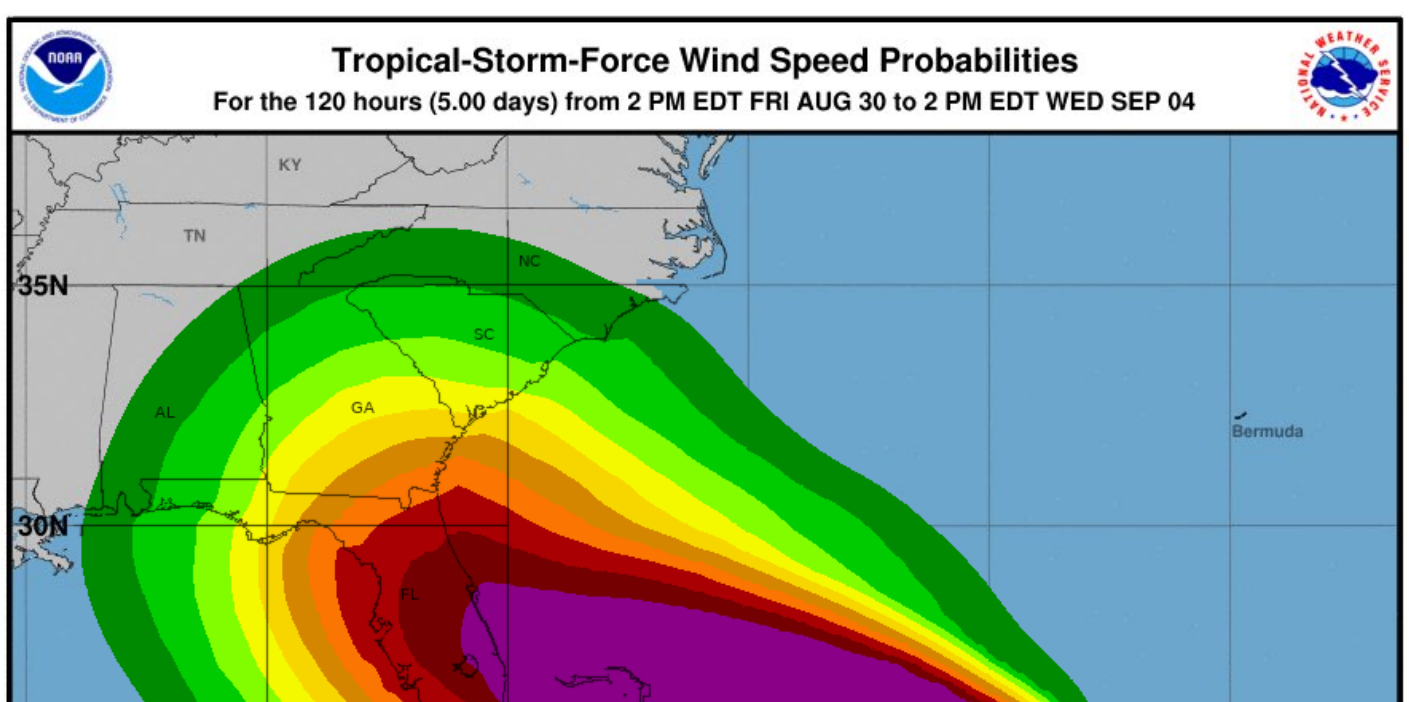

gist_rainbow
rainbow
jet
nipy_spectral
gist_ncar

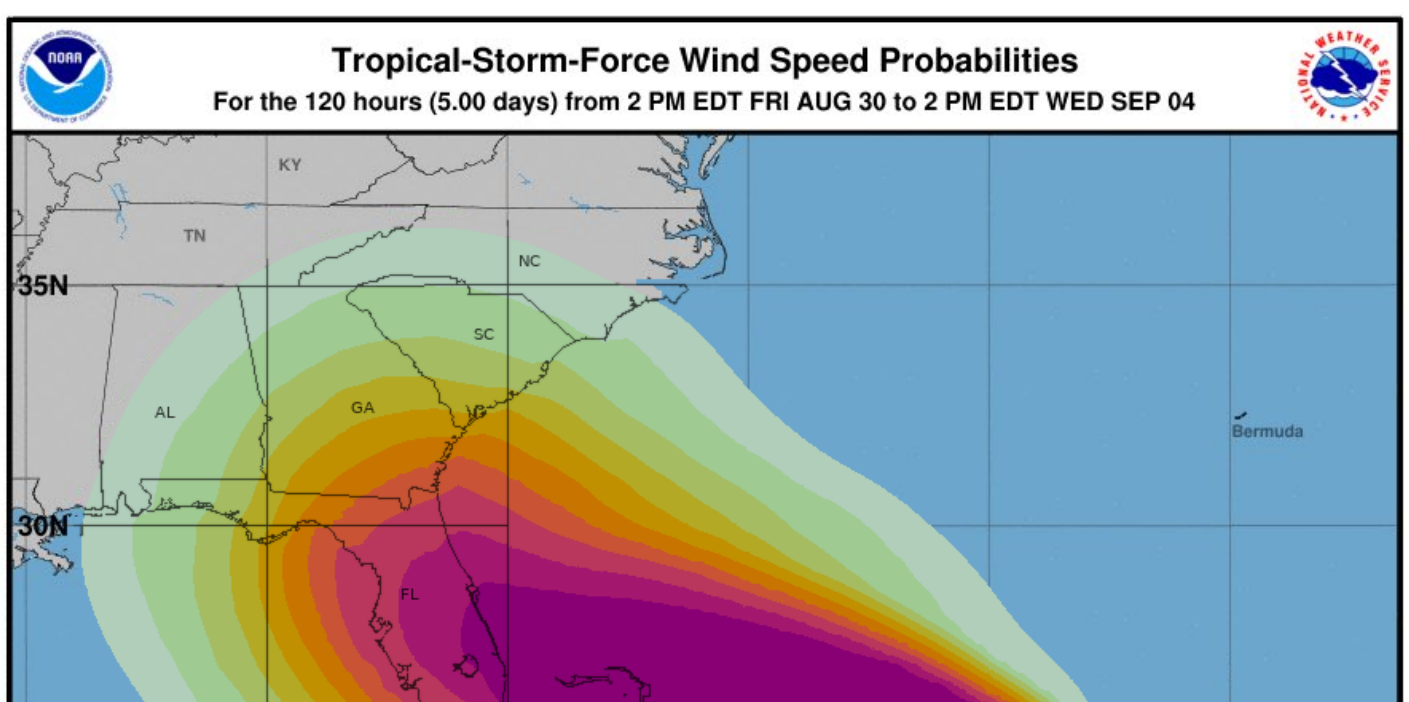

gist_rainbow
rainbow
jet
nipy_spectral
gist_ncar

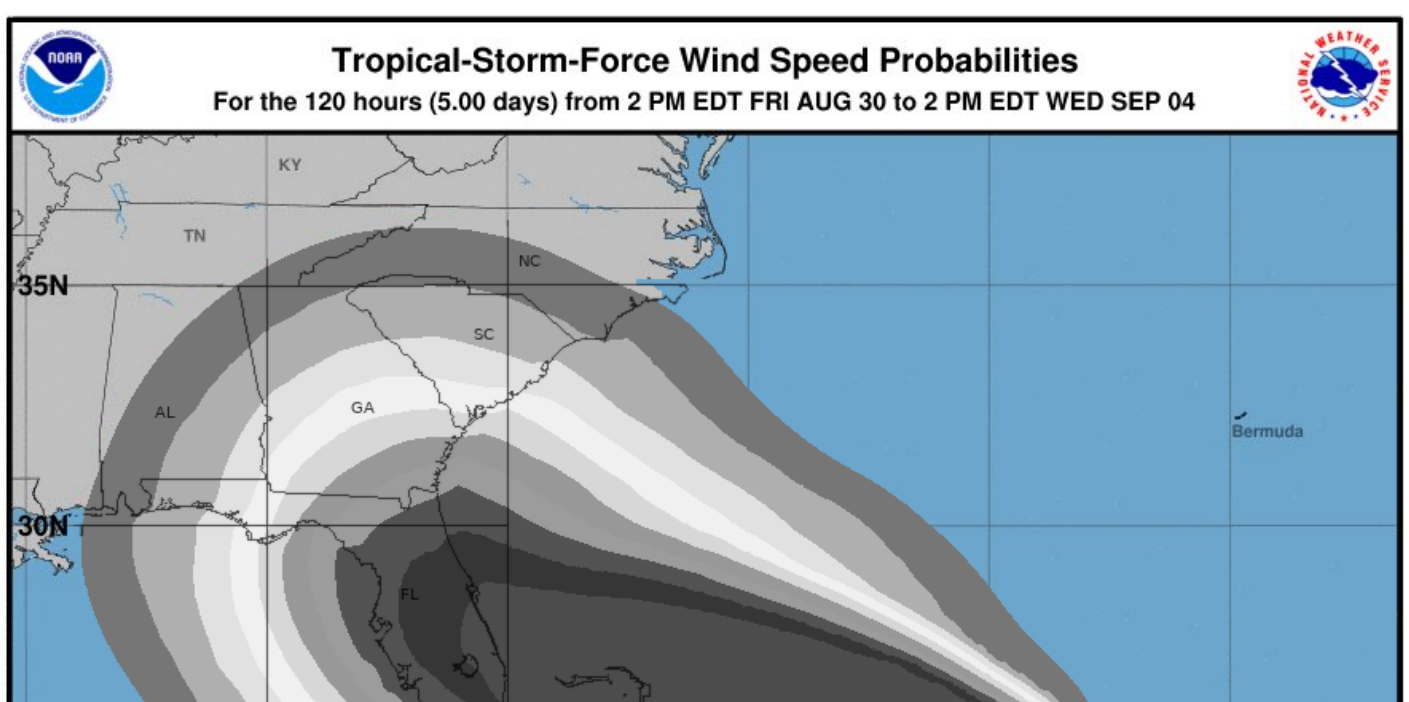

Perceptually Uniform Sequential colormaps
viridis
plasma
inferno
magma
cividis

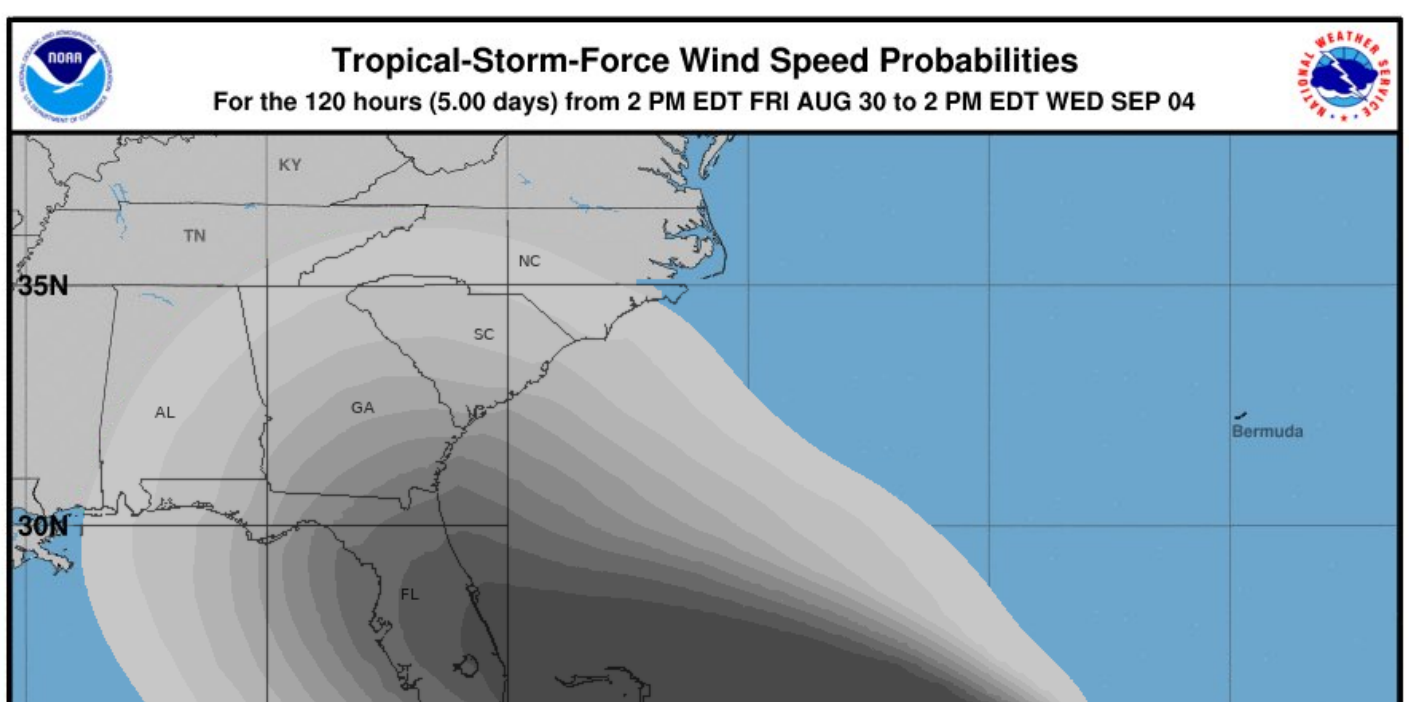

/ Chroma
80
100
120
180
360
e


/ Chroma
60
80
100
180
360
e


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
KY
TN
NC
35N
SC
AL
GA
Bermuda
30N
FL


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
KY
TN
NC
35N
SC
AL
GA
Bermuda
30N
FL


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
KY
TN
NC
35N
SC
AL
GA
Bermuda
30N
FL


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
KY
TN
NC
35N
SC
AL
GA
Bermuda
30N
FL