



**DATA VISUALIZATION,
#DATAVIZ, MATPLOTLIB,
SEABORN, BOKEH &
PLOTLY, EDWARD
TUFTE, PIE CHARTS**

LEARNING OBJECTIVES

- Make a decent plot with Matplotlib
- Use Seaborn for prettier plots
- Use Seaborn for data insights
- What makes a good #dataviz
- Bokeh, Plot.ly

PRE WORK & REVIEW

LAST LESSON REVIEW

- Correlation vs Causation
- Linear regression
- R^2 , P-value, ...
- Confounders / Causality

LAST SESSION

QUESTIONS REGARDING LAST CLASS

- When does correlation imply causation?
- What is R-squared
- Why use Adj. R-squared?
- P-Value?
- 4 Conditions for linear regression?
- Condition for causality?

TODAY

THE AMAZING WORLD OF DATAVIZ

TODAY

- Matplotlib
- Seaborn

Then

- Bokeh
- Plotly
- Tableau

AND A SURPRISE!

MATPLOTLIB

MATPLOTLIB

- Matplotlib is a plotting library for the Python
- Matplotlib 1.5.x supports Python versions 2.7 through 3.5
- [Matplotlib.org](http://matplotlib.org)

TO THE NOTEBOOK!

[Local notebook](#)

Inspired from this [Matplotlib tutorial](#)

[More matplotlib](#)

THEMES AND COLORS

COLORS

http://matplotlib.org/examples/color/named_colors.html

COLOR PALETTES

[list of color maps](#)

```
my_cmap = matplotlib.cm.get_cmap('rainbow')  
imshow(i, cmap=my_cmap)
```

MATPLOTLIB THEMES

https://raw.githubusercontent.com/rasbt/matplotlib-gallery/master/images/formatting_4.png

```
plt.style.use('fivethirtyeight')  
  
print(plt.style.available)
```

WHAT MAKES A NICE PLOT?

NICER PLOT ANYONE?

LESS IS MORE

Perfection is achieved not when there is nothing more to add, but when there is nothing left to take away.

Antoine de St Exupery

- Rules for better figures
- Illustration: simplifying
Excell

TEN SIMPLE RULES FOR BETTER FIGURES

- Rule 1: Know Your Audience
- Rule 2: Identify Your Message
- Rule 3: Adapt the Figure to the Support Medium
- Rule 4: Captions Are Not Optional
- Rule 5: Do Not Trust the Defaults
- Rule 6: Use Color Effectively
- Rule 7: Do Not Mislead the Reader
- Rule 8: Avoid “Chartjunk”
- Rule 9: Message Trumps Beauty
- Rule 10: Get the Right Tool

EDWARD TUFTE

American statistician and professor of political science, statistics, and computer science at Yale University

1983 Edward Tufte, [The Visual Display of Quantitative Information](#)

- Expert in Visual communication of information
- Chartjunk: all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information

6 PRINCIPLES OF EDWARD TUFTE

1. The representation of numbers, as physically measured on the surface of the graphic itself, should be **directly proportional** to the numerical quantities represented.
2. Clear, detailed, and thorough **labeling** should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.
3. Show **data variation**, not design variation.
4. In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
5. The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
6. Graphics must not quote data **out of context**.

6 PRINCIPLES EDWARD TUFTES

Principles of Information Display for Visualization Practitioners

Examples:

- <http://www.npr.org/2011/07/03/137536111/by-the-numbers-todays-military>
- <http://www.informationisbeautiful.net/visualizations/the-billion-dollar-gram/>
- <https://www.technologyreview.com/s/425120/space-over-time/>
- <https://upload.wikimedia.org/wikipedia/commons/2/29/Minard.png>

See also [Data Visualization Principles: Lessons from Tufte](#)

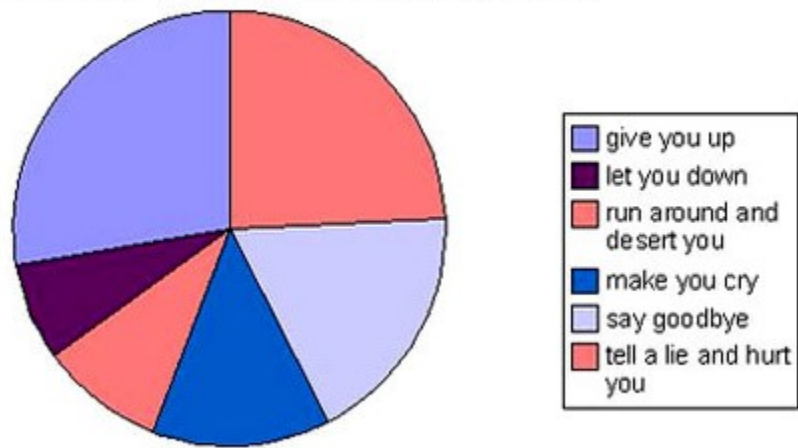
PIE CHARTS

ANY OPINION?

- Pie charts are simple and efficient
- Pie charts are confusing and should be avoided

PIE CHARTS

Rick Astley would never:



Never Gonna Give You Up
Rick Astley

0:00

PIE CHARTS

Pie charts are evil.

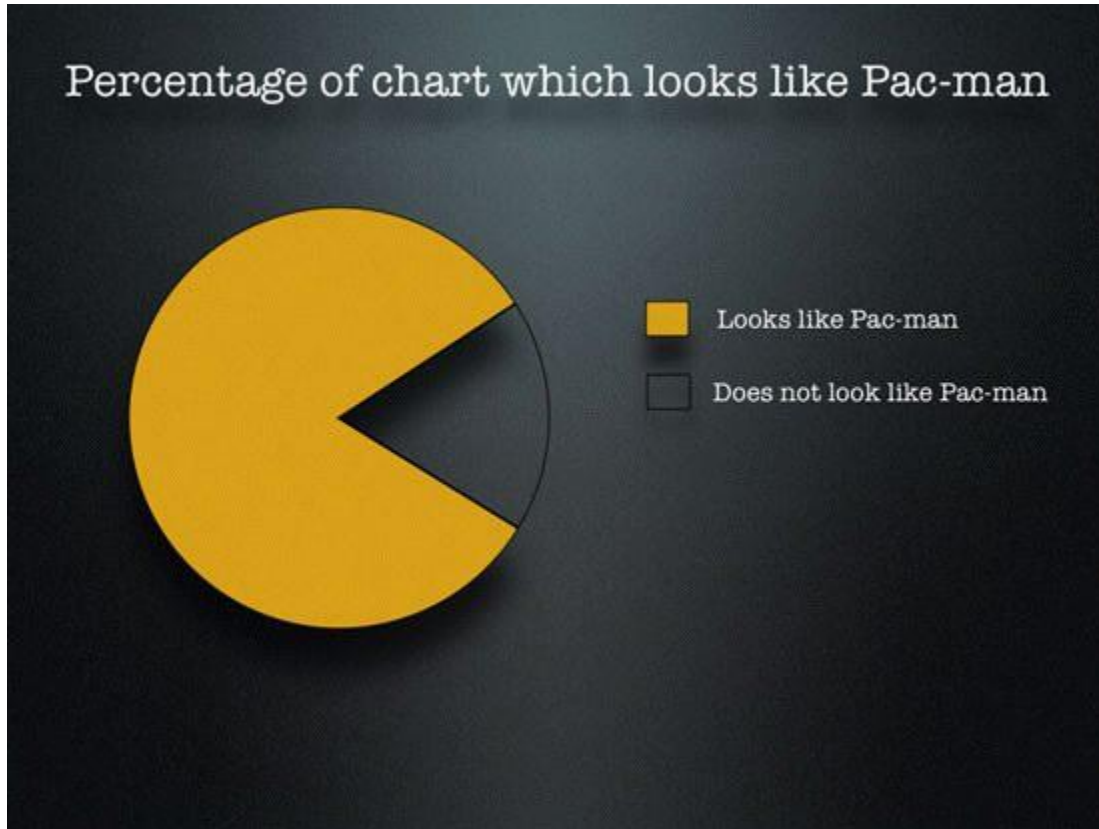
They represent much of what is wrong with the poor design of many websites and software applications. They're also innefective, misleading, and innacurate. Using a pie chart as your graph of choice to visually display important statistics and information demonstrates either a lack of knowledge, laziness, or poor design skills.

From [Countdown of Top 10 Reasons to Never Ever Use a Pie Chart](#)

WHY ?

[The Worst Chart In The World](#)

PIE CHARTS



See also [Save the Pies for Dessert](#) for more examples on the limitations of Pie charts.

MATPLOTLIB LAB

MATPLOTLIB LAB

- Open this [notebook](#)

SEABORN

SEABORN

SEABORN

[Seaborn](#) is a library for making attractive and informative statistical graphics in Python.

It is built on top of matplotlib and tightly integrated with the PyData stack, including support for numpy and pandas data structures and statistical routines from scipy and statsmodels.

SEABORN

- [Gallery](#)

```
iris = sns.load_dataset("iris")  
sns.pairplot(iris);
```

- [Seaborn
Datasets](#)

THE TIPS DATASET

One waiter recorded information about each tip he received over a period of a few months working in one restaurant.

He collected the following variables:

- tip in dollars,
- bill in dollars,
- sex of the bill payer,
- whether there were smokers in the party,
- day of the week,
- time of day,
- size of the party

In all he recorded 244 tips.

LAB 2 - SEABORN

Lab 2 - Seaborn

BOKEH

BOKEH

BOKEH

[Bokeh](#) is a Python interactive visualization library that targets modern web browsers for presentation. Its goal is to provide elegant, concise construction of novel graphics in the style of D3.js, and to extend this capability with high-performance interactivity over **very large** or **streaming datasets**

[Demo](#)

[Tutorial](#)

PLOTLY

PLOTLY

PLOTLY

- Interactive
- BI focused
- Google drive for charts!
- Excel, [Python](#), R compatible
- [Dashboards](#)
- Matplotlib and Seaborn
- [Example 1: scatter plots](#)
- [Example 2: Notebooks](#)

LESSON REVIEW

COURSE

BEFORE NEXT CLASS

5 QUESTIONS ABOUT TODAY

EXIT TICKET

EXIT TICKET

LINKS

[Tutorial | Basic Bokeh | EuroSciPy 2015 | Bryan Van de Ven](#)

[Bokeh is here Interactive Visualization for Python by Fabio Pliger](#)

[A Simple Time Series Analysis Of The S&P 500 Index](#)