

BUAN 3500: Data Visualization and Descriptive Analytics in Business

Data Visualization

Lauren M. Nelsen, Ph.D.

University of Colorado Colorado Springs

September 4, 2025

Reminders

Reminder: Follow the instructions in the syllabus to download Tableau if you haven't already!

(You can download Tableau for free as a student.)

Reminder: Be sure you're reading the assigned sections in the textbook and the assigned pages in Tufte.

Tufte

- Tufte: pg. 13-51: Graphical Excellence
- Tufte: pg. 53-77: Graphical Integrity

Tufte – Graphical Excellence

Graphical Excellence:

“the efficient communication of complex quantitative ideas”

Thoughts/Comments/Questions?

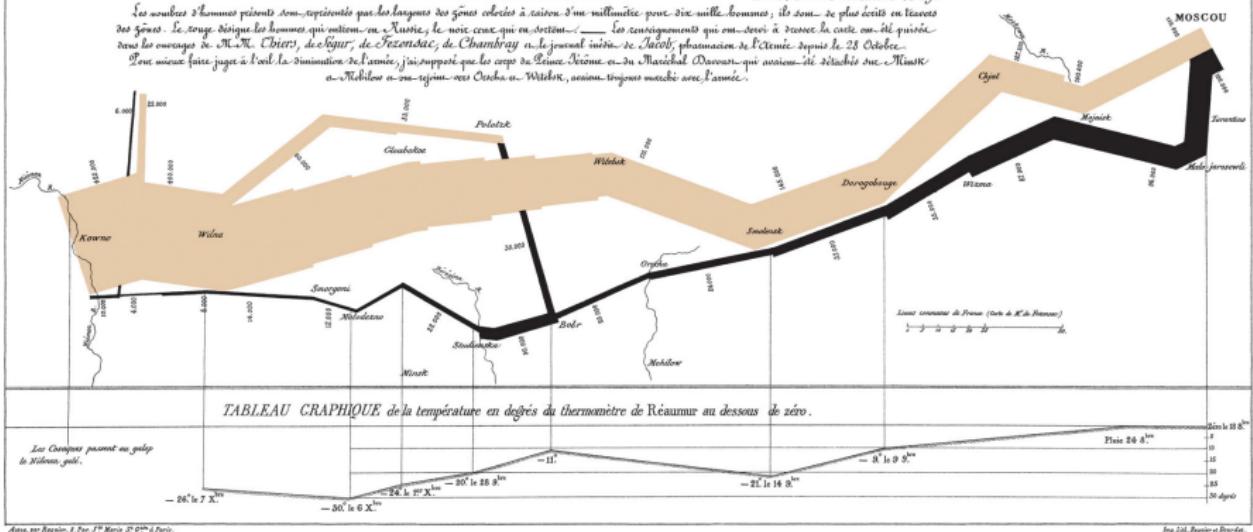
Tufte – Graphical Excellence

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

Brassée par M. Minard, Ingénieur Général des Ponts et Chaussées au Ministère.

Paris, le 20 Novembre 1869.

Les nombre d'hommes perdus sont... évidemment que les longueurs des zones colorées à sauter d'une millième pour dix mille hommes; ils sont... de plus écrits ou levés des zones... Le rouge désigne les hommes qui restent en Russie, le noir ceux qui en sortent... Les renseignements qui me servent à dresser la carte ont été pris dans les ouvrages de M. M. Chabri, de Léger, de Fresenac, de Chambray, le Journal intime de Napoléon, pharmacien de l'Armée depuis le 28 Octobre 1812 jusqu'à l'arrivée à l'armée russe, j'ai supposé que les corps de l'Armée étaient dispersés depuis le 28 Octobre à Nébilia et que l'armée vers Orelia et Nébilia, n'avaient toujours marché avec l'armée.

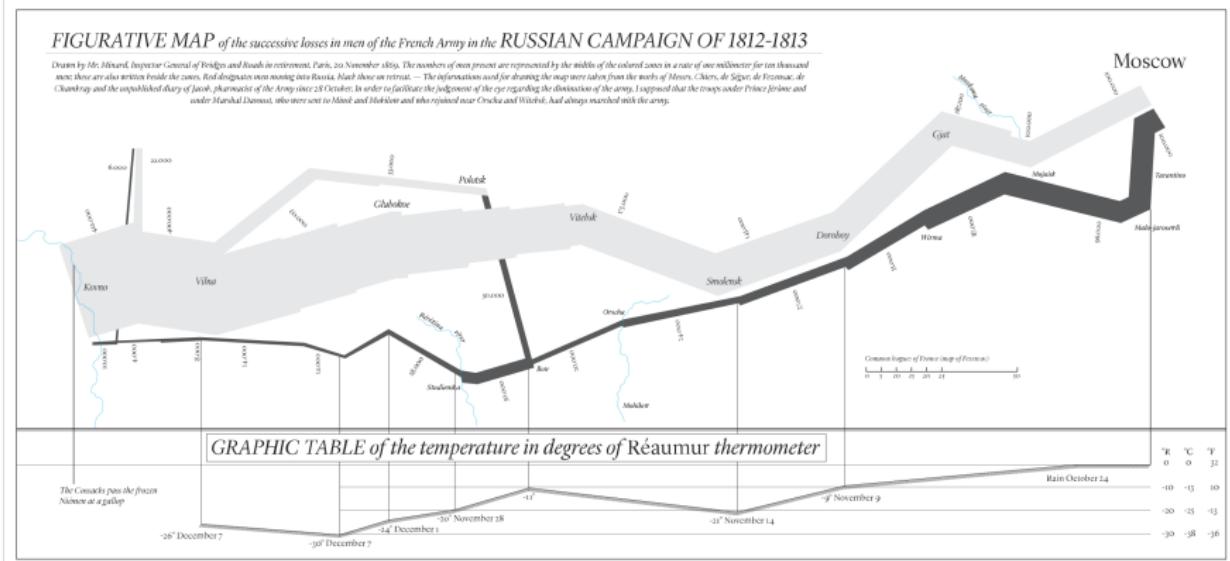


Arch. per Register, E. Par. 1^{re} Marq. D² 2^{me} à Paris.

Imp. Lib. Pigeot et Deshayes.

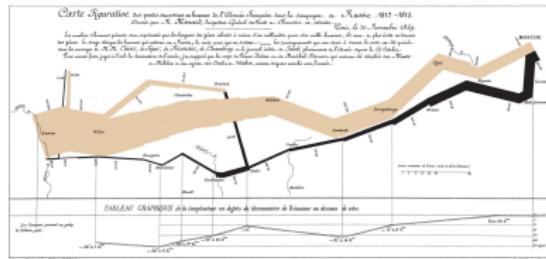
Charles Joseph Minard's portrait of Napoleon's 1812 Russian Campaign
(English translation and redrawing is in the book)

Tufte – Graphical Excellence



A modern redrawing of Minard's graphic

Tufte – Graphical Excellence



Charles Joseph Minard's portal of Napoleon's 1812 Russian Campaign
Tufte says this "may well be the best statistical graphic ever drawn". Why?

- It tells a story
 - Shows six variables in a 2-D surface

Tufte – Graphical Excellence

Let's look at the map shown at the link below. A map is a data visualization – how much information (and what kinds) does it portray?

<https://bit.ly/3BmWnvR>

Tufte – Graphical Excellence

Tufte's Principles of Graphical Excellence (p. 51):

- Graphical excellence is the well-designed presentation of interesting data – a matter of substance, of statistics, and of design.
- Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.
- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- Graphical excellence is nearly always multivariate.
- Graphical excellence requires telling the truth about the data.

Tufte – Graphical Integrity

“Graphical excellence begins with telling the truth about the data.”

Tufte (p. 53)

Graphical Integrity

YouTube Video:

How to spot a misleading graph - Lea Gaslowitz

<https://youtu.be/E91bGT9BjYk>

Graphical Integrity

Let's look at several graphics that are lying or are misleading.
(You might have seen some of these before!)

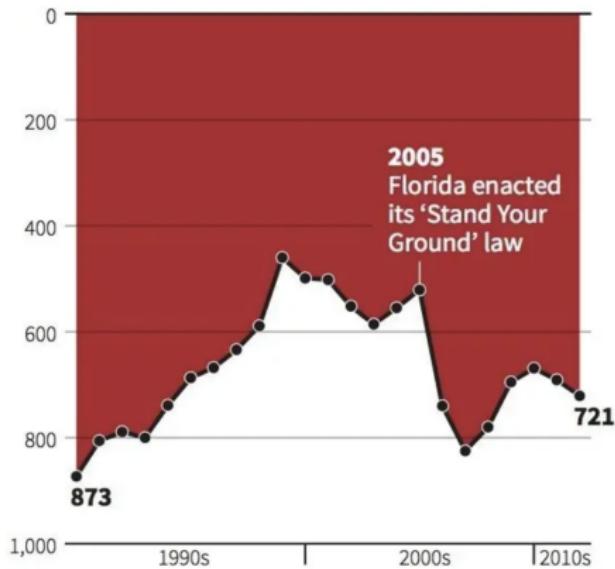
The images on the following slides are taken from the following websites:

- <https://bit.ly/3rYJiYG>
- <https://bit.ly/3YwVj3B>
- <https://bit.ly/3DS7ILf>

Graphical Integrity

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

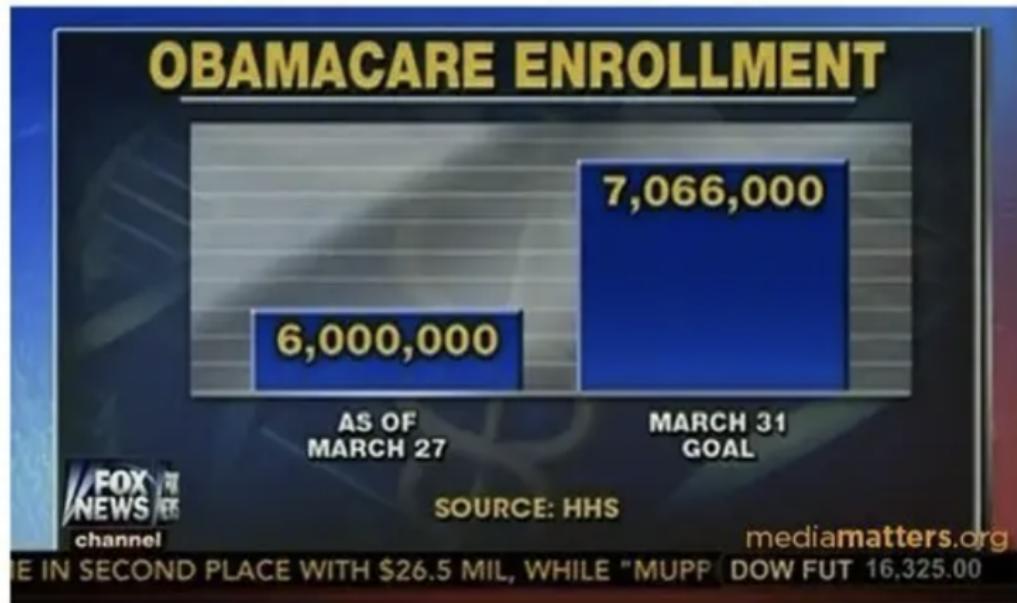
C. Chan 16/02/2014

Reuters / Via reddit.com

REUTERS

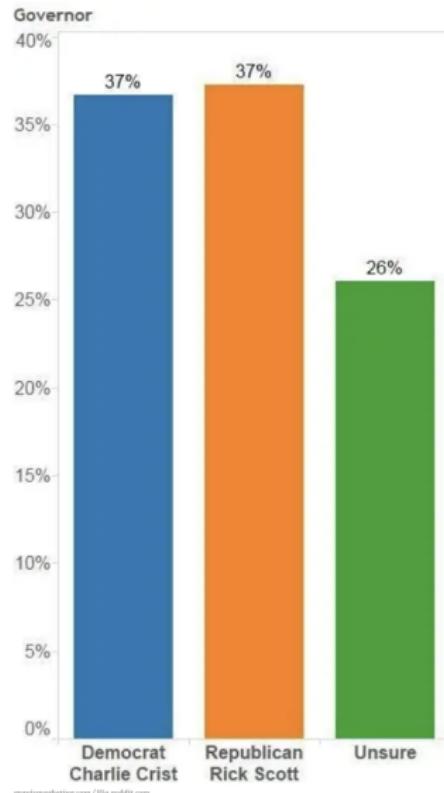
Graphical Integrity

Obamacare Signups, According to Fox News

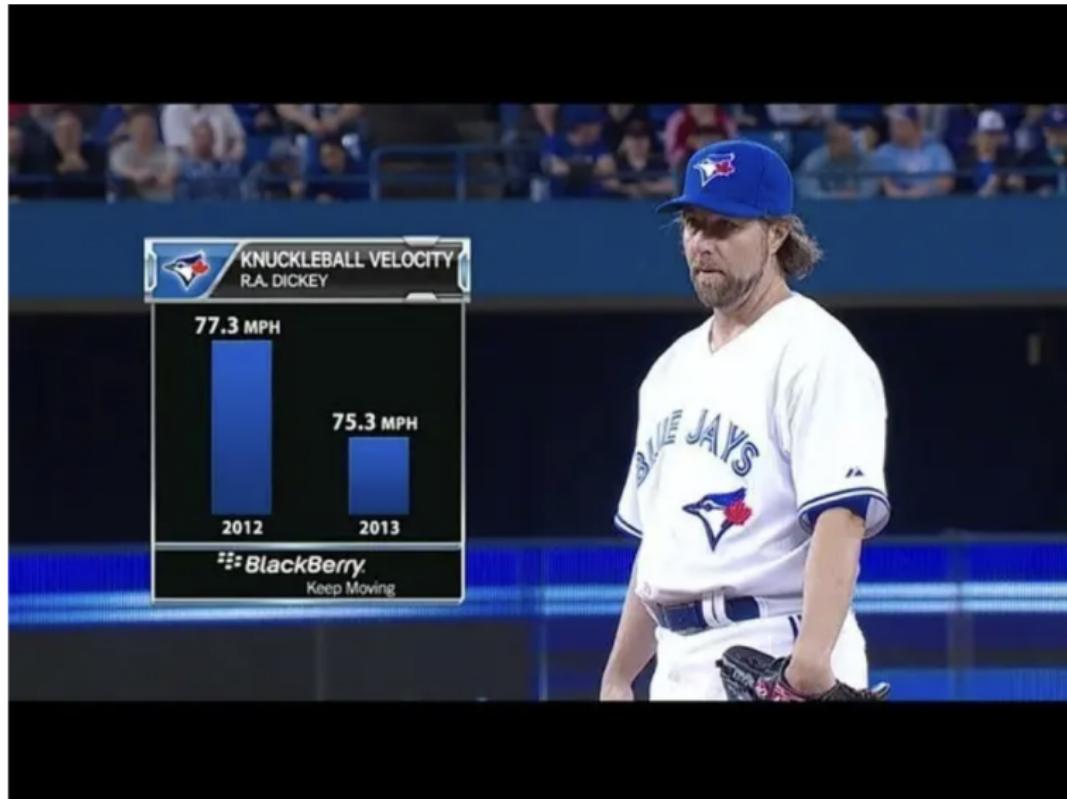


Source: Media Matters of America

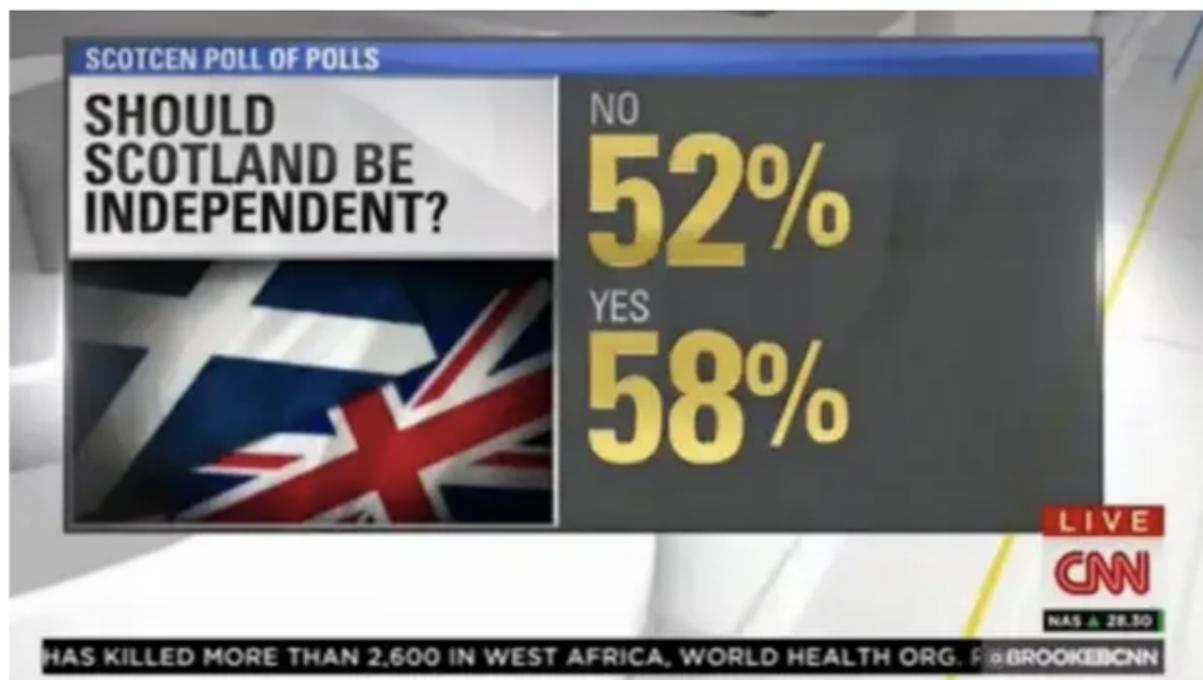
Graphical Integrity



Graphical Integrity



Graphical Integrity



Graphical Integrity





Graphical Integrity



Graphical Integrity

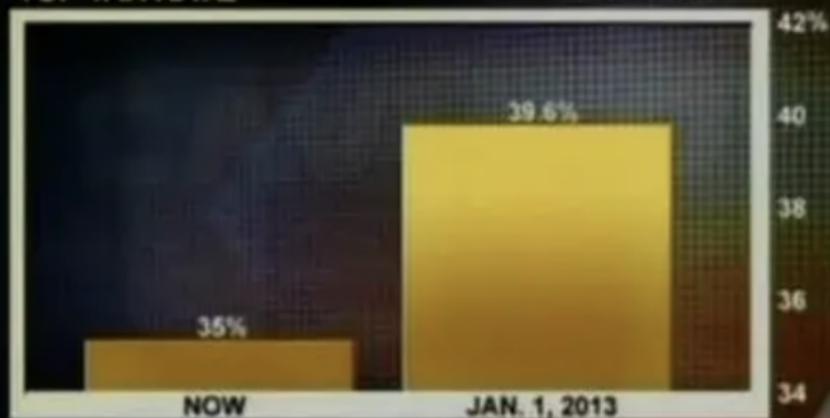
The figure is a screenshot of a web-based map titled "US Wildfire Activity Web Map" from ArcGIS. The map displays wildfire activity across North America, with a significant concentration of fires in the western United States and Canada. Key features include:

- Geography:** Shows the outlines of the United States and Canada, along with major lakes like Great Bear Lake, Slave Lake, Athabasca, Lake Superior, and Hudson Bay.
- Cities:** Labels major cities such as Los Angeles, San Francisco, Seattle, Vancouver, Edmonton, Calgary, Chicago, Detroit, St. Louis, Atlanta, and Miami.
- Wildfires:** A dense pattern of red and orange dots representing wildfire locations, with higher concentrations in the western US and southern Canada.
- Basemap:** A light green base map showing terrain and water bodies.
- Legend:** Located on the left side, it includes icons for Add, Layers, Tables, Basemap, Charts, Legend, Bookmarks, Save and open, Map properties, Share map, Create app, Print, and Information.
- Search Bar:** A search bar with a magnifying glass icon is located in the bottom right corner.
- Navigation:** Includes standard map controls for zooming in and out (+, -), panning, and orientation.

Graphical Integrity

IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

WITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T

DOW 13008.68 ▼ 64.33

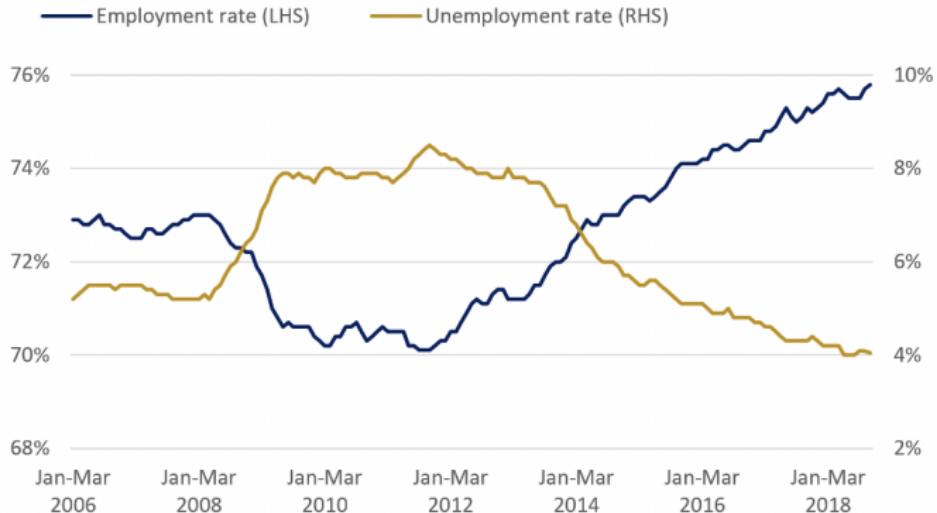
S&P 1379.32 ▼ 5.98

NASDAQ 2939.52 ▼ 6.32

Graphical Integrity

Employment and unemployment rates

UK, seasonally adjusted, January to March 2006 to September to November 2018

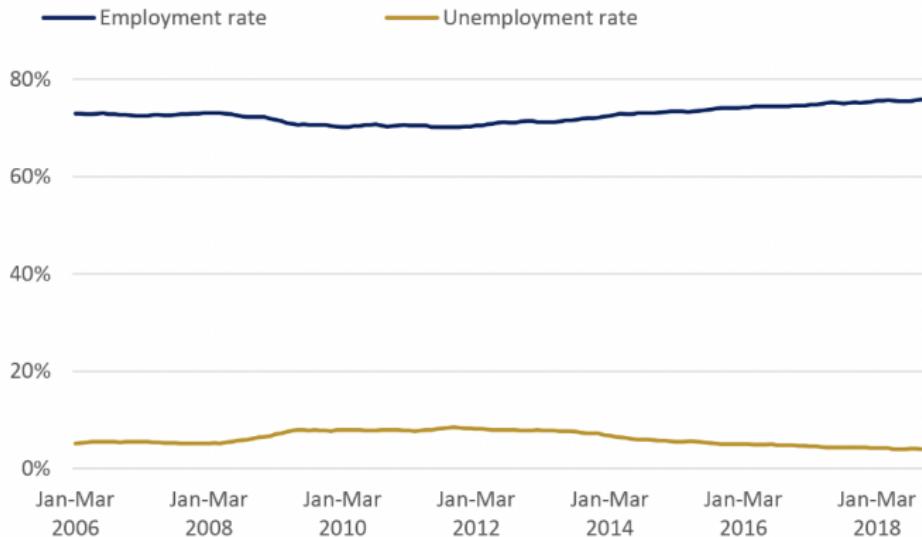


Source: [Labour market economic commentary: January 2019, ONS](#)

Graphical Integrity

Employment and unemployment rates

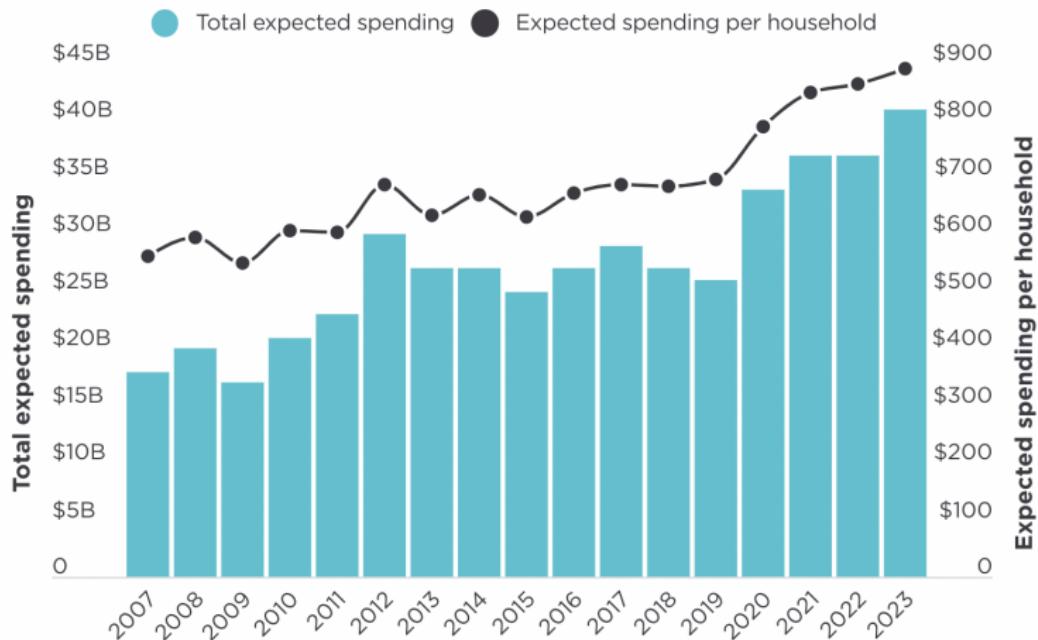
UK, seasonally adjusted, January to March 2006 to September to November 2018



Source: [Labour market economic commentary: January 2019, ONS](#)

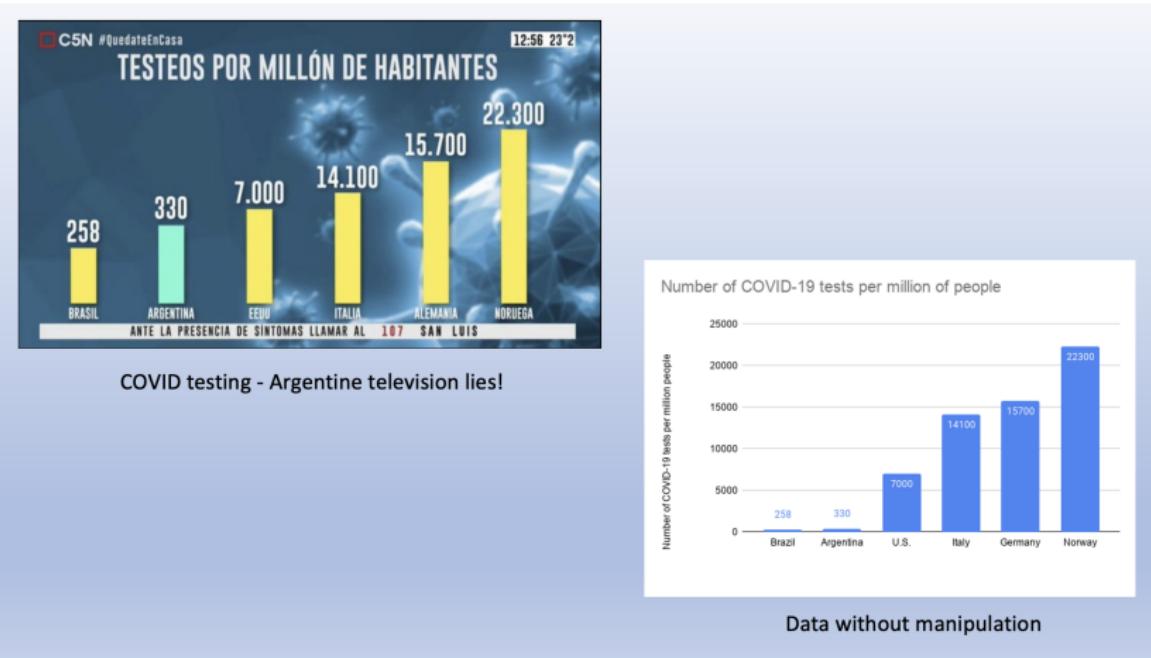
Graphical Integrity

Planned back-to-school spending



Source: NRF's Annual 2023 Back-to-School Survey, conducted by Prosper Insights & Analytics

Graphical Integrity



(Image from Dave Leupp)

Graphical Integrity



Later bars are arbitrary heights not truly reflecting the data values.



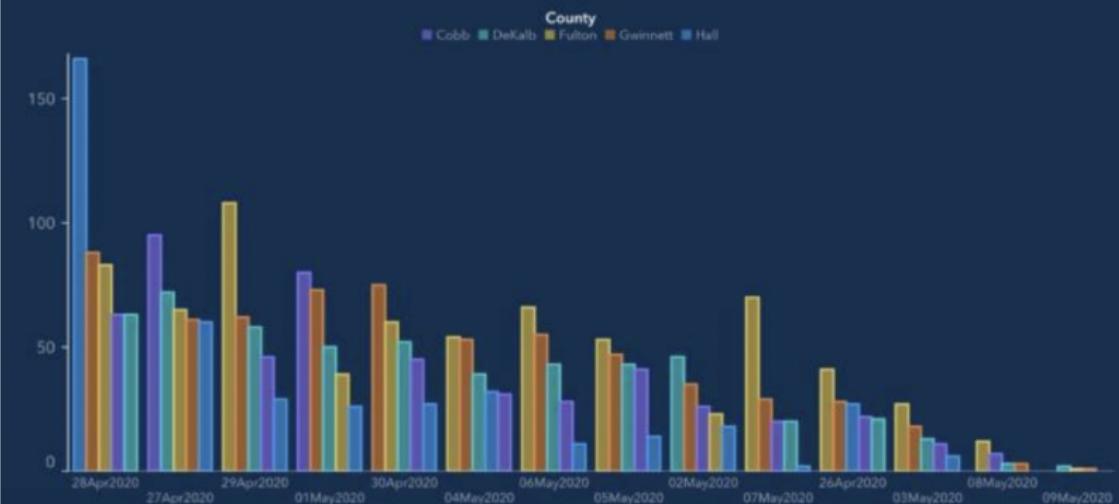
Data without manipulation (no flattening)

(Image from Dave Leupp)

Graphical Integrity

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.

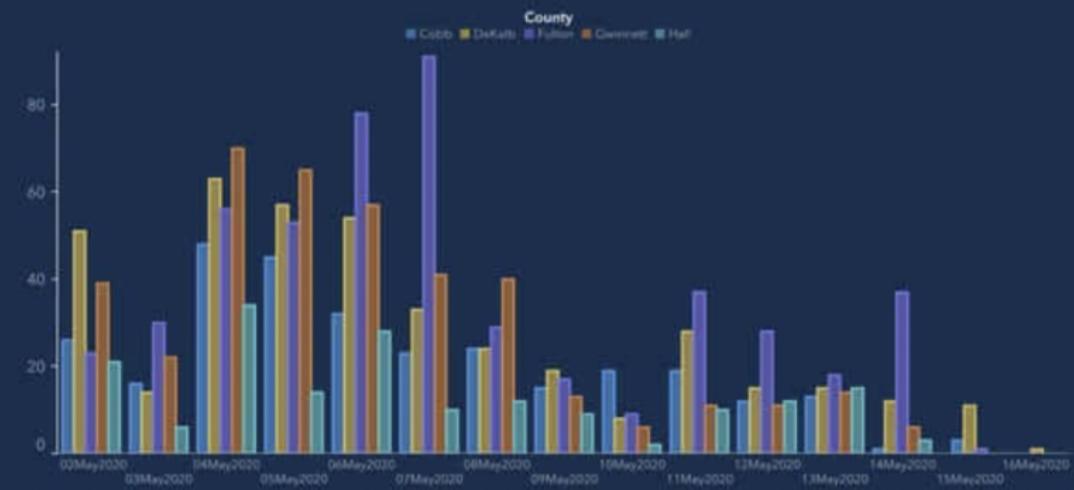


Source: Vox

Graphical Integrity

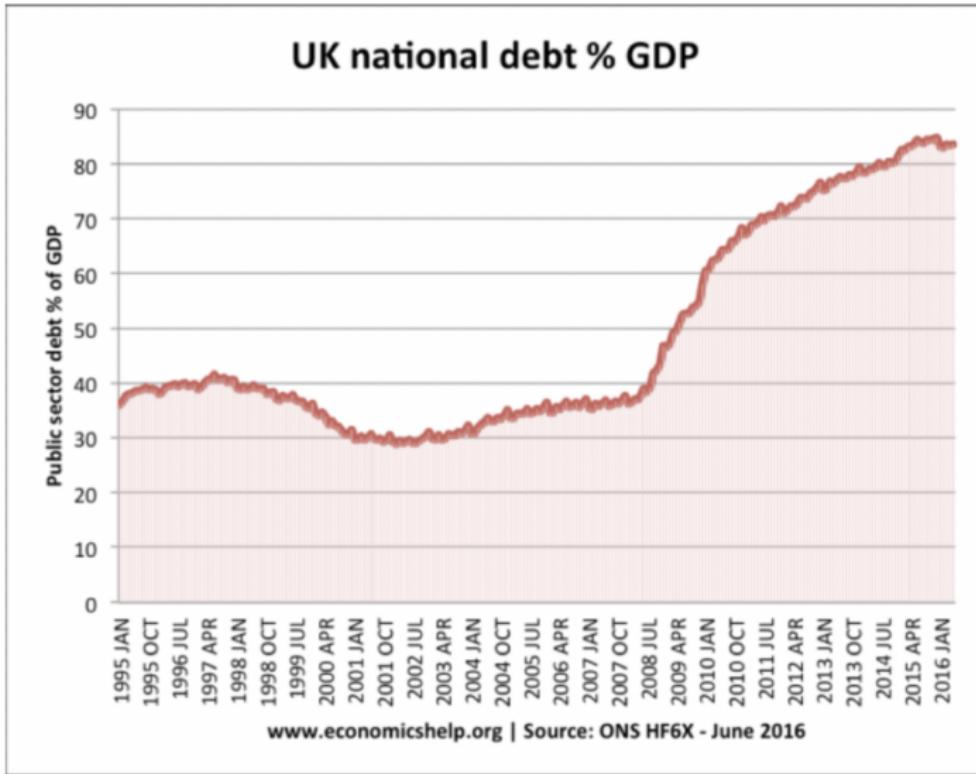
Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.



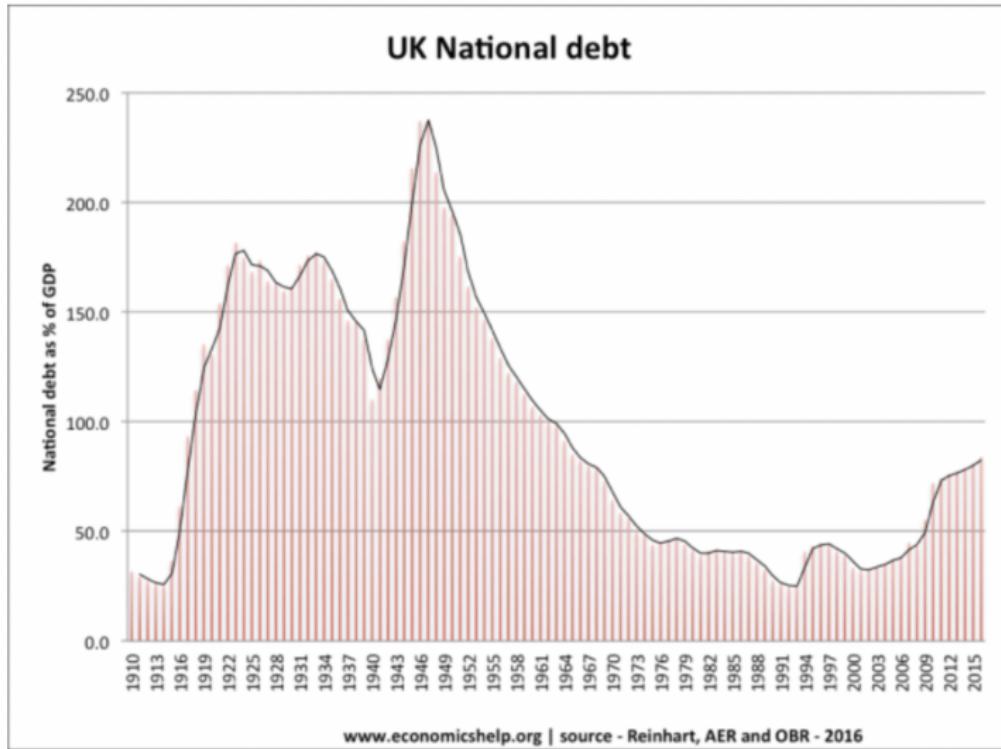
Source: Business Insider

Graphical Integrity



Source: www.economicshelp.org

Graphical Integrity



Source: www.economicshelp.org

Graphical Integrity

- Data graphics can either intentionally or unintentionally distort the data they represent.
- Tufte gives many other examples that you should carefully study!
- Always be aware of this danger in your own work!

“Graphical excellence begins with telling the truth about the data.”

Guidelines

Primary Guidelines for Data Visualization:¹

- Show the data
- Reduce the clutter
- Integrate the graphics and text
- Avoid the spaghetti chart
- Start with gray

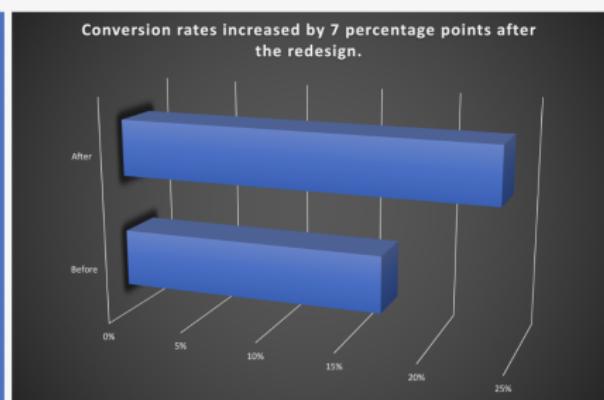
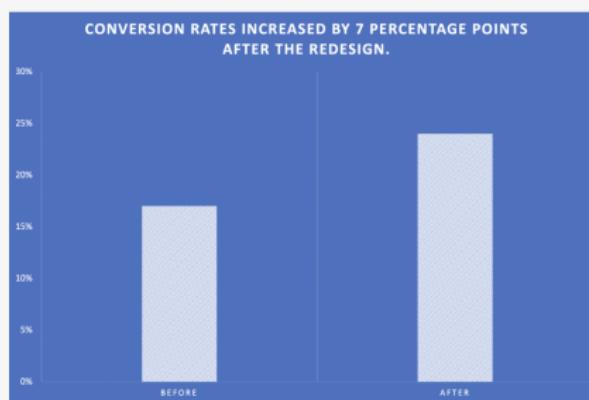
¹ "Better Data Visualizations, A Guide for Scholars, Researchers, and Wonks" by Jonathan Schwabish

Guidelines

- Show the data
 - (This does not mean we show ALL the data all the time!)

Guidelines

- Reduce the clutter



Bad: Two examples of Excel's chart styles. Funky colors, fonts, 3D effects, gradients, shadows, and textures don't add informational value to your visual — in fact, they're just distracting.

Guidelines

- Integrate the graphics and text

"The annotation layer is the most important thing we do...otherwise it's a case of 'here it is, you go figure it out.'"

-Amanda Cox, Data Editor at the *New York Times*

Guidelines

- Avoid the spaghetti chart

<https://www.data-to-viz.com/caveat/spaghetti.html>

Sometimes we face the challenge of including lots of data in a single graph, but we don't need to try to pack everything into a single graph.

Guidelines

- Start with gray
 - <https://bit.ly/3shWs2R>

Overview of Data Visualization

Effective Design Techniques

- **data-ink ratio:** measures the proportion of what Tufte terms “data-ink” to the total amount of ink used in a table or chart
(non-data ink is ink used in a table or chart that serves no useful purpose in conveying the data to the audience)

Overview of Data Visualization

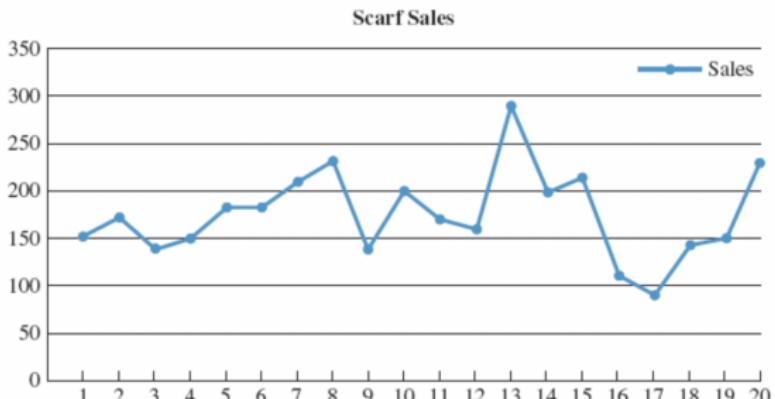
Table 3.1 Example of a Low Data-Ink Ratio Table

Scarf Sales			
Day	Sales (units)	Day	Sales (units)
1	150	11	170
2	170	12	160
3	140	13	290
4	150	14	200
5	180	15	210
6	180	16	110
7	210	17	90
8	230	18	140



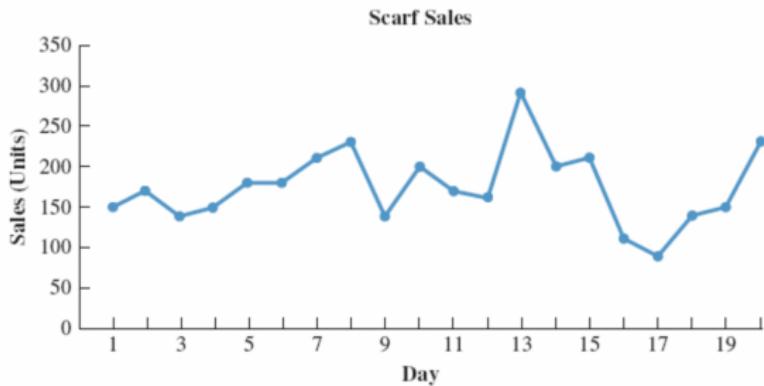
Overview of Data Visualization

Figure 3.3 Example of a Low Data-Ink Ratio Chart



Overview of Data Visualization

Figure 3.4 Increasing the Data-Ink Ratio by Adding Labels to Axes and Removing Unnecessary Lines and Labels



Overview of Data Visualization

Table 3.2 Increasing the Data-Ink Ratio by Removing Unnecessary Gridlines

Scarf Sales			
Day	Sales (units)	Day	Sales (units)
1	150	11	170
2	170	12	160
3	140	13	290
4	150	14	200
5	180	15	210
6	180	16	110
7	210	17	90
8	230	18	140



Tables

Our textbook says: “**The first decision in displaying data is whether or a table or a chart will be more effective.**”

Questions:

- Is Minard's depiction of Napoleon's invasion of Russia a table or a chart?
- Is a map a table or a chart?

Displaying data is a creative process. Tables and charts are not the only options!

Tables

When tables might be a good choice:

- Tufte says that tables usually outperform graphics when depicting small data sets of 20 numbers or less. (p. 56)
- (Also see Tufte pg. 178-180.)

Tables should be used when:

- the reader needs to refer to specific numerical values
- the reader needs to make precise comparisons between different values and not just relative comparisons
- the values being displayed have different units or very different magnitudes

Tables

Table design:

- Data-ink ratio
 - Don't clutter with unnecessary grid lines, colors, fonts, etc.

Figure 3.7 from the text shows variations of the same data table:

Figure 3.7 Comparing Different Table Designs

Design A:

	Month						Total
	1	2	3	4	5	6	
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design B:

	Month						Total
	1	2	3	4	5	6	
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design C:



	Month						Total
	1	2	3	4	5	6	
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design D:

	Month						Total
	1	2	3	4	5	6	
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Tables

Crosstabulation (Pivot Tables): provides a tabular summary of data for two variables

- a way to reorganize and summarize data in a table (or database in general)
- very broad and flexible uses

(You should have used Pivot Tables in Excel in previous classes, so this should be review.)

Tables

PivotTables in Excel

Example

The data file “restaurant” contains data from Zagat’s Restaurant Review. Data on the quality rating, meal price, and the usual wait time for a table during peak hours were collected for a sample of 300 Los Angeles area restaurants.

- What if we want to know the number of restaurants in each combination of Quality and Price?
 - We could manually count how many restaurants are rated Good with mean prices between \$10 and \$19 for example, and so on, and create a table like Table 3.7 in the book:

Table 3.7 Crosstabulation of Quality Rating and Meal Price for 300 Los Angeles Restaurants

Quality Rating	Meal Price				Total
	\$10–19	\$20–29	\$30–39	\$40–49	
Good	42	40	2	0	84
Very Good	34	64	46	6	150
Excellent	2	14	28	22	66
Total	78	118	76	28	300

Tables

Example

Restaurant Example Continued:

Let's use PivotTables!

Charts

Examples of commonly used charts to display and analyze data:

- Scatter charts (plots)
- Line charts
- Bar charts and column charts

You have done all of these before – your challenge in this course will be to apply good data visualization principles when you make them.

(Be sure to read this section in the textbook, because we will not work through everything together in class!)

Charts

Note: Excel default charts normally DO NOT follow good design principles.

Charts

Scatter charts:

Example

Open the file “Ajax” on Canvas. We’re going to make a scatter plot using some of this data. Let’s think about how we can improve on the default chart that Excel displays so that the data stands out.

Charts

Line Charts:

Example

Open the file “KirklandRegional” on Canvas that shows regional sales data by month for air compressors at Kirkland Industries. We’re going to investigate some of the charts we can make with this data. (These steps are in the textbook if you want to follow along.)

Charts

Bar Charts and Column Charts:

Example

Open the file “AccountsManaged” on Canvas. We’re going to create a sorted bar chart with data labels for this data.

Charts

Pie Charts:

For the last example, we could have made a pie chart:

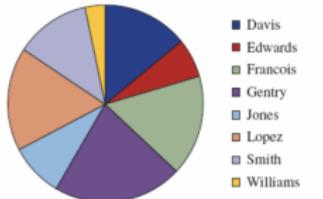
Figure 3.25

Bar Chart with Data Labels for Accounts Managed Data



Figure 3.26

Pie Chart of Accounts Managed



Question: What thoughts do you have when you compare the figures above?

Charts

Pie Charts:

Not good!

- It's harder for our brains to compare the size of the wedges.
(Comparing the lengths of the bars is much easier!)
- It's recommended to not use pie charts in most situations.
- Bar charts are often better for comparing categorical data.

Charts

3-D Charts:

Also usually bad!

- Excel makes it easy to create 3-D bar, line, pie, and other charts.
BUT in most cases this effect adds unnecessary detail that does not help explain the data.
 - The third dimension is not data – so it just adds noise (chartjunk)

Charts

3-D Charts:

Just because we **can** create a certain in Excel does not mean we **should!**

Example

Open the file “AutoProduction” in Canvas. Let’s see how we can create bad pie charts and 3-D charts in Excel (and why they are not helpful ways to visualize the data).

Bubble charts and heat maps:

Definition

- A **bubble chart** is a graphical means of visualizing three variables in a two-dimensional graph
 - A **heat map** is a two-dimensional graphical representation of data that uses different shades of color to indicate the magnitude.
-
- Can be useful when used properly, but can take a lot of work to create.
 - (These might be easier to create in Tableau!)

Charts

Example

Bubble chart example: Open the file “billionaires” on Canvas. We’re going to create a bubble chart using Excel to further exam the data.

Example

Heat map example: Open the file “samestoresales” on Canvas. We’re going to create a heat map in Excel.

Charts

Note the “sparklines” in the previous example.

Sparklines is a term coined by Tufte for “data-intense, design-simple, word-sized graphics” (in his book, *Beautiful Evidence*)

Charts

“Additional Charts for Multiple Variables”: Be sure to read the coverage of these topics in the book. (We might not have time to talk about these together in class.)

- These are useful, but you need to be very careful in their design.
- (The reader needs to understand them!)

Charts

Example

Let's look at an example of what is called a "parallel coordinates" plot.
(There is a homework problem about this type of plot.)

<https://bit.ly/47IP54P>

This example uses something called Python to create the visualization.

Charts

PivotCharts in Excel: Excel pairs PivotCharts with PivotTables to summarize and analyze data with both a crosstabulation and charting.

Example

Open the “restaurant” file again. We’ll work through how to create a PivotChart together. (If we don’t have time for this in class, then work through this on your own following the steps in the textbook!)

Data Visualization in Tableau

Note: This part of the textbook is on pg. 141-156.

Tableau Desktop is a powerful visualization tool that is widely used for data visualization.



Data Visualization in Tableau

- Tableau can do far more than we will ask it to do.
- It is capable of connecting to a huge variety of data sources, from Excel files, to databases, to text files, to cloud services like Google Analytics and SalesForce

Data Visualization in Tableau

Tableau has some jargon that can be confusing for beginner users.

- “Measures”: are numeric variables
- “Dimensions”: are categorical variables

Tableau is already installed on classroom computers, but you should also have already downloaded a copy of Tableau Desktop onto your own computer.

(You can do this for free using our class activation key that is in the announcement on Canvas.)

Data Visualization in Tableau

Example

We're going to see how to create several visualizations in Tableau using the files listed below that you can find on Canvas:

- “NewHouseSalesUS”
- “electronics”
- “KirklandRegional”
- “AccountsManaged”
- “billionaires”
- “NYCityData”
- “global100”
- “WorldGDP2014”
- “AuditTime”
- “HomeSalesStacked”

Data Visualization in Tableau

Visual Analysis Best Practices

Simple Techniques for Making Every Data Visualization Useful and Beautiful

[https://www.tableau.com/sites/default/files/media/
whitepaper_visual-analysis-guidebook_0.pdf](https://www.tableau.com/sites/default/files/media/whitepaper_visual-analysis-guidebook_0.pdf)

Data Visualization in Tableau

Tableau is a lot of fun to use!

If you want to play around with it, here are some sources for data that you could use:

- Colorado Information Marketplace:
<https://data.colorado.gov/>
- UC Irvine Machine Learning Repository
<https://archive.ics.uci.edu/>
- Pew Research Center
<https://www.pewresearch.org/internet/datasets/>
- Buzzfeed News
<https://github.com/orgs/BuzzFeedNews/repositories?type=all>
- FiveThirtyEight
<https://data.fivethirtyeight.com/>
- U.S. Government Open Data
<https://data.gov/>