

DataViz for Data Science

Concepts and Technologies

Who am I and who you are?

What to expect from each other?

What is dataviz for you?

Why/when dataviz is important?

My relationship with dataviz

The screenshot shows the homepage of the [r/dataisbeautiful](#) subreddit on Reddit. The interface includes a sidebar with various icons, a top navigation bar with links to reddit and the subreddit, a search bar, and a user profile for 'motams' with 1 karma. The main content area displays several posts, with the first one being a Battle post about visualizing information on all 802 Pokemon. To the right, there's a 'Community Details' sidebar showing 13.3m Subscribers and 7.0k Online users, both of which are highlighted with a red border. Below that, a description of the subreddit as a place for visual representations of data is provided, along with 'SUBSCRIBED' and 'CREATE POST' buttons. At the bottom, a 'USER FLAIR PREVIEW' section shows the user 'motams'.

Posts Wiki Posting Rules

VIEW SORT HOT

COMMUNITY DETAILS

r/dataisbeautiful

13.3m Subscribers 7.0k Online

A place for visual representations of data: Graphs, charts, maps, etc. DataIsBeautiful is for visualizations that effectively convey information. **Aesthetics are an important part of information visualization, but pretty pictures are not the aim of this subreddit.**

SUBSCRIBED

CREATE POST

USER FLAIR PREVIEW

motams

Posts Wiki Posting Rules

VIEW SORT HOT

COMMUNITY DETAILS

r/dataisbeautiful

13.3m Subscribers 7.0k Online

A place for visual representations of data: Graphs, charts, maps, etc. DataIsBeautiful is for visualizations that effectively convey information. **Aesthetics are an important part of information visualization, but pretty pictures are not the aim of this subreddit.**

SUBSCRIBED

CREATE POST

USER FLAIR PREVIEW

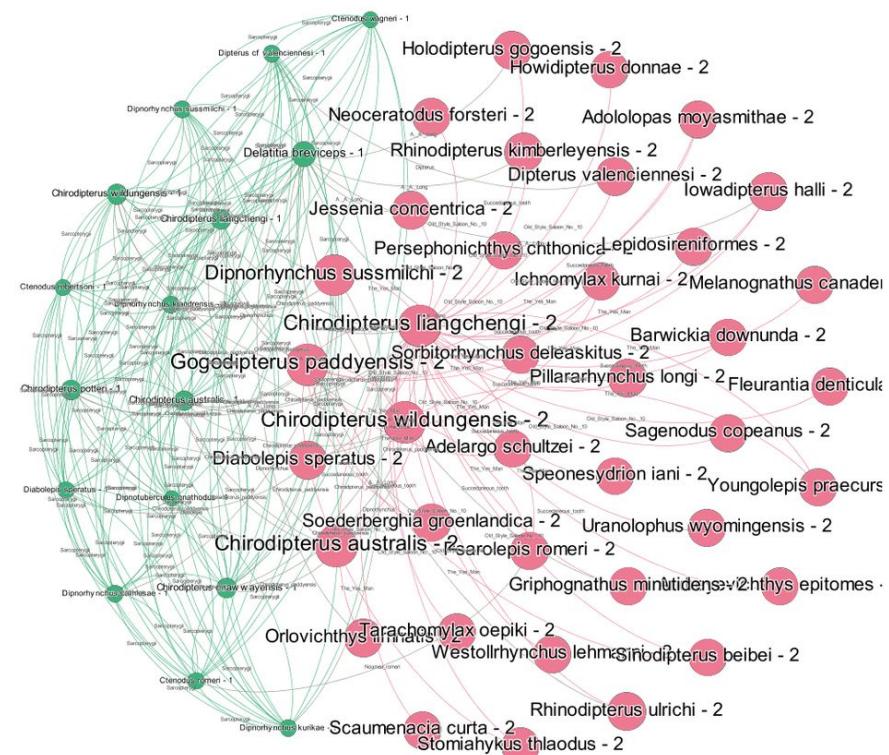
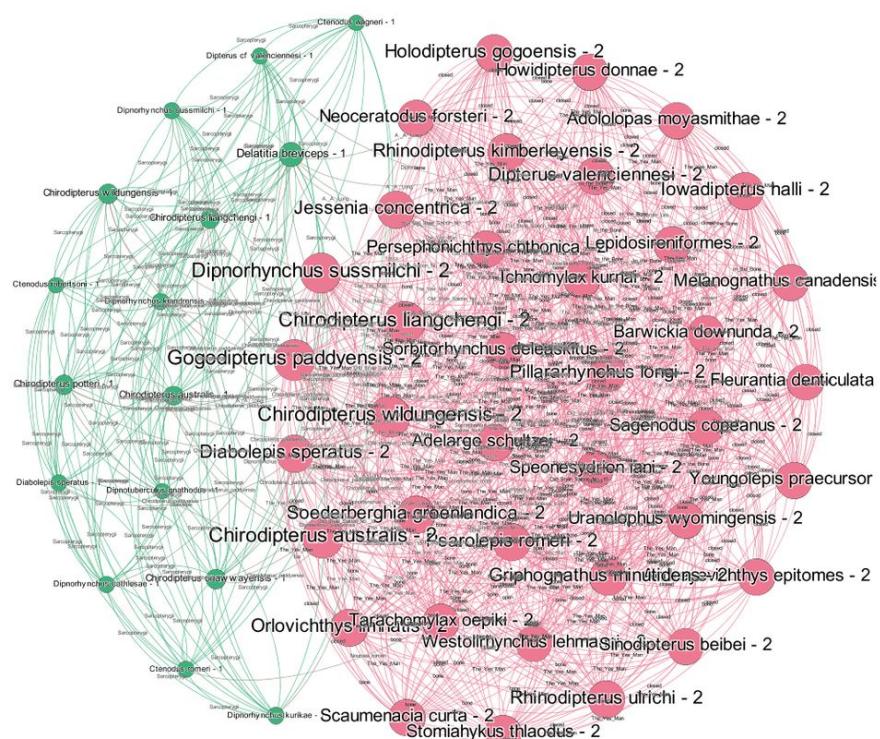
motams

My relationship with dataviz (/r/dataisbeautiful)

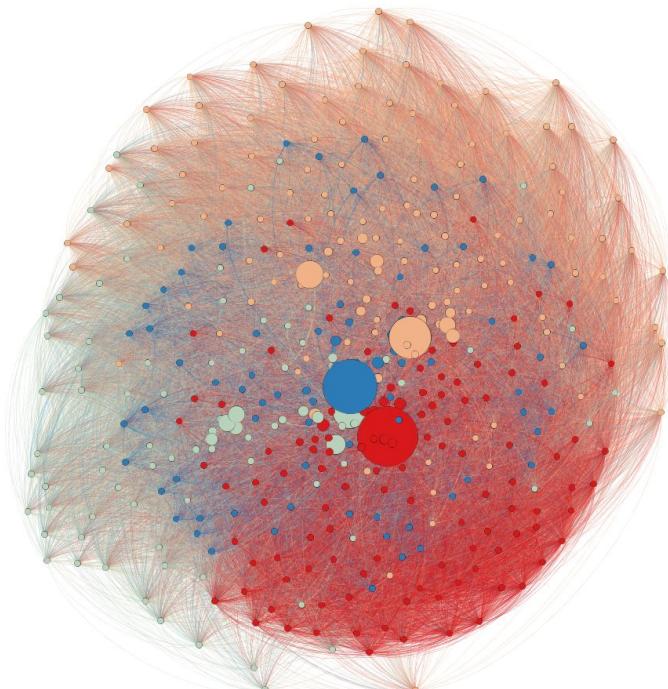
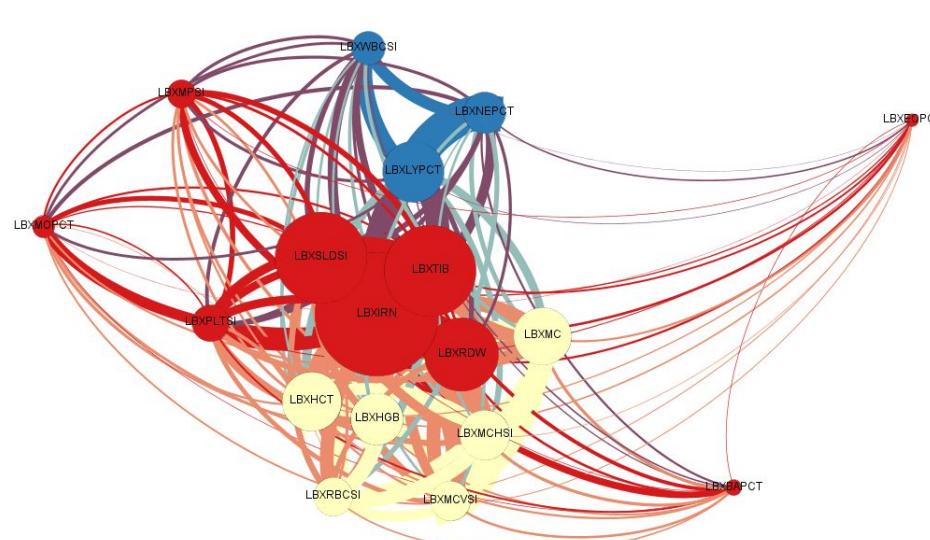
The screenshot shows the /r/dataisbeautiful subreddit on the Reddit website. The page displays a list of posts, each featuring a thumbnail image, the post title, the author, the number of upvotes, the number of comments, and various interaction buttons like 'Share', 'Save', 'Give Gold', 'Hide', and 'Report'. The posts are arranged in a descending order of upvotes. The sidebar on the right includes links for 'About', 'Careers', 'Press', 'Advertise', 'Blog', 'Help', 'The Reddit App', 'Reddit Gold', and 'Reddit Gifts'. It also contains legal links for 'Content Policy', 'Privacy Policy', 'User Agreement', and 'Mod Policy', along with a copyright notice for '© 2018 Reddit, Inc. All rights reserved'.

- [Battle] DataViz Battle for the month of April 2018: Visualize every line from every scene in The Office [OC] r/dataisbeautiful · Posted by u/AutoModerator 3 months ago
- 107 81 Comments Share Save Give Gold Hide Report
- oc Wife let you live a long life before they kill you. Roman Emperors dataset visualization - DataViz Battle of June [OC] r/dataisbeautiful · Posted by u/DataPro19 2 months ago
- 13 7 Comments Share Save Give Gold Hide Report
- oc Submission for June 2018 DataViz Battle [OC] r/dataisbeautiful · Posted by u/Schwarz_1991 2 months ago
- 2 2 Comments Share Save Give Gold Hide Report
- [Battle] DataViz Battle for the month of January 2018: Visualize the Growth Rates of Algae r/dataisbeautiful · Posted by u/AutoModerator 8 months ago
- 98 147 Comments Share Save Give Gold Hide Report
- [Battle] DataViz Battle for the month of February 2018: Visualize the Legal Status of Same-sex Marriage by US State and Year r/dataisbeautiful · Posted by u/AutoModerator 7 months ago
- 109 144 Comments Share Save Give Gold Hide Report
- [Battle] DataViz Battle for the month of March 2018: Visualized Over 100,000 Stars r/dataisbeautiful · Posted by u/AutoModerator 6 months ago
- 112 62 Comments Share Save Give Gold Hide Report
- oc Visualizing votes for contestants of the Battle for BFDI team BEEP. [OC] r/dataisbeautiful · Posted by u/ColossalCocaine 3 months ago
- 7 2 Comments Share Save Give Gold Hide Report
- [Lounge] This week is a Bye Week for the DatViz Battles. Use this thread for off-topic discussion, smack talk, and cool suggestions! r/dataisbeautiful · Posted by u/AutoModerator 4 months ago
- 17 12 Comments Share Save Give Gold Hide Report
- 12.8k Malicious The Battle for your TV - Cable vs Netflix viewership [OC] r/dataisbeautiful · Posted by u/DataZombiez 1 year ago
- 12.8k 1.0K Comments Share Save Give Gold Hide Report
- oc Do Numbers Always Win? Analyzing troop strengths and losses in 486 historical battles [OC] r/dataisbeautiful · Posted by u/roninachi_beta 4 months ago
- 20 3 Comments Share Save Give Gold Hide Report
- oc Who Drives the Show? Word Counts over Seasons, April DataViz Battle submission [OC] r/dataisbeautiful · Posted by u/drew_123 4 months ago
- 1 BACK TO TOP

My relationship with dataviz



My relationship with dataviz



My relationship with data science



Tweets
1,978

Following
640

Followers
14.2K

Likes
963

Lists
2

Moments
0

Edit profile

Matheus Mota

@matheusmota

Another primate hurtling through space on top of an ordinary rock orbiting a common star in an average galaxy.
Interested in Data Science & Engineering.

📍 Brazil

💻 Joined March 2009

⌚ Born on August 7, 1986

🖼 138 Photos and videos



Tweets

Tweets & replies

Media

You Retweeted



Elias Tandel @etandel · Sep 14

Adriano datascientist

🌐 Translate Tweet



言论

11

心 23

✉

You Retweeted

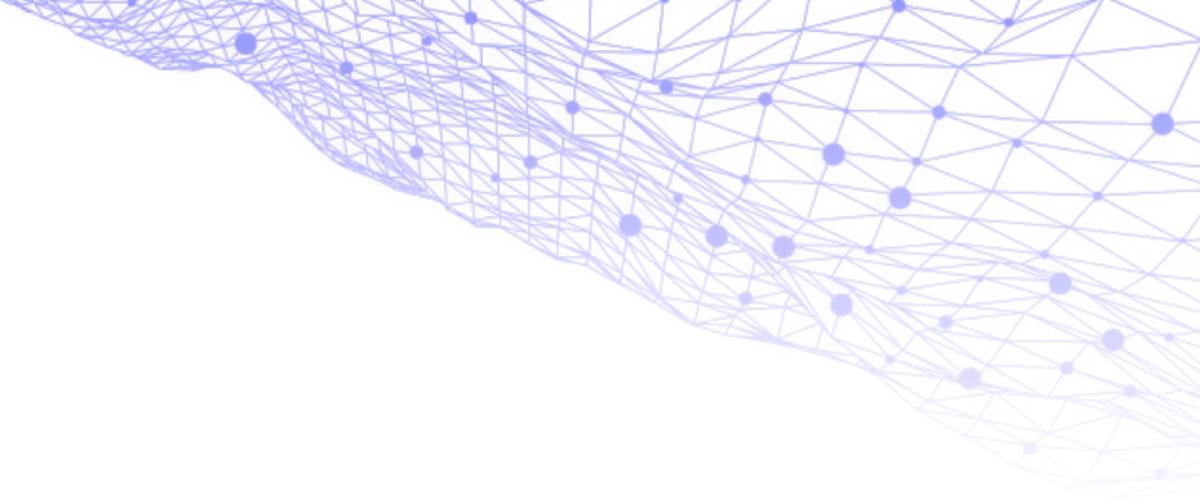


Silvia Benza @sbbareiro · 13 Mar 2017

My relationship with data science

<https://towardsdatascience.com/>

The screenshot shows a web browser displaying a blog post on the Towards Data Science website. The header includes the site's logo, navigation links for HOME, DATA SCIENCE, MACHINE LEARNING, PROGRAMMING, VISUALIZATION, PICKS, and CONTRIBUTE, and social sharing icons. The main content features a photo of the author, Matheus Mota, and a brief bio: "Data Engineer interested in Data Science aaS, Semantic Web, IoT, and Blockchain." Below this is the title "JupyterLab is the data science UI we have been looking for". The text discusses the Project Jupyter tool, highlighting its polyglot nature, web-based interface, and support for reproducibility and scientific computing across multiple programming languages via notebooks. It also notes the language-agnostic behavior of Jupyter and its growing community, mentioning extensions like a kernel for SAS maintained by SAS itself. At the bottom, there are two screenshots of the Jupyter Notebook interface. The left screenshot shows the "Welcome to jupyter" page. The right screenshot shows a notebook titled "Exploring the Lorenz System" with code for solving differential equations and a plot of the Lorenz attractor.



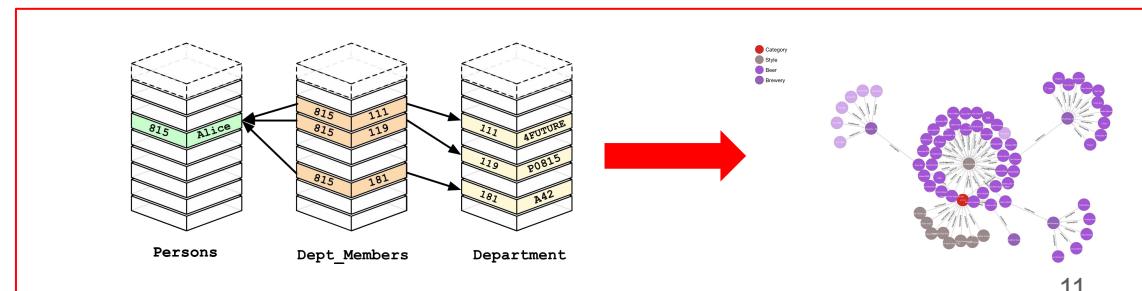
Course Structure

Agenda

- New items for your datascience backpack
 - environment, reproducibility, story-telling
- Exploratory Analysis
 - Variables
 - Distributions -> understanding latent characteristics of large lists
 - Graphical Structures
 - Line Plot, Bar chart, Box plot, Histogram, Pie chart, Scatter plot, Violin Plot
- Libraries
 - matplotlib
 - pandas.plot()
 - seaborn
 - plotly
- Tools
 - Tableau
 - Voyager

Agenda

- New items for your datascience backpack
 - environment, reproducibility, story-telling
- Exploratory Analysis
 - Variables
 - Distributions -> understanding latent characteristics of large lists
 - Graphical Structures
 - Line Plot, Bar chart, Box plot, Histogram, Pie chart, Scatter plot, Violin Plot
- Libraries
 - matplotlib
 - pandas.plot()
 - seaborn
 - plotly
- Extra
 - Radar Chart, Error Chart etc.
 - Other JS and/or R Libraries
 - **Graphs/Graph Database**
- Tools
 - Tableau
 - Voyager



Evaluation

There will be :

- Two extra class activities ~~\$exercises = [ex1, ex2]~~
- A final group work \$fw
- An extra work \$ew

$$\$exs = \text{average}(\$exercises)$$

$$\max(\$ew)=2$$

$$\$final_score = \$fw * 0.40 + \$exs * 0.60 + \$ew$$

$$\$final_score = \$fw + \$ew$$

Tools/Technologies/Platforms



kaggle



Tools/Technologies/Platforms



kaggle



kaggle

binder



Tools/Technologies/Platforms



python



kaggle



kaggle

binder



Tools/Technologies/Platforms

Seaborn



matplotlib



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



kaggle



kaggle

binder



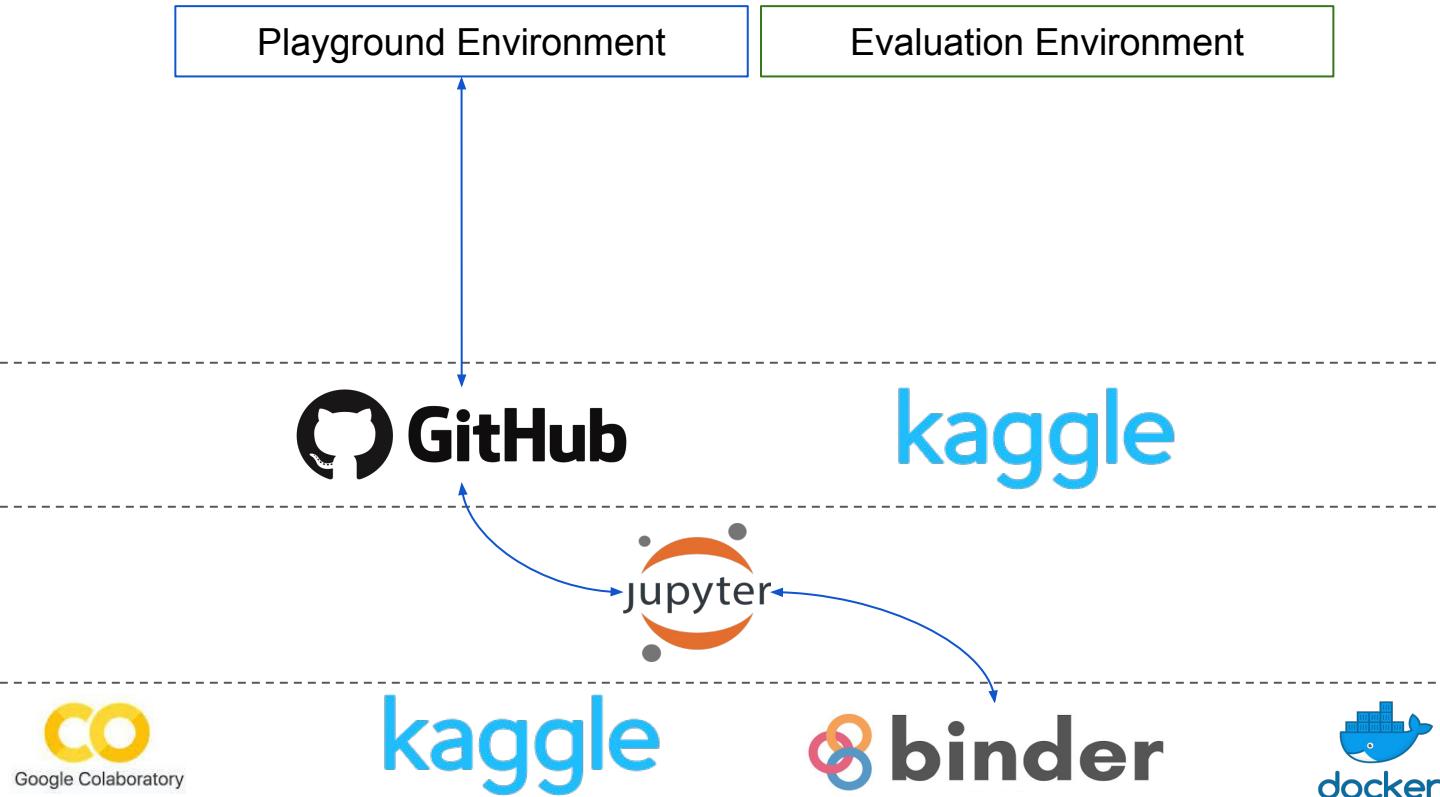
Tools/Technologies/Platforms

Playground Environment

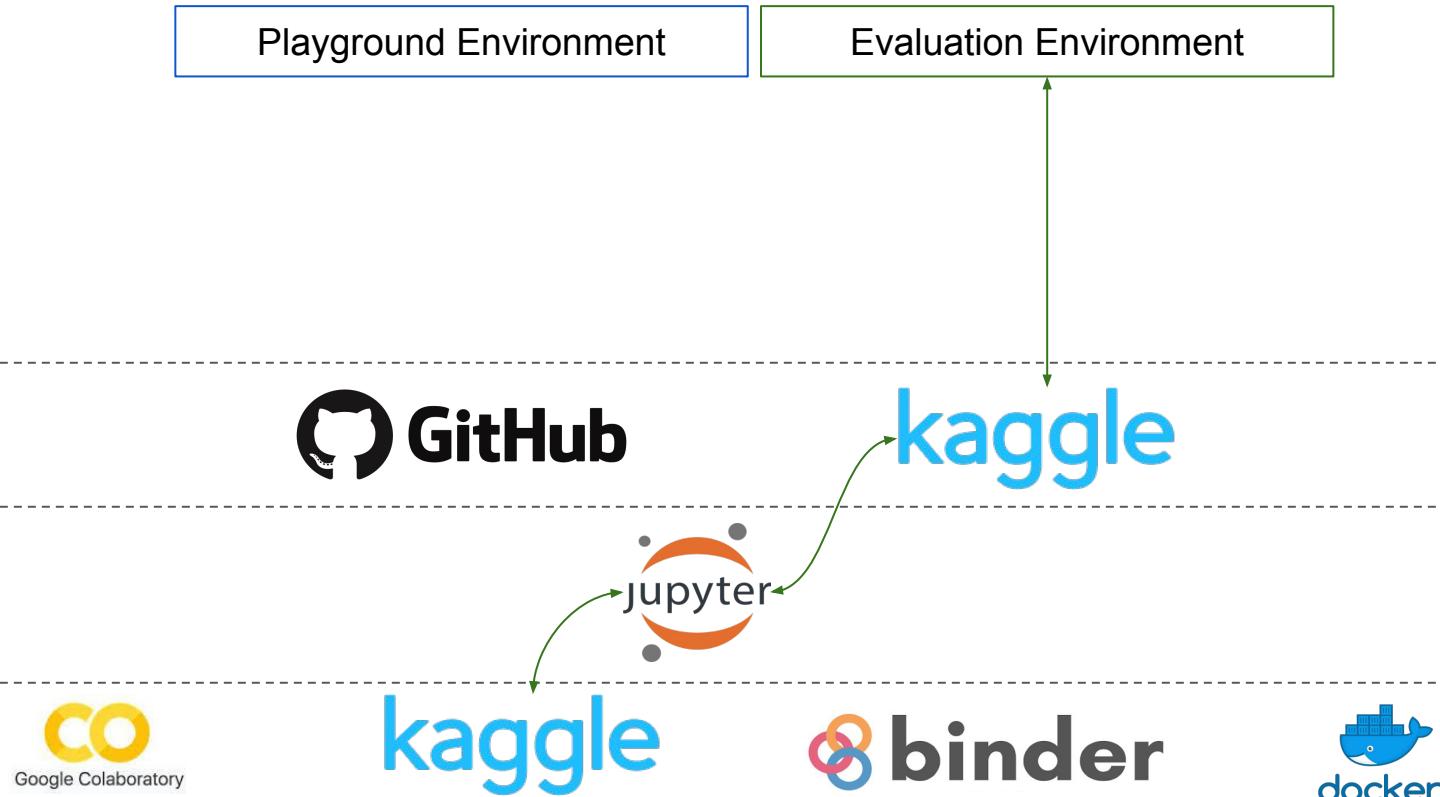
Evaluation Environment



Tools/Technologies/Platforms



Tools/Technologies/Platforms



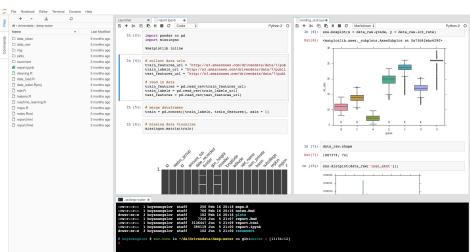
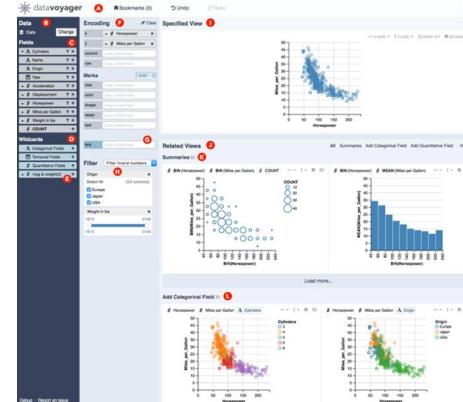
Tools/Technologies/Platforms



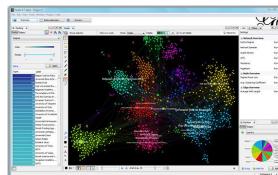
Grafana



datavoyager

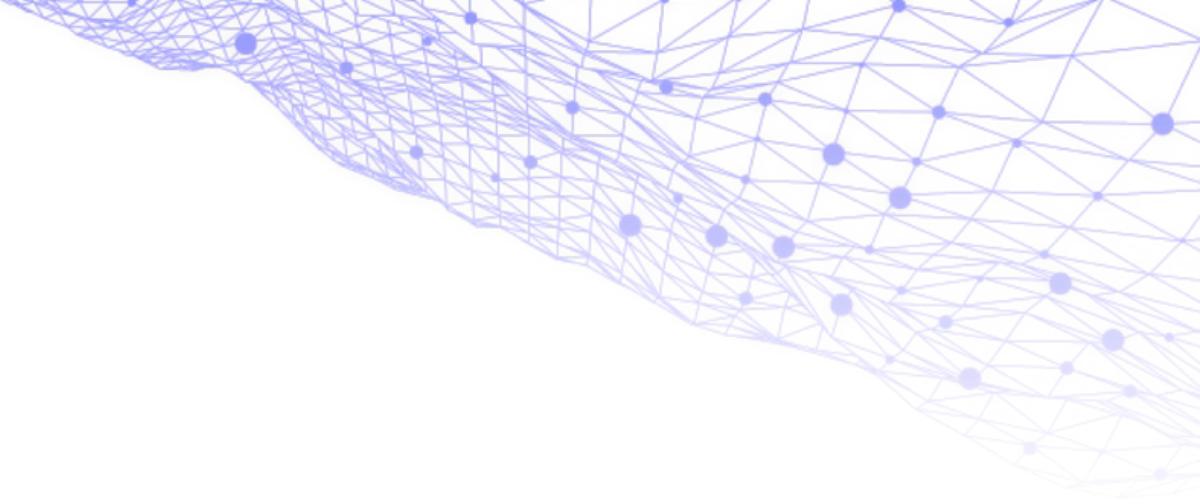


Gephi



neo4j





New items for your datascience backpack

environment, reproducibility, story-telling

Data Science Tools

IP[y]: Notebook spectrogram Last Checkpoint: a few seconds ago (autosaved) IPython (Python 3)

File Edit View Insert Cell Kernel Help

Cell Toolbar: None

Simple spectral analysis

An illustration of the [Discrete Fourier Transform](#) using windowing, to reveal the frequency content of a sound signal.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad k = 0, \dots, N - 1$$

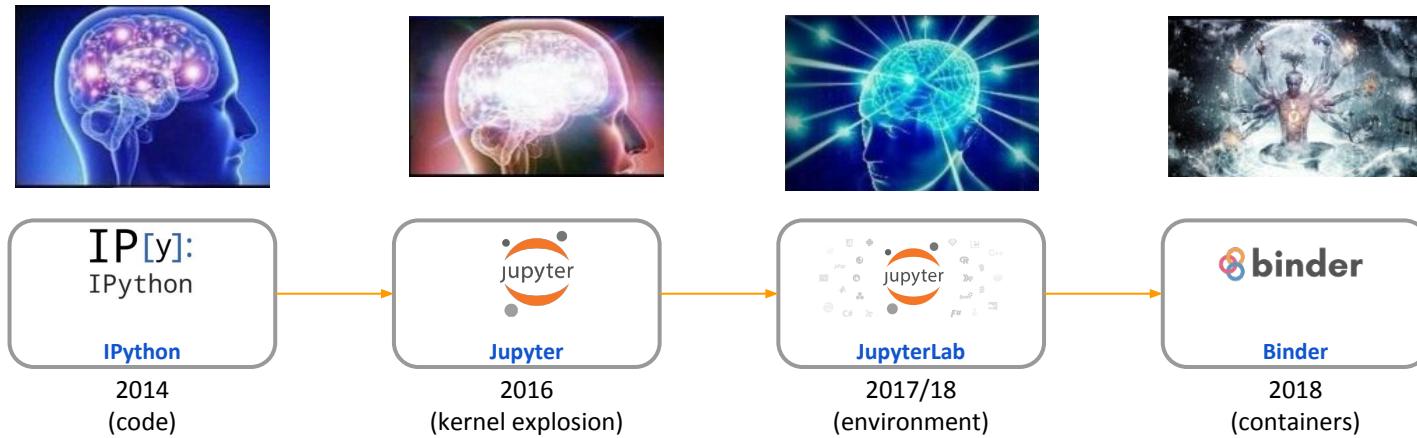
We begin by loading a datafile using SciPy's audio file support:

```
In [1]: from scipy.io import wavfile  
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view its spectral structure using matplotlib's builtin specgram routine:

```
In [2]: %matplotlib inline  
from matplotlib import pyplot as plt  
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))  
ax1.plot(x); ax1.set_title('Raw audio signal')  
ax2.specgram(x); ax2.set_title('Spectrogram');
```

Why Notebooks



History

IP[y]:
IPython

IPython

2014
(code)

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 515 > Issue 7525 > Toolbox > Article

E-alert RSS Facebook Twitter

NATURE | TOOLBOX

Interactive notebooks: Sharing the code

The free IPython notebook makes data analysis easier to record, understand and reproduce.

Helen Shen

05 November 2014

PDF Rights & Permissions



What matters in science — and why — free in your inbox every weekday.

Sign up

Listen

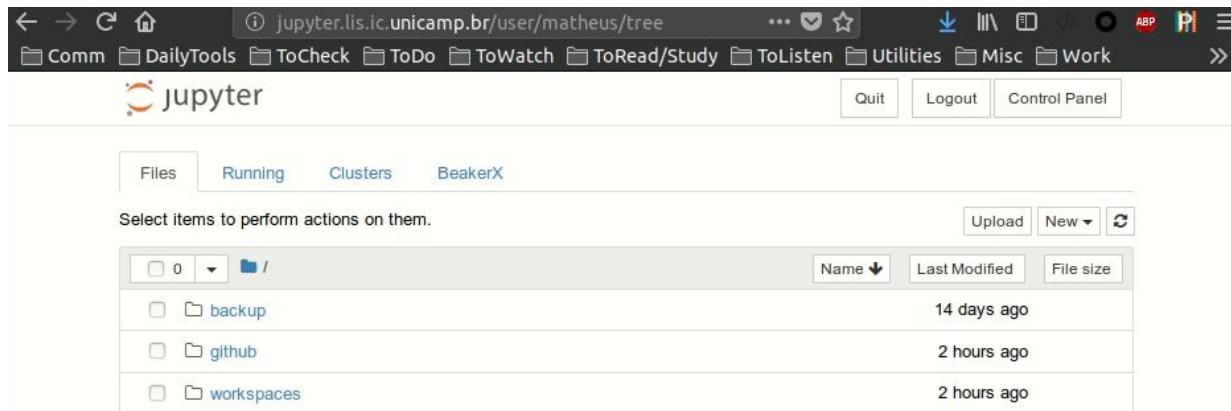
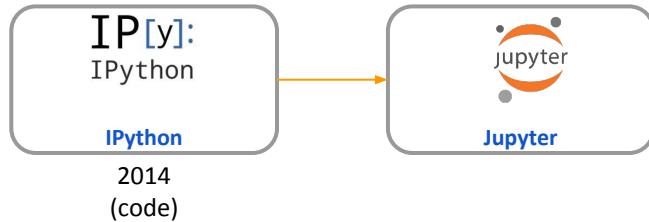
Nature Podcast

Our award-winning show features highlights from the week's edition of Nature, interviews with the people behind the science, and in-depth commentary and analysis from journalists around the world.

Science jobs from naturejobs

South China Normal University sincerely invite overseas talented scholars to apply for the

History

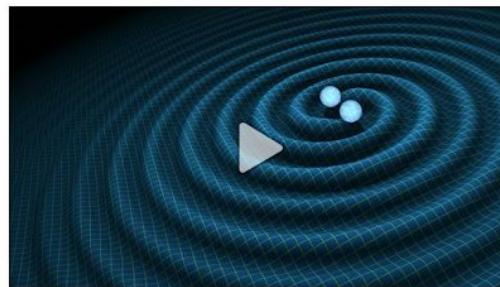


History



Gravitational waves detected -- and
that's creating waves in science

By Todd Leopold, CNN
Updated 0412 GMT (1212 HKT) February 12, 2016



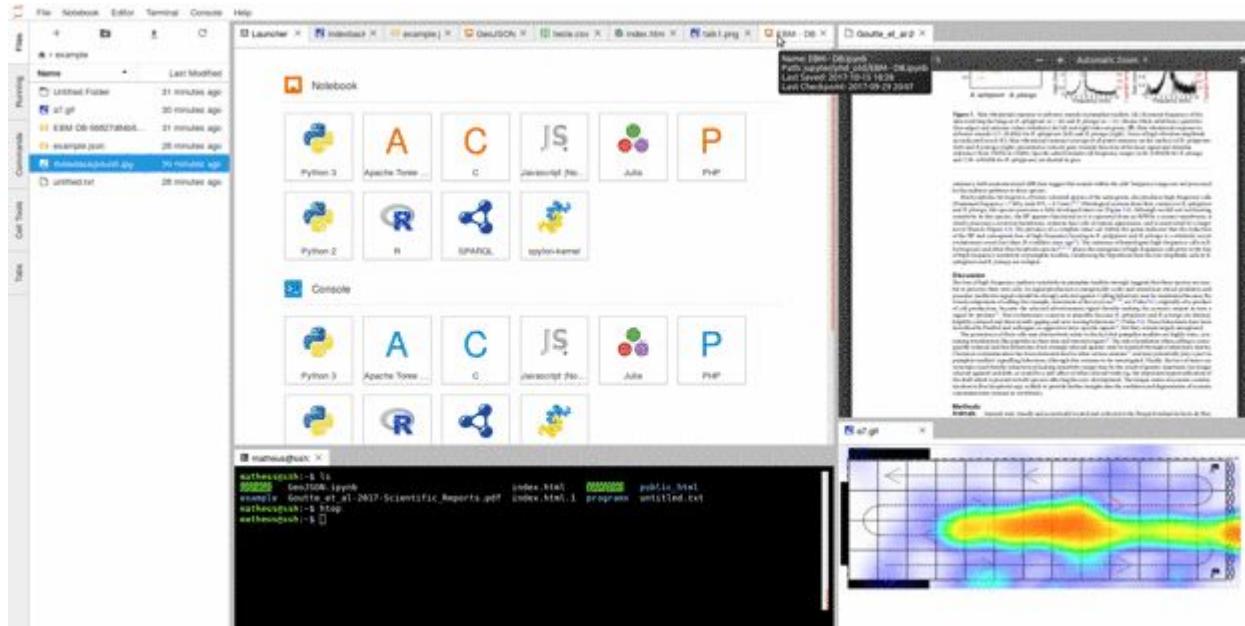
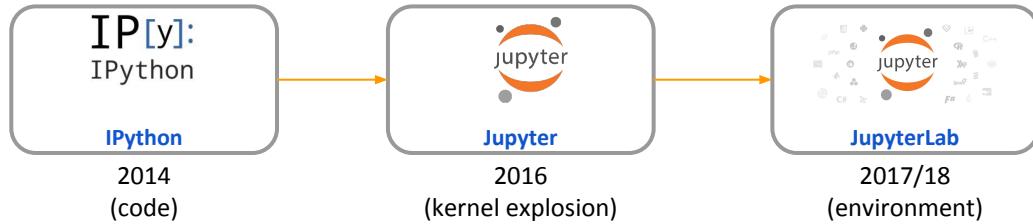
Story highlights

NEW: Discovery is "unlike anything we've ever detected before," one physicist says
Scientists: "We have detected gravitational waves!"
Observatory that discovered them is like a "cosmic microphone"

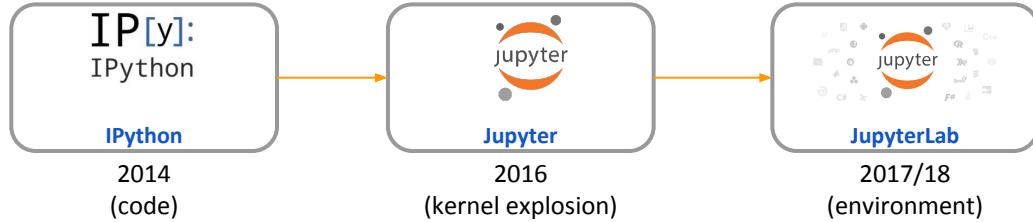
(CNN) — Einstein was right.

Just over 100 years after he published his general theory of relativity, scientists have found what Albert Einstein predicted as part of the theory: gravitational waves.
"We have detected gravitational waves. We did it," said David Reitze, executive director of LIGO, the Laser

History

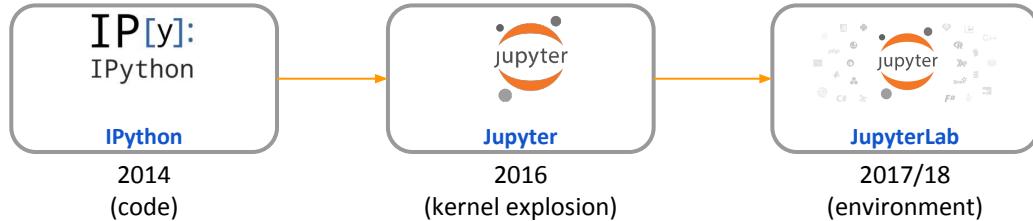


History



Jupyter Lab Demo

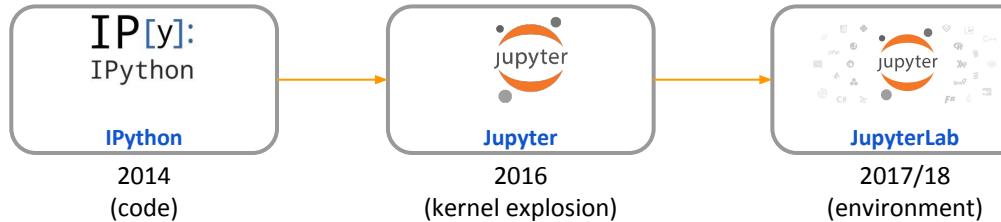
History



The screenshot shows the ACM Communications website with the following details:

- URL:** <https://cacm.acm.org/opinion/articles/226698-the-scientific-paper-is-obsolete/fulltext>
- Page Title:** COMMUNICATIONS OF THE ACM
- Section:** OPINION
- Article Title:** The Scientific Paper Is Obsolete
- Author:** By The Atlantic
- Date:** April 6, 2018
- Comments:** (link)
- Share Options:** View As: (PDF, EPUB), Share: (Email, Print, LinkedIn, Google+, Twitter, Facebook, etc.)
- Image:** A photograph of a flame.
- Text:** The scientific paper—the actual form of it—was one of the enabling inventions of modernity. Before it was developed in the 1600s, results were communicated privately in letters, ephemerally in lectures, or all at once in books.
- Text:** From The Atlantic
- Text:** View Full Article
- Right Sidebar:**
 - SIGN IN for Full Access**
 - User Name
 - Password
 - Forgot Password?
 - Create an ACM Web Account
 - SIGN IN
- MORE NEWS & OPINIONS**
 - If There's Life on Saturn's Moon Enceladus, It Might Look Like This
 - One Space Between Each Sentence, They Said. Science Just Proved Them Wrong.
 - When Will We Learn?
- ACM RESOURCES**
 - What's New in IBM Lotus
 - Notes 8
 - Courses

Past, Present and Future



Beyond Interactive: Notebook Innovation at Netflix

By [Michelle Ufford](#), [M Pacer](#), [Matthew Seal](#), and [Kyle Kelley](#)

Notebooks have rapidly grown in popularity among data scientists to become the de facto standard for quick prototyping and exploratory analysis. At Netflix, we're pushing the boundaries even further, reimagining what a notebook can be, who can use it, and what they can do with it. And we're making big investments to help make this vision a reality.

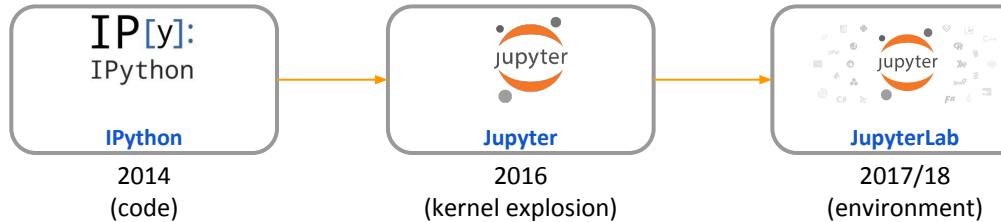
In this post, we'll share our motivations and why we find Jupyter notebooks so compelling. We'll also introduce components of our notebook infrastructure and explore some of the novel ways we're using notebooks at Netflix.

If you're short on time, we suggest jumping down to the Use Cases section.

Motivations

Data powers Netflix. It permeates our thoughts, informs our decisions, and challenges our assumptions. It fuels experimentation and innovation at

Past, Present and Future



Applause from you and 1,400 others

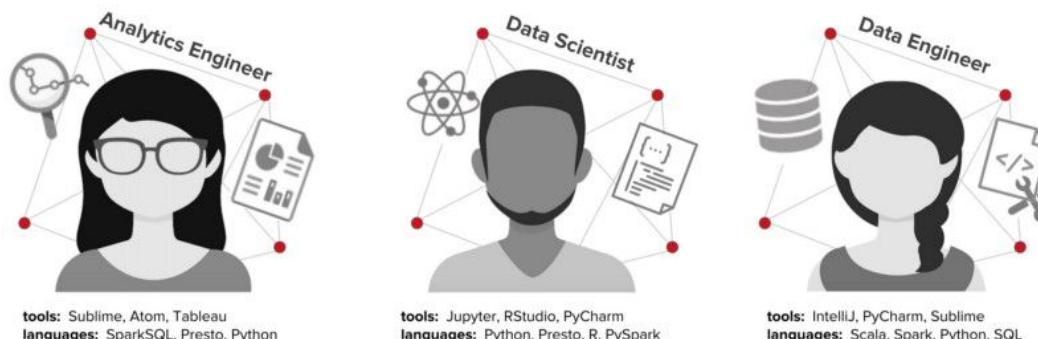
Netflix Technology Blog [Follow](#)

Learn more about how Netflix designs, builds, and operates our systems and engineering organizations

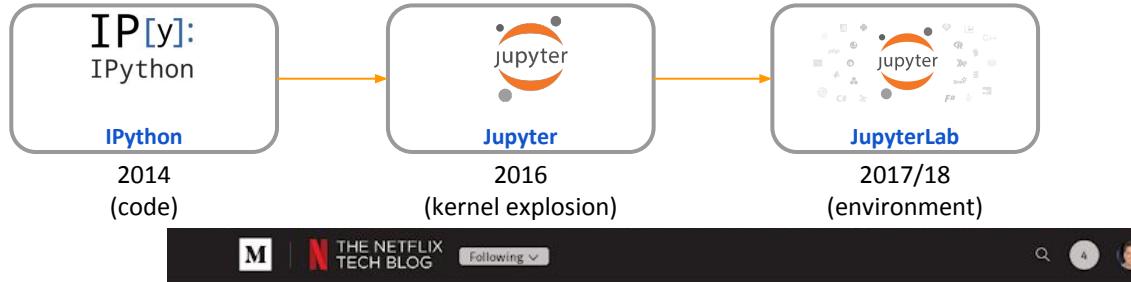
Aug 16 · 13 min read

Beyond Interactive: Notebook Innovation at Netflix

By [Michelle Ufford](#), [M Pacer](#), [Matthew Seal](#), and [Kyle Kelley](#)



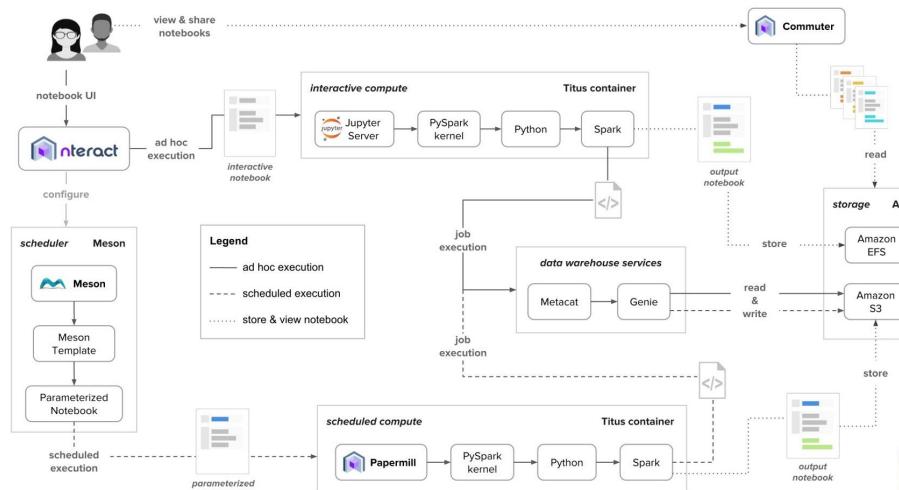
Past, Present and Future



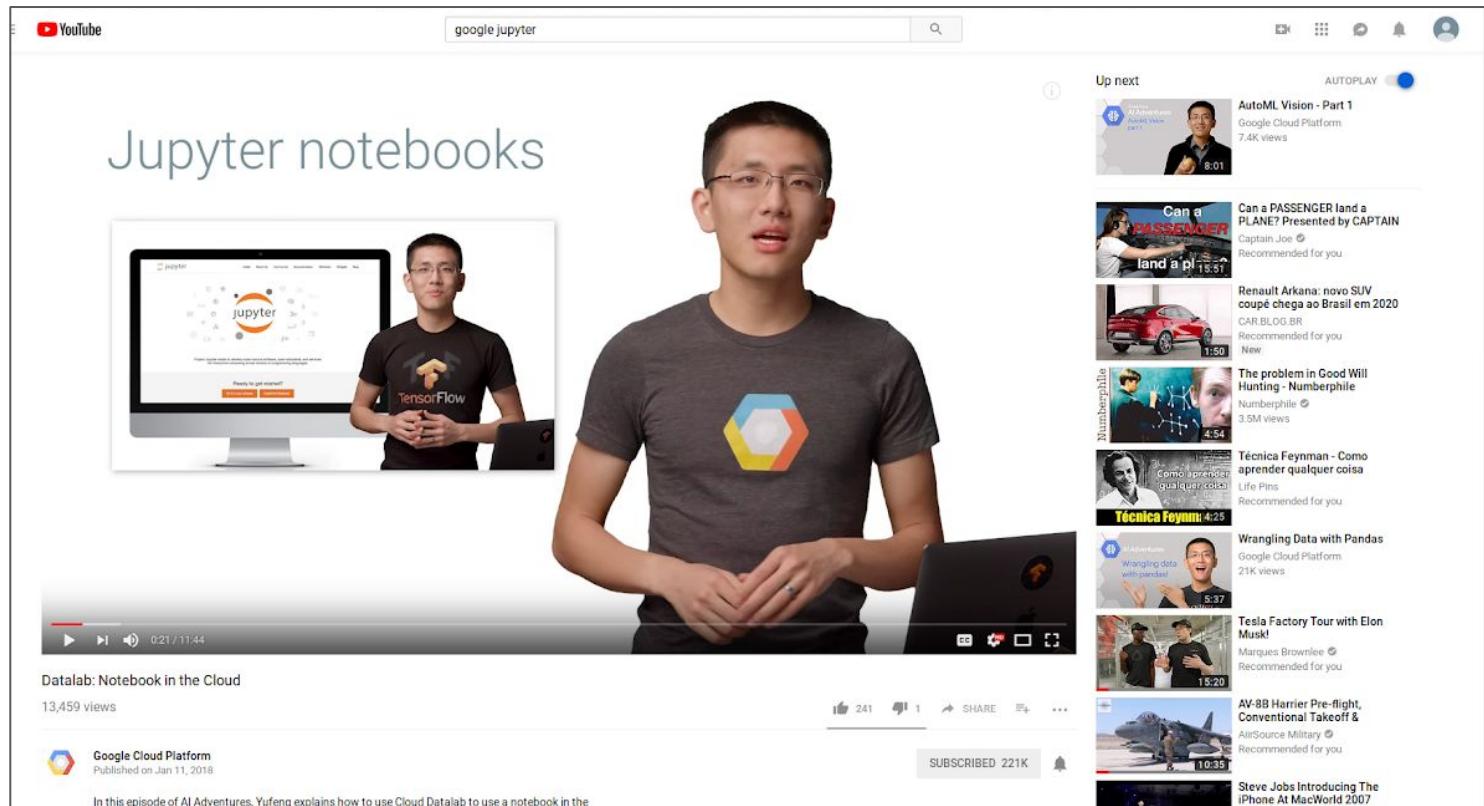
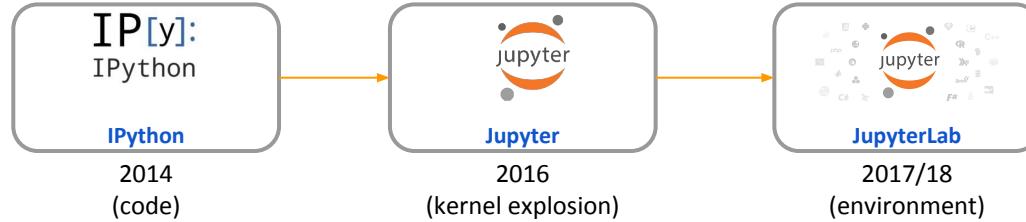
Applause from you and 1,400 others

Netflix Technology Blog [Follow](#)
Learn more about how Netflix designs, builds, and operates our systems and engineering organizations
Aug 16 · 13 min read

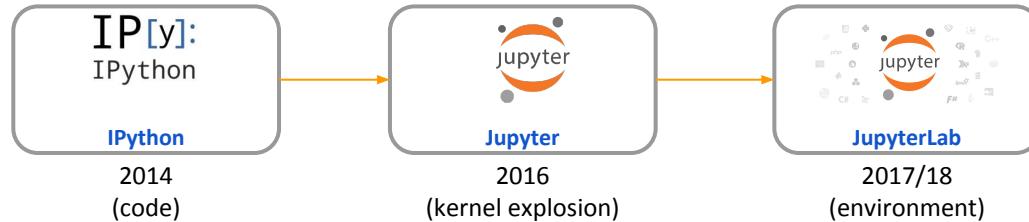
Beyond Interactive: Notebook Innovation at Netflix



Now and Future



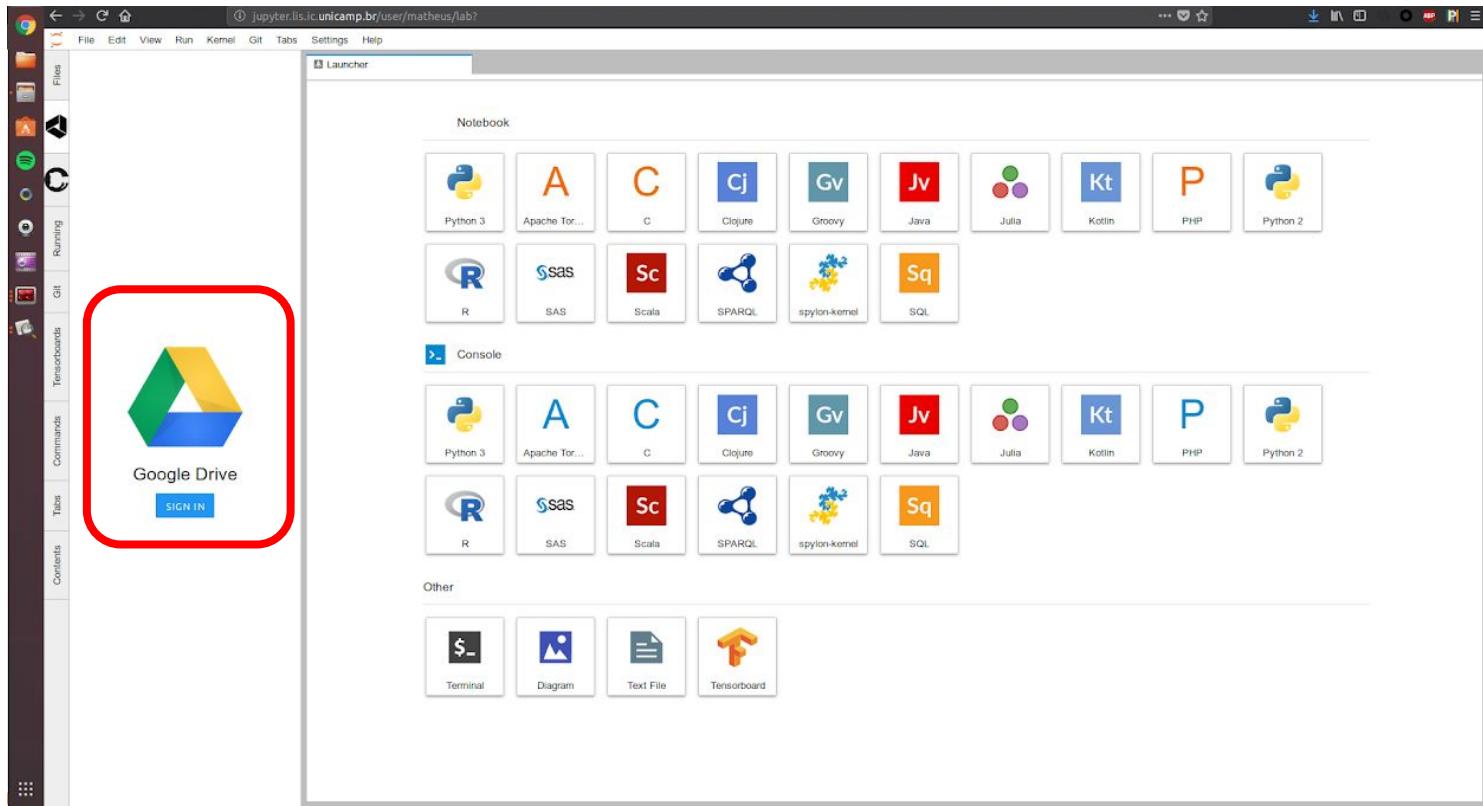
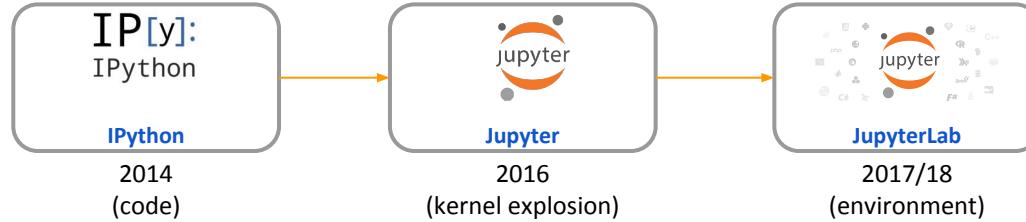
Now and Future



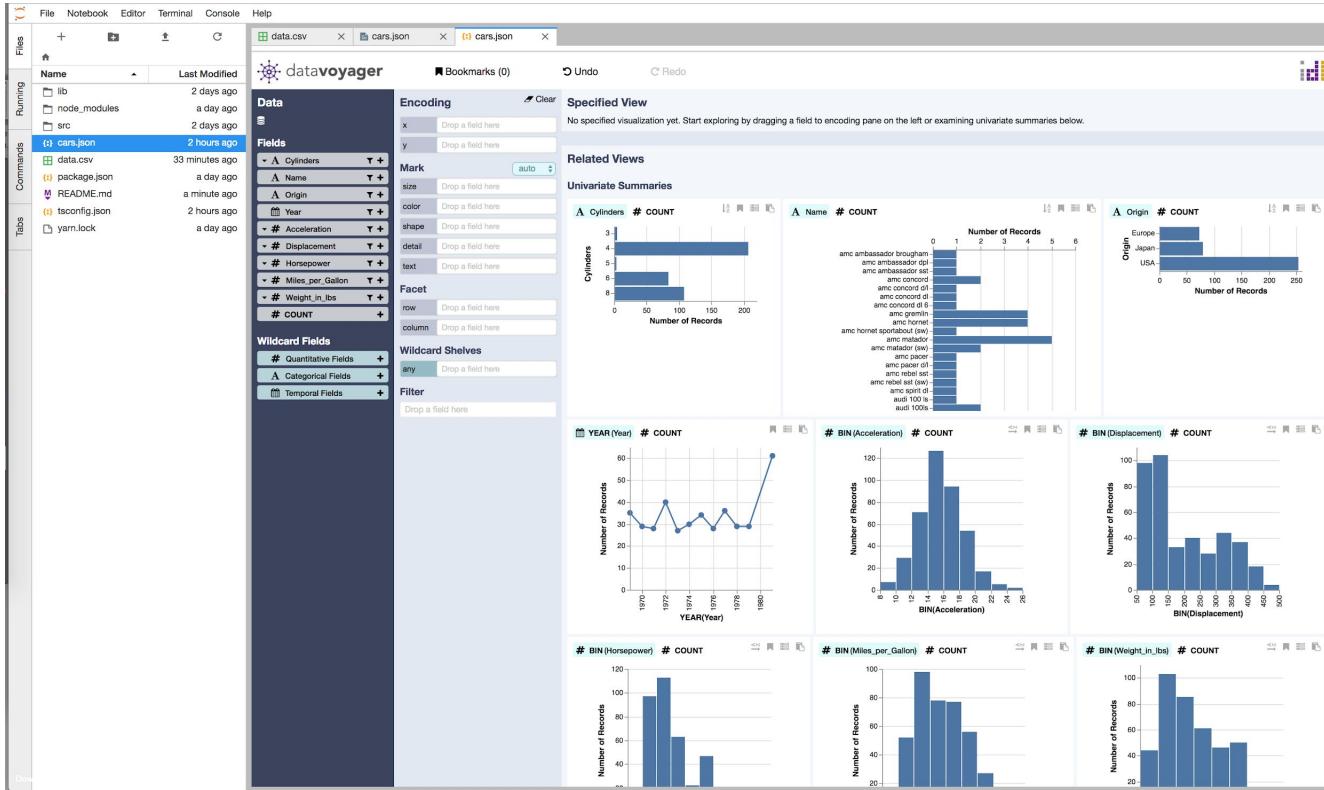
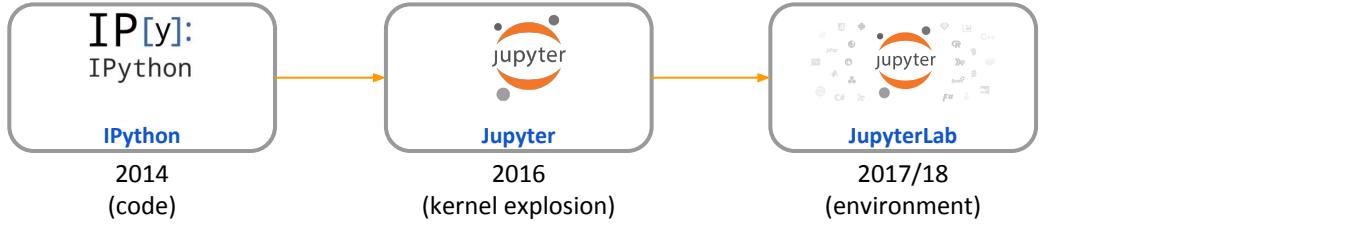
Screenshot of the Google Research Colaboratory interface, showing the following features:

- Getting Started**: Overview of Colaboratory, Loading and saving data, Local files, Drive, Sheets, Google Cloud Storage, Importing libraries and installing dependencies, Using Google Cloud BigQuery, Forms, Charts, Markdown, & Widgets, TensorFlow with GPU, Machine Learning Crash Course: Intro to Pandas & First Steps with TensorFlow.
- Highlighted Features**: Seedbank (Discover interactive machine learning examples), TensorFlow execution (Execute TensorFlow code in your browser).
- Code Snippets**: A snippet shows matrix addition:
$$\begin{bmatrix} 1. & 1. & 1. \\ 1. & 1. & 1. \end{bmatrix} + \begin{bmatrix} 1. & 2. & 3. \\ 4. & 5. & 6. \end{bmatrix} = \begin{bmatrix} 2. & 3. & 4. \\ 5. & 6. & 7. \end{bmatrix}$$
- TensorFlow execution**: A code cell imports tensorflow, creates two 2x3 matrices, adds them, and evaluates the result.
- GitHub**: A link to GitHub.

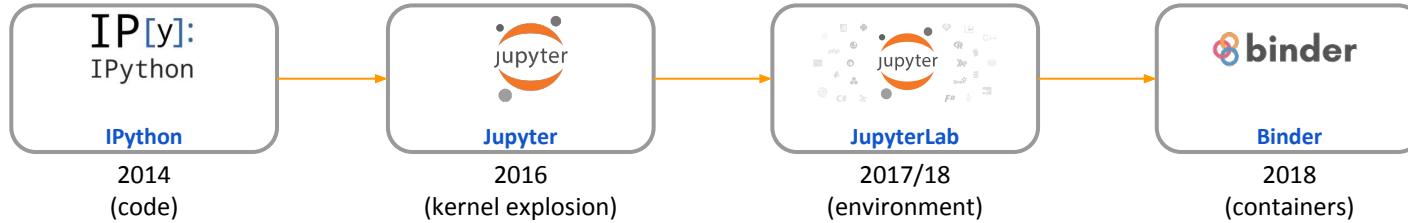
Now and Future



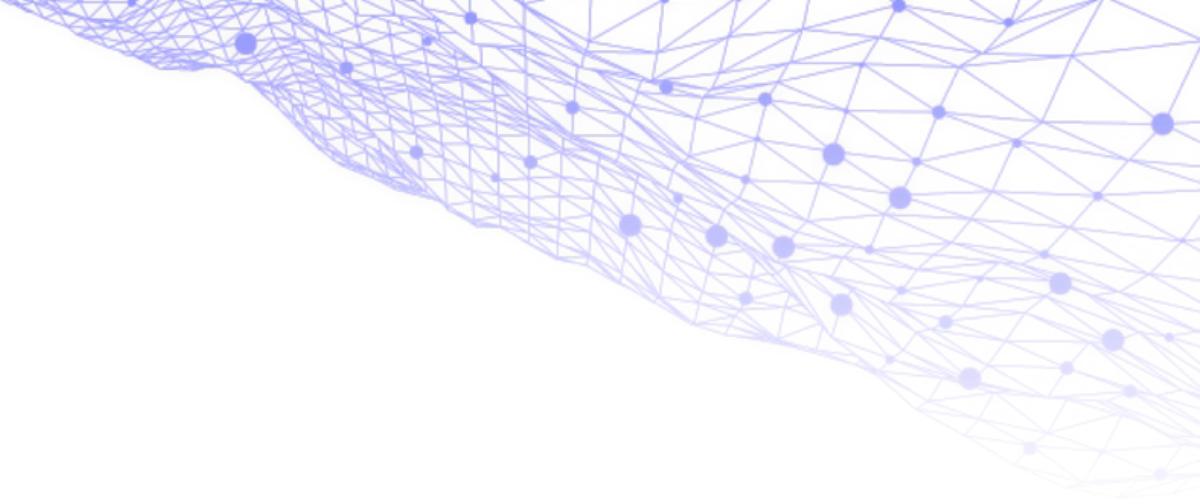
Now and Future



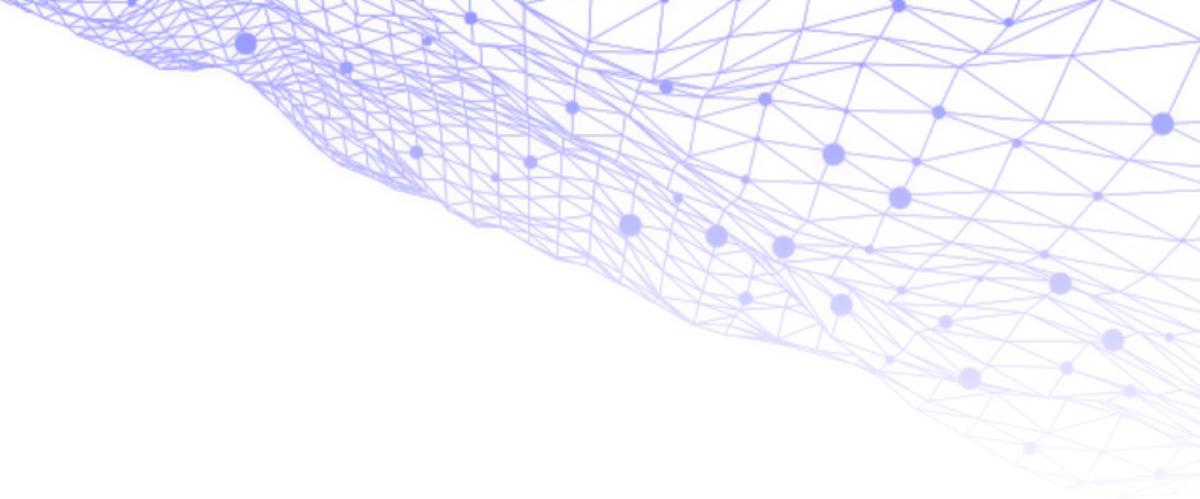
Now and Future



<https://github.com/thedatasociety/lab-dataviz>



Let's understand our environments



Dataviz relationship with Data Science

Dataviz? Why? Where?

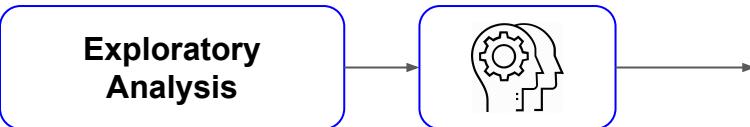
A Ultra Super Simple Data Science Process

Exploratory
Analysis



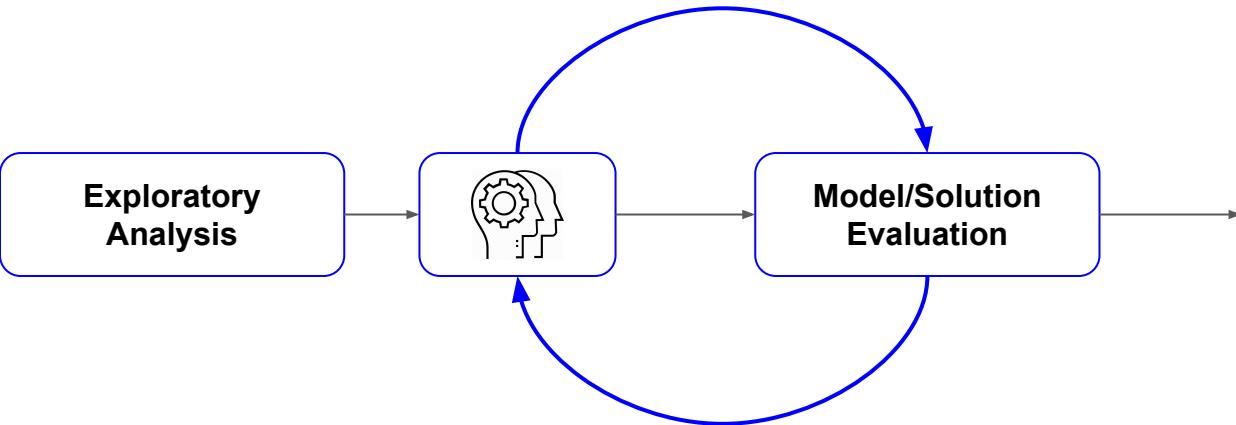
Dataviz? Why? Where?

A Ultra Super Simple Data Science Process



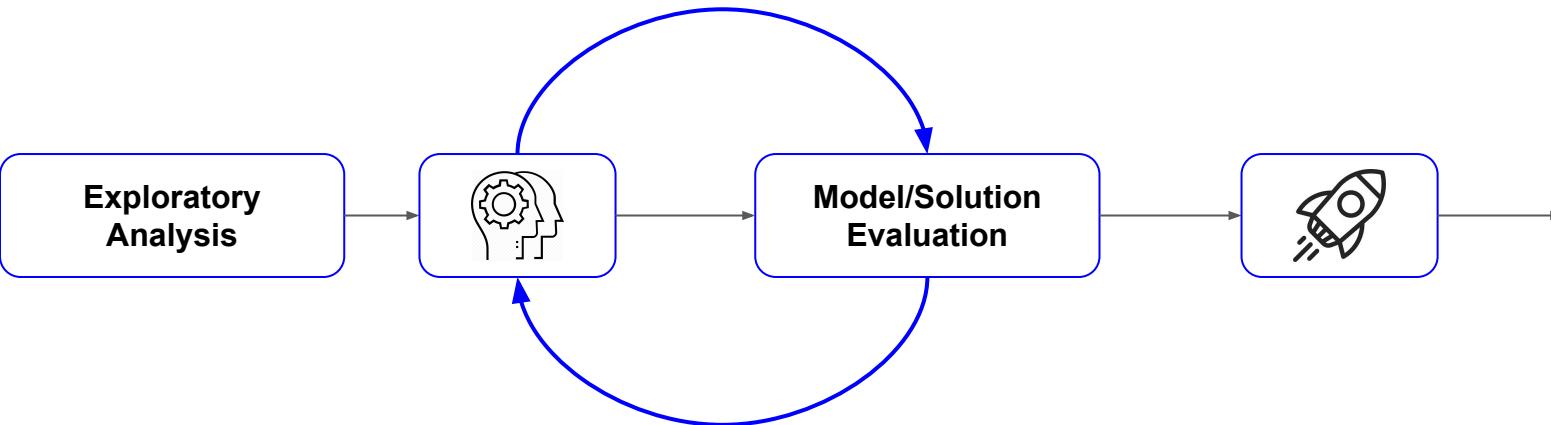
Dataviz? Why? Where?

A Ultra Super Simple Data Science Process



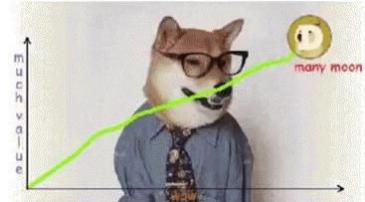
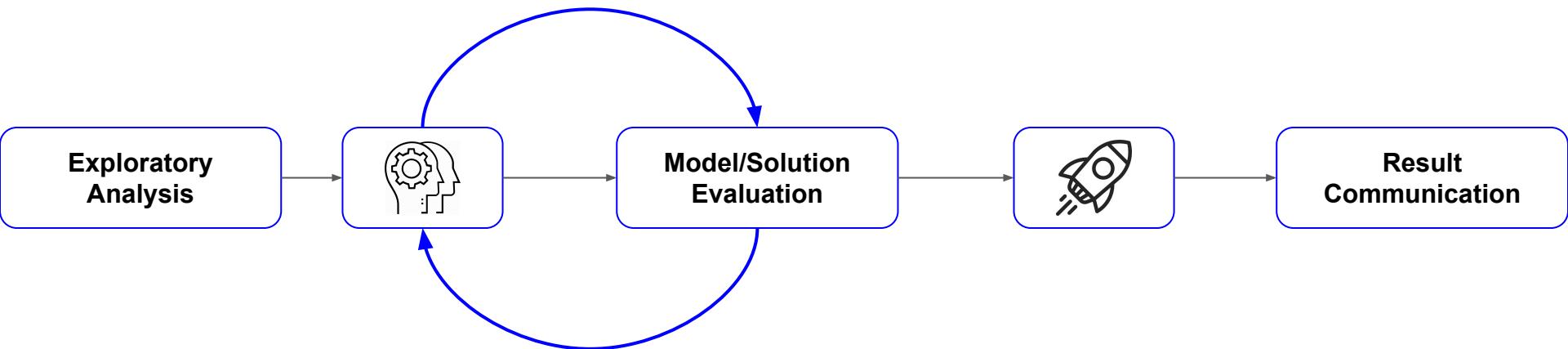
Dataviz? Why? Where?

A Ultra Super Simple Data Science Process



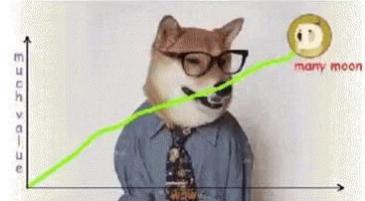
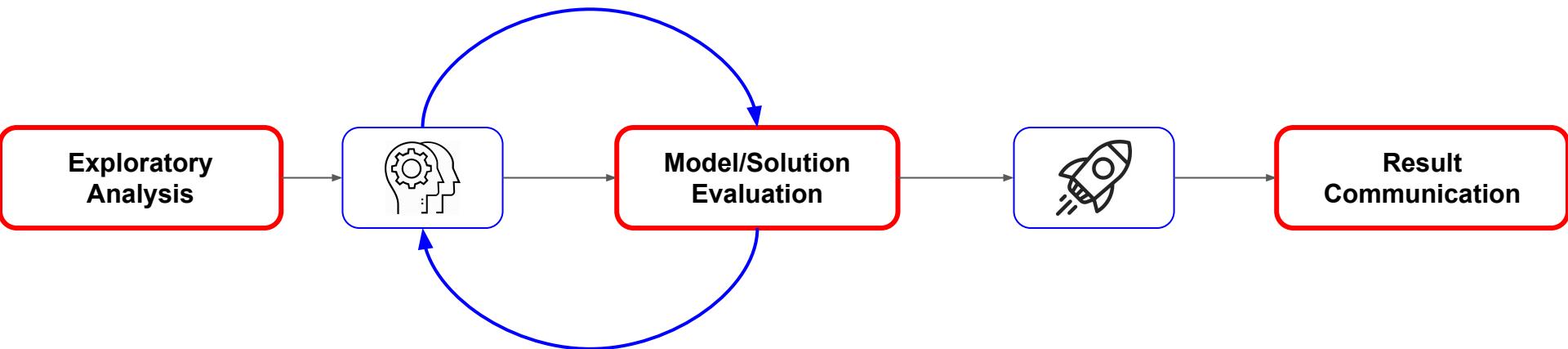
Dataviz? Why? Where?

A Ultra Super Simple Data Science Process



Dataviz? Why? Where?

A Ultra Super Simple Data Science Process



Dataviz? Why? Where?

Data Science Processes

Team Data Science Process Lifecycle [1]

CRISP-DM [2]

KDD [3]

Dataviz? Why? Where?

Data Science Processes

Team Data Science Process Lifecycle [1]

CRISP-DM [2]

KDD [3]

The screenshot shows a blog post on a website. At the top, there's a navigation bar with links for 'Sign in', 'Get started', and several categories: 'QUEM É PAULO VASCONCELLOS?', 'PORTFÓLIO', 'NOTÍCIAS', 'DICAS E TUTORIAIS', and 'CONHEÇA O DATA HACKERS!'. Below the navigation, the author's profile picture and name 'Paulo Vasconcellos' are displayed, along with a brief bio: 'Bem-vindo ao blog onde você encontrará projetos e tutoriais de Ciência de Dados usando Data...'. There are also 'Follow' and 'Follow' buttons for social media. The main title of the post is 'CRISP-DM, SEMMA e KDD: conheça as melhores técnicas para exploração de dados'. Below the title, a snippet reads: 'Saiba como explorar e extrair valores de seus dados usando as três maiores metodologias usadas em Data Mining e EDA'. Another snippet below that says: 'Nem sempre é fácil retirar Insights de dados. As vezes, ficamos horas olhando para aquele CSV, aquela tabela de SQL ou aquele JSON maldito, esperando que uma informação nova e útil apareça do nada na nossa frente. Felizmente, técnicas de análise exploratória de dados foram desenvolvidas no decorrer dos anos, desde que o termo foi cunhado em 1977 por John Tukey, em seu livro "Exploratory Data Analysis".'. At the bottom, there's a call-to-action: 'Never miss a story from Paulo Vasconcellos, when you sign up for Medium. Learn more' with a 'GET UPDATES' button.

Dataviz? Why? Where?

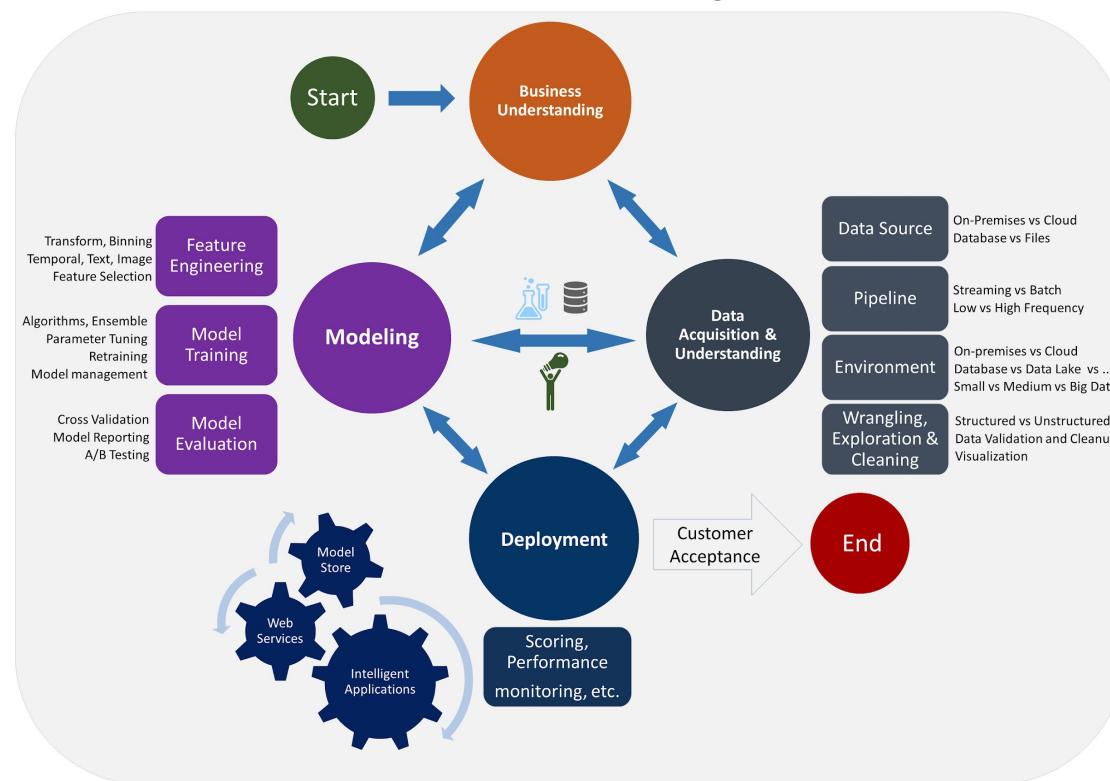
Data Science Processes

Team Data Science Process Lifecycle [1]

CRISP-DM [2]

KDD [3]

Data Science Lifecycle



Dataviz? Why? Where?

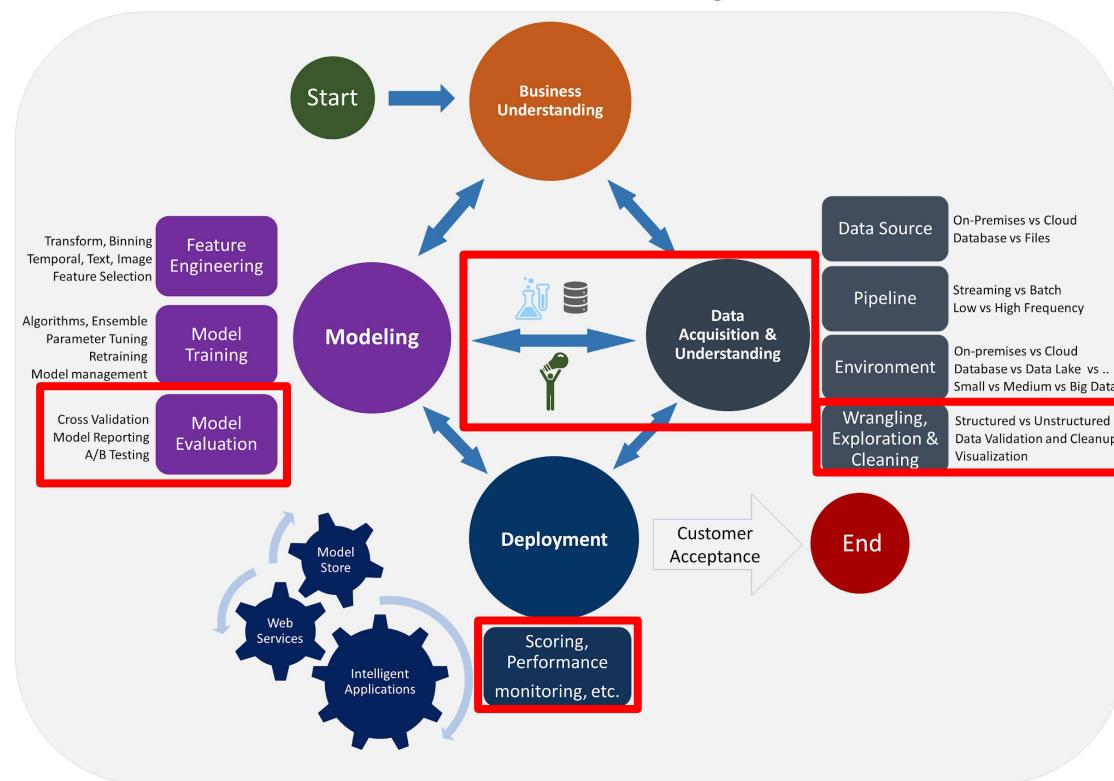
Data Science Processes

Team Data Science Process Lifecycle [1]

CRISP-DM [2]

KDD [3]

Data Science Lifecycle



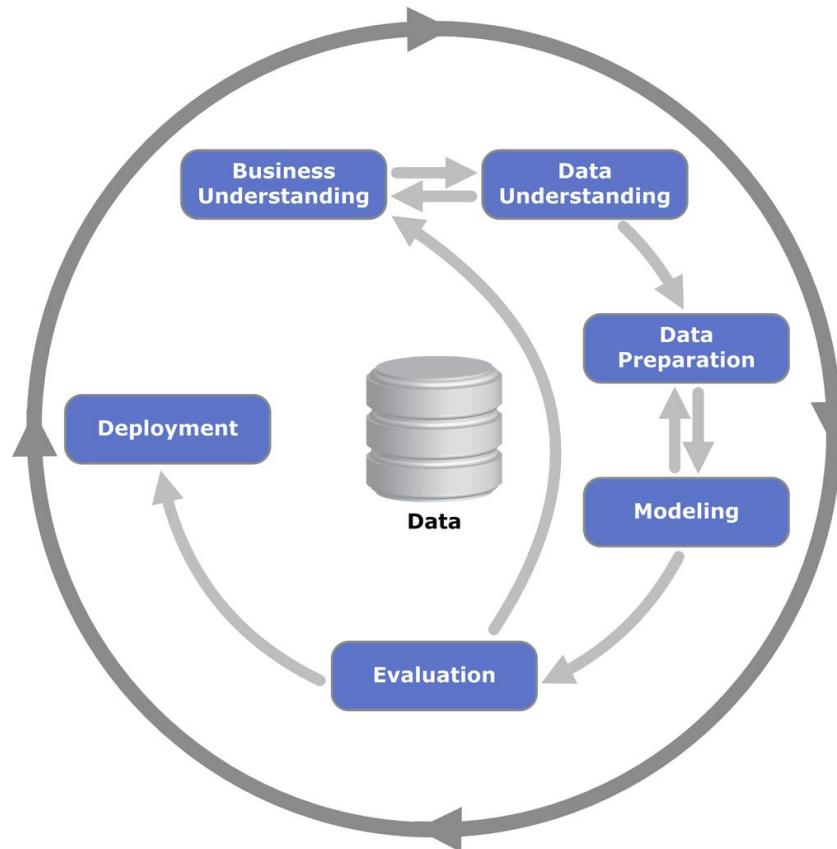
Dataviz? Why? Where?

Data Science Processes

Team Data Science Process lifecycle [1]

CRISP-DM [2]

KDD [3]



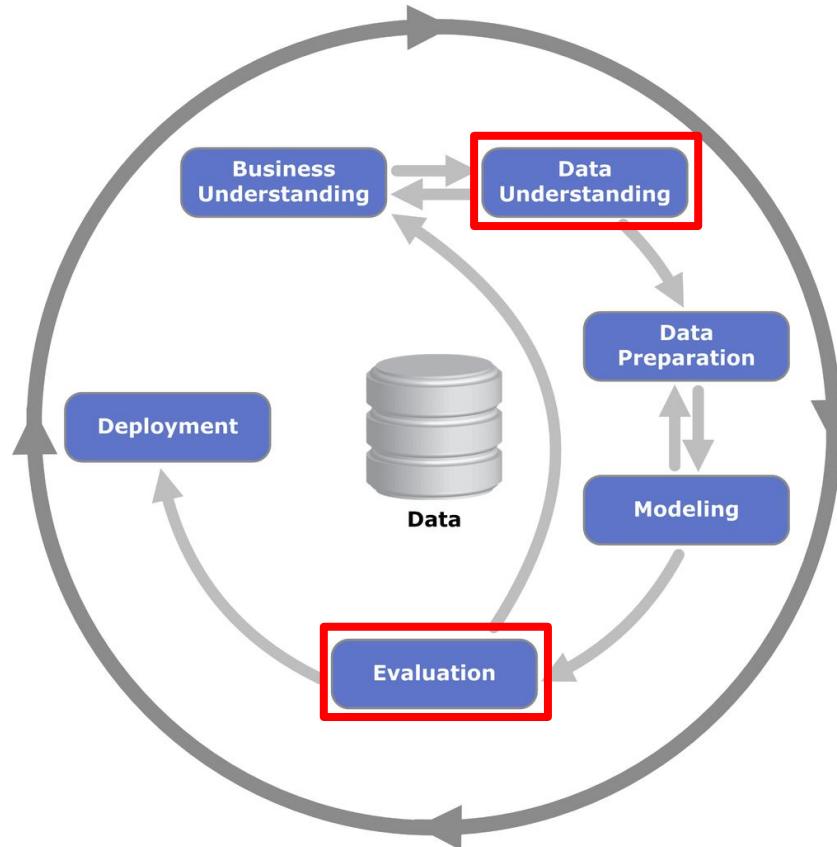
Dataviz? Why? Where?

Data Science Processes

Team Data Science Process lifecycle [1]

CRISP-DM [2]

KDD [3]



Dataviz? Why? Where?

Data Science Processes

Team Data Science Process lifecycle [1]

CRISP-DM [2]

KDD [3]

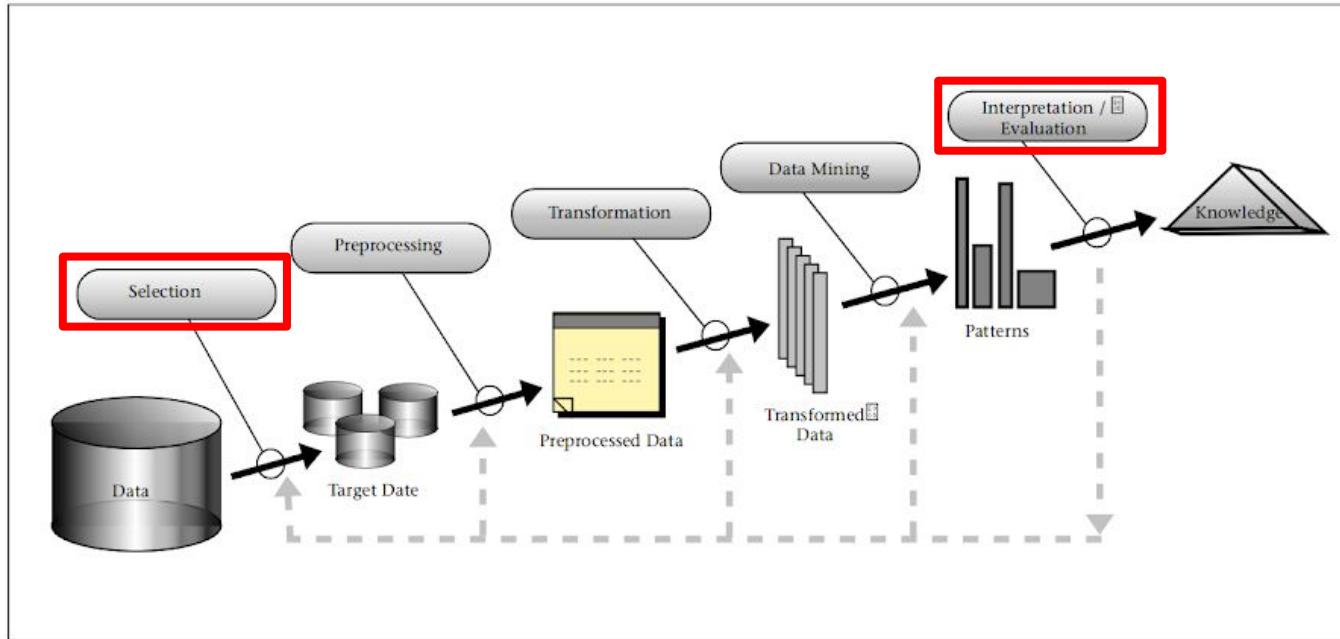
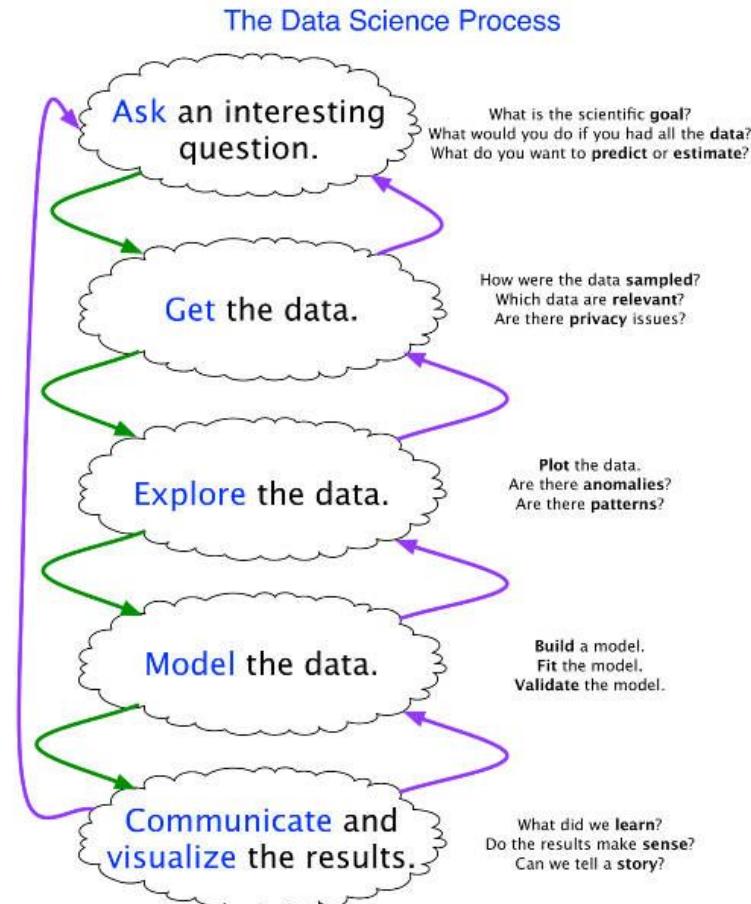
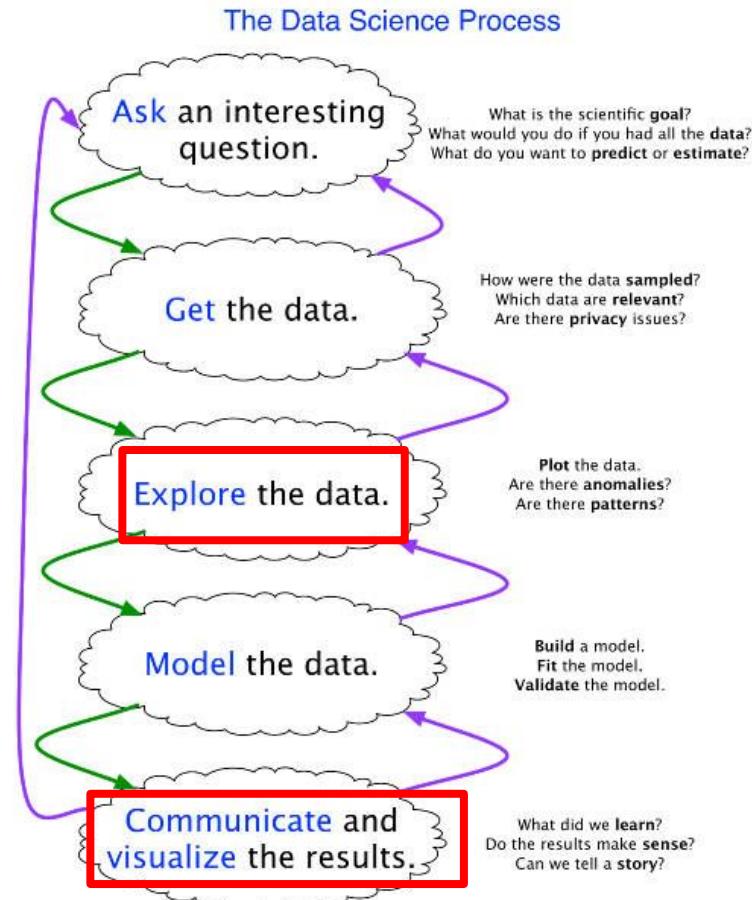


Figure 1. An Overview of the Steps That Compose the KDD Process.

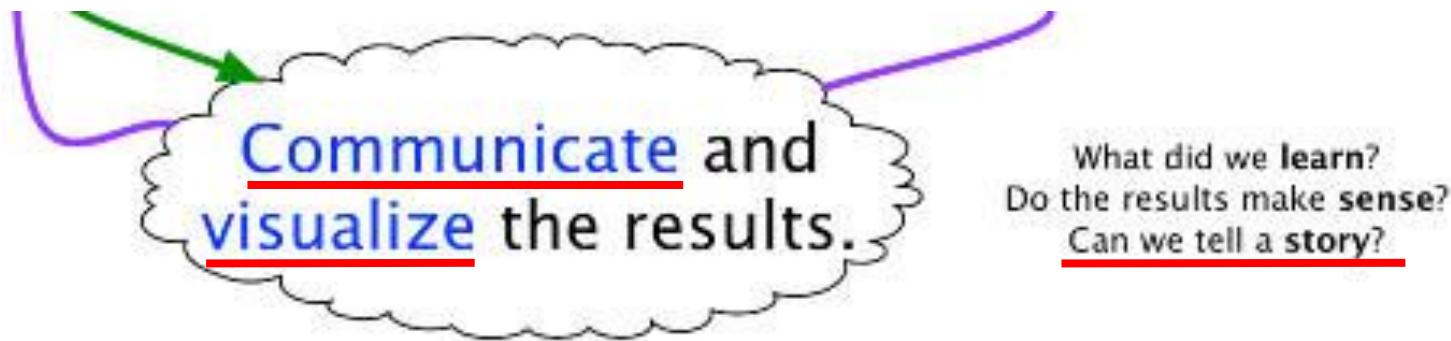
Dataviz? Why? Where?



Dataviz? Why? Where?



Dataviz? Why? Where?



Dataviz? Why? Where?

Data Science

Dataviz? Why? Where?

Science

“knowledge about a phenomena obtained and tested through scientific method” [4]

systematic pursuit of knowledge involving the recognition and formulation of a problem,
the collection of data through observation and experiment,



Dataviz? Why? Where?

Science

“knowledge about a phenomena obtained and tested through scientific method” [4]

systematic pursuit of knowledge involving the **recognition** and **formulation** of a **problem**,
the **collection of data** through **observation** and **experiment**,
the formulation, analyzing and **testing** of **hypotheses**,
and the **communication** of the results back to the scientific community



Dataviz? Why? Where?

Science



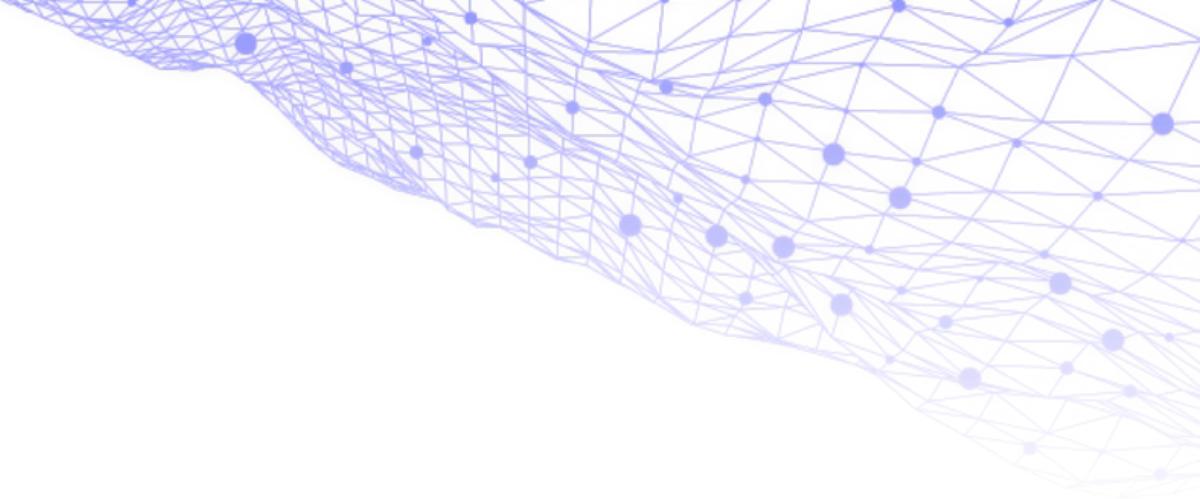
Dataviz? Why? Where?

Method!



Science works because it is systematic, reproducible, evidence-based.

Data Science is all about applying such **fundamentals** to problems in companies, governments etc.



What is and why we need dataviz?

What is / why dataviz?

- Our brain is not that good in extracting useful information from numbers
- Visual structures have the power of summarizing large collections/list of numbers
- Good visual structures lead to “faster” conclusions

What is / why dataviz?

“Data visualization provides a powerful way to communicate a data-driven finding. In some cases, the visualization is so convincing that no follow up analysis is required.”

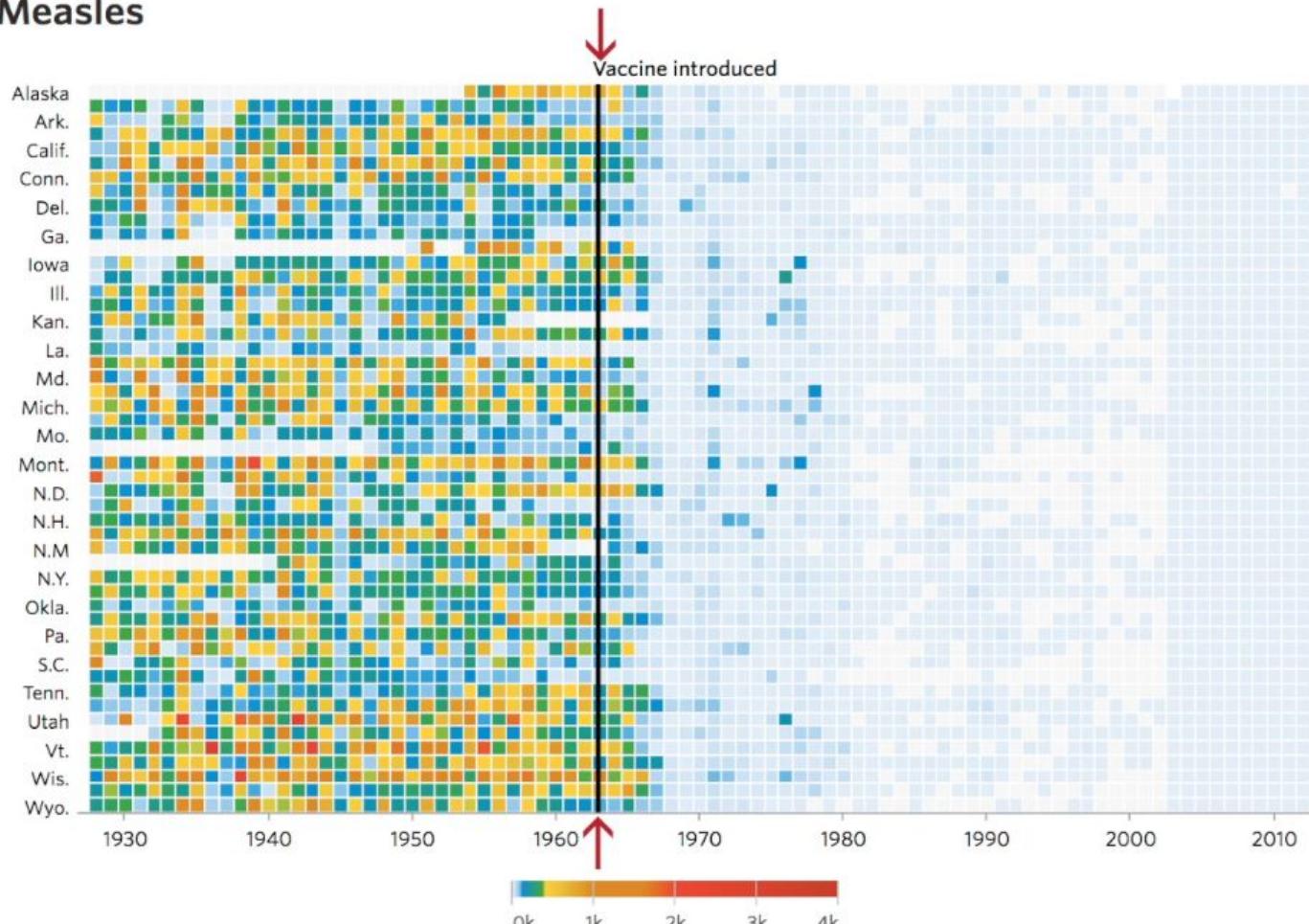
Prof. Rafael Irizarry, Professor of Applied Statistics at Harvard

“The greatest value of a picture is when it forces us to notice what we never expected to see.”

John Tukey, mathematician (FFT algorithm, box plot etc.)

What is dataviz - Argument Highlighting

Measles

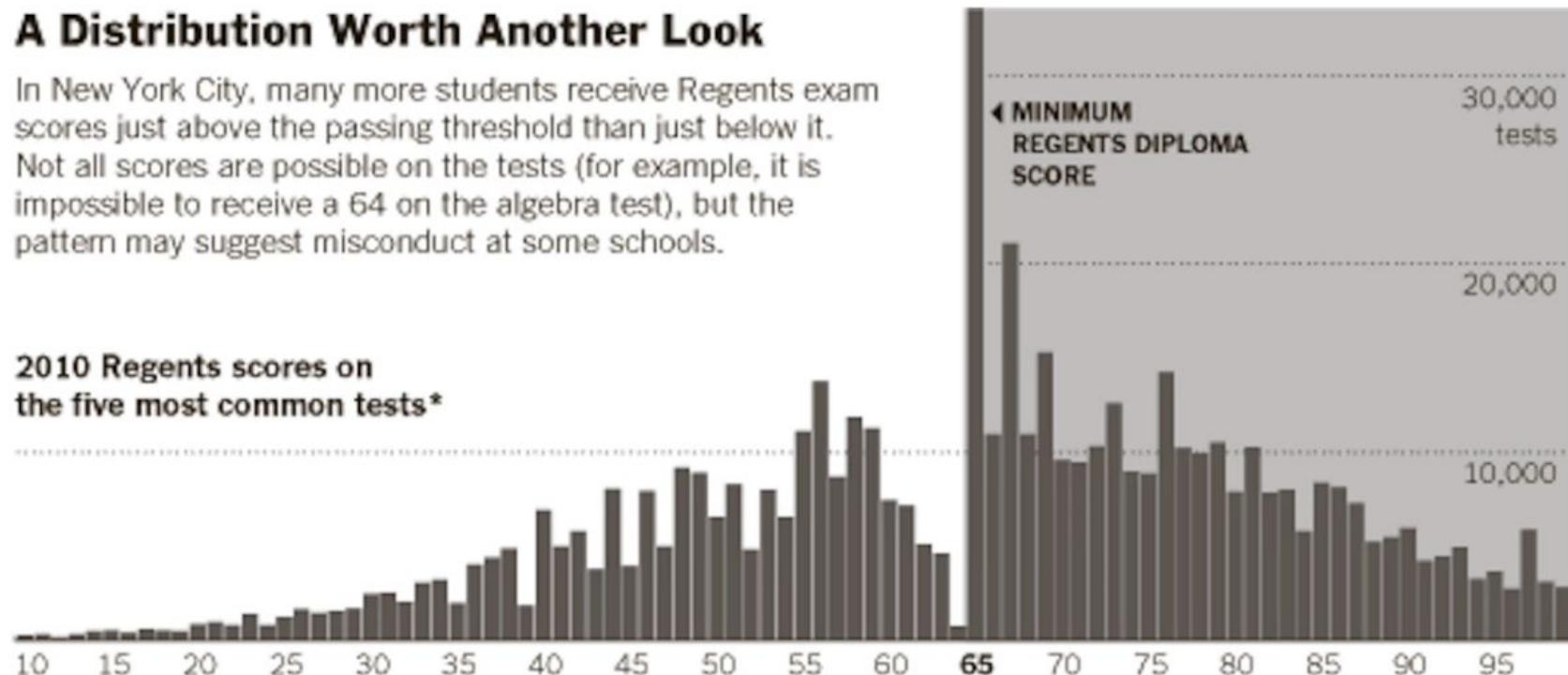


What is dataviz - Problem Statement

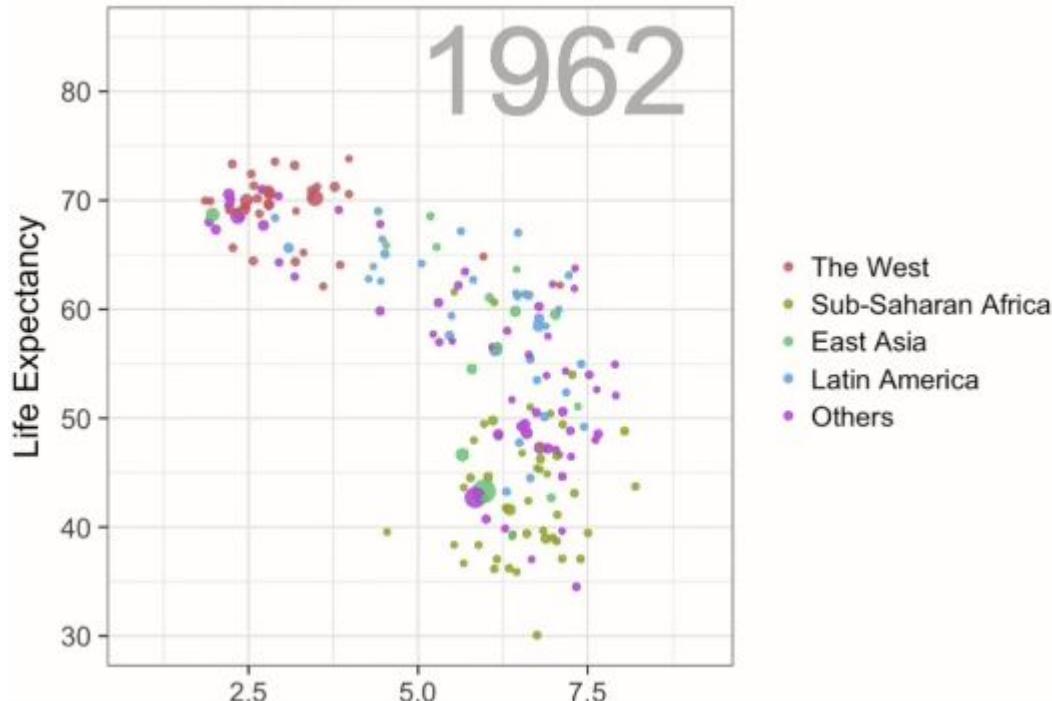
A Distribution Worth Another Look

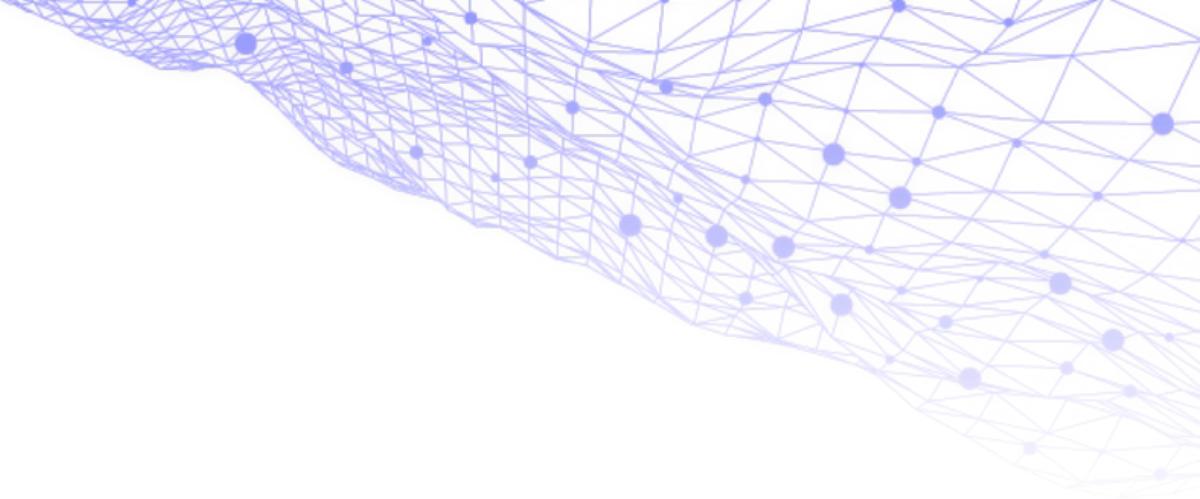
In New York City, many more students receive Regents exam scores just above the passing threshold than just below it. Not all scores are possible on the tests (for example, it is impossible to receive a 64 on the algebra test), but the pattern may suggest misconduct at some schools.

2010 Regents scores on
the five most common tests*



What is dataviz - Summarization / Tendency



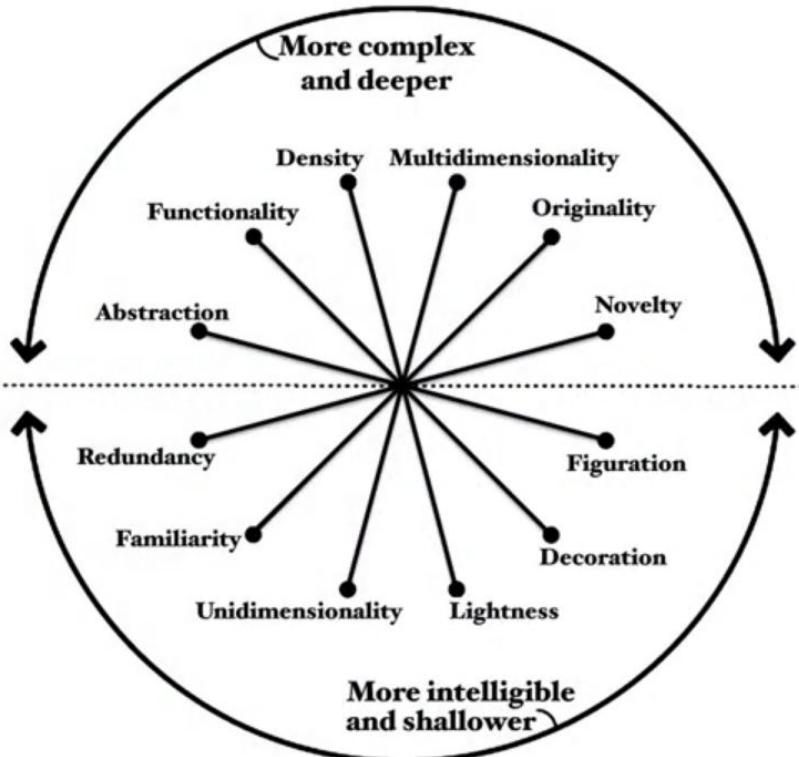


Tools for Thinking About Design

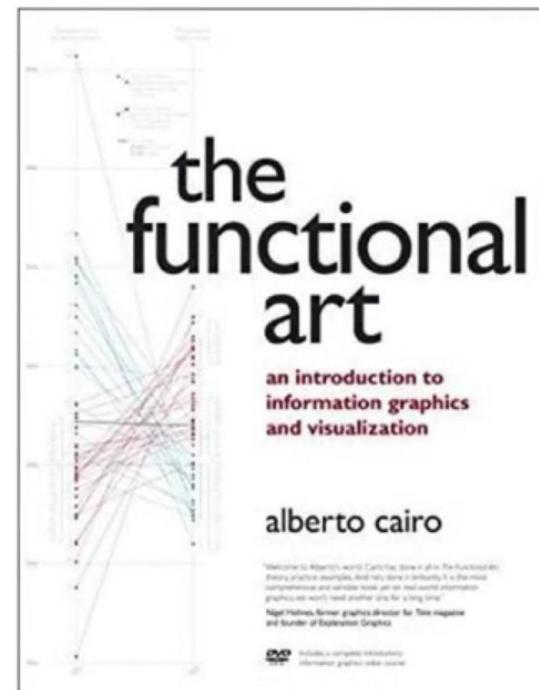
The visualization Wheel

The Visualization Wheel

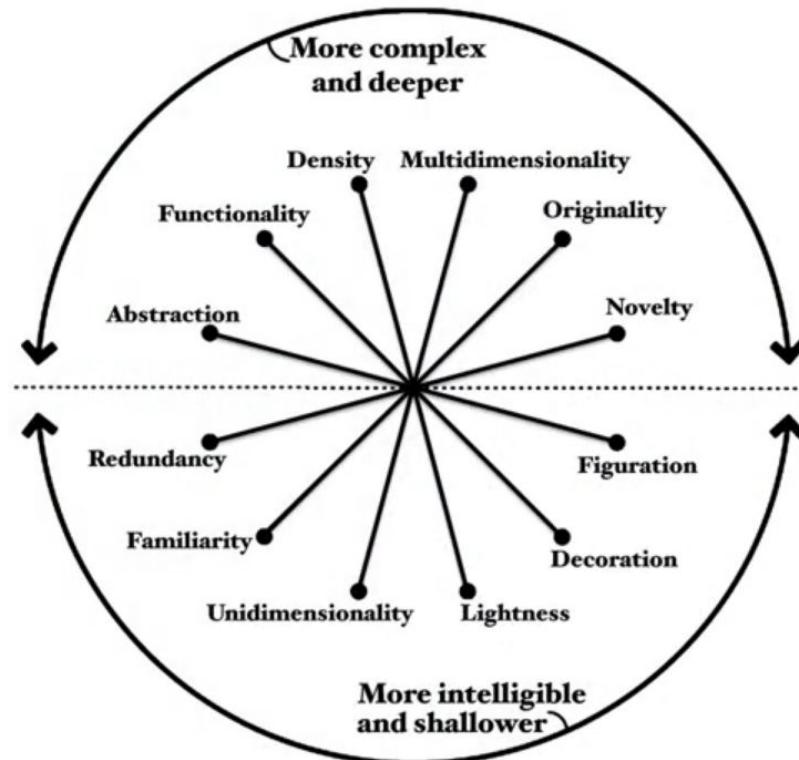
Thinking about design tradeoffs when building information graphics



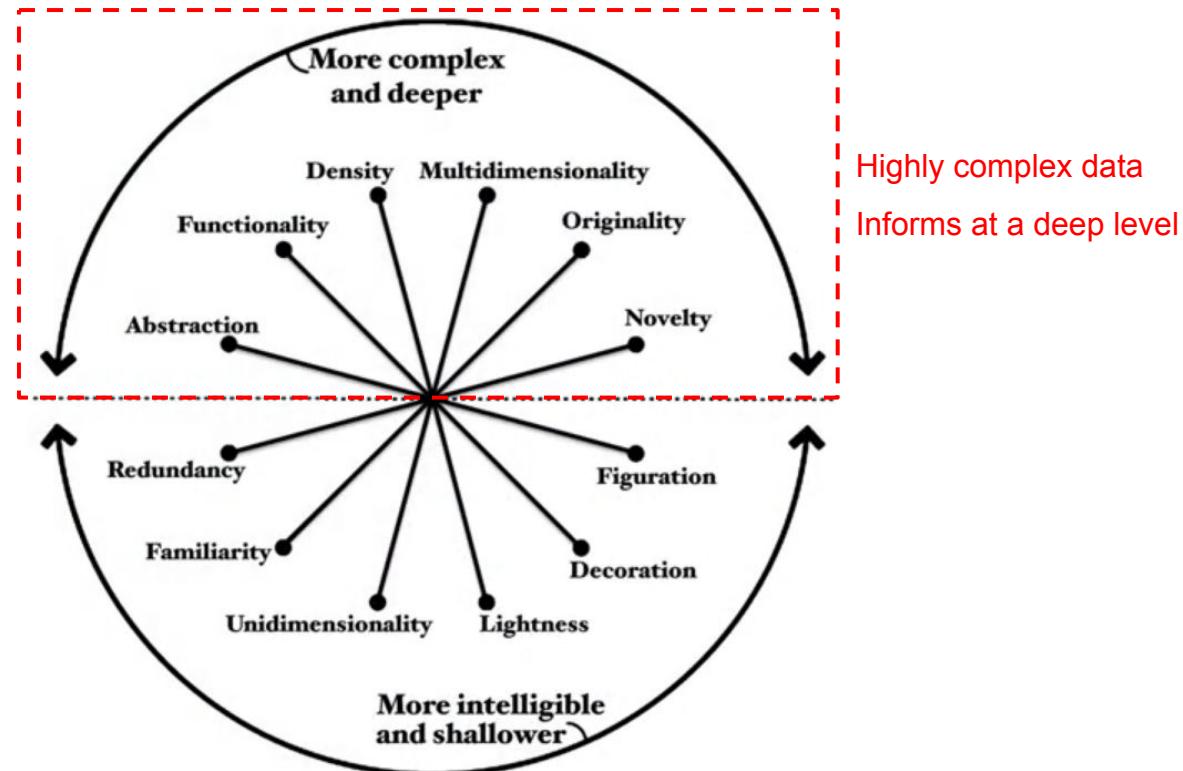
- Alberto Cairo; the functional art



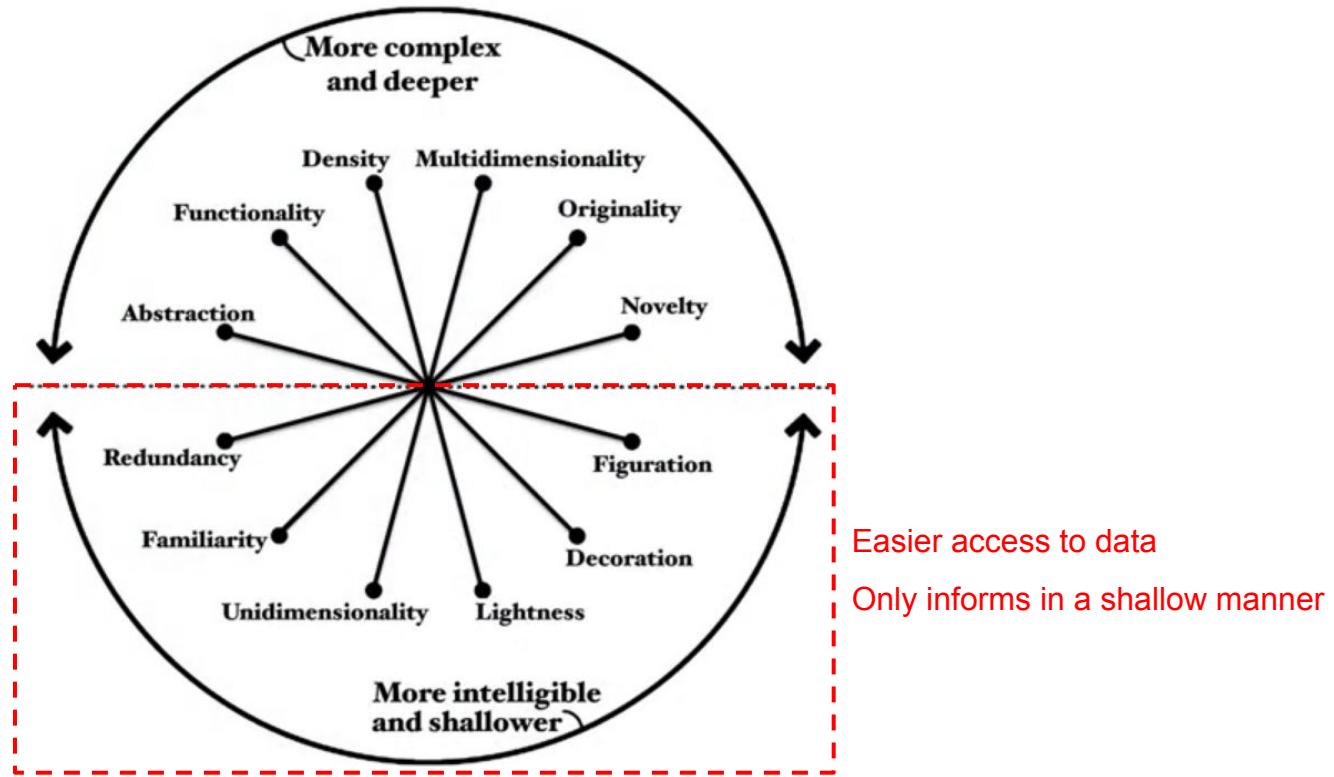
The Visualization Wheel



The Visualization Wheel

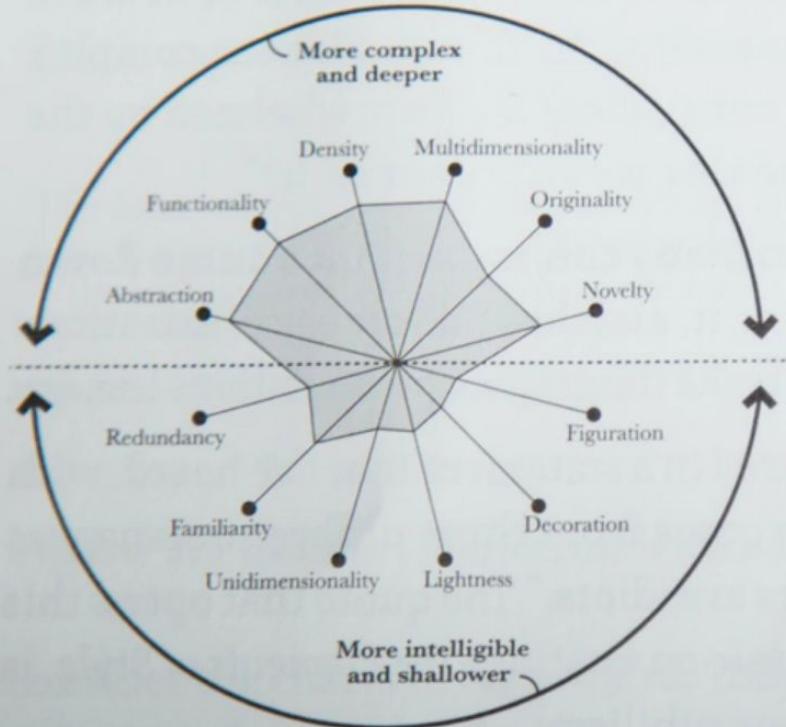


The Visualization Wheel



The Visualization Wheel

The wheel preferred by scientists and engineers



The wheel favored by artists, graphic designers, and journalists

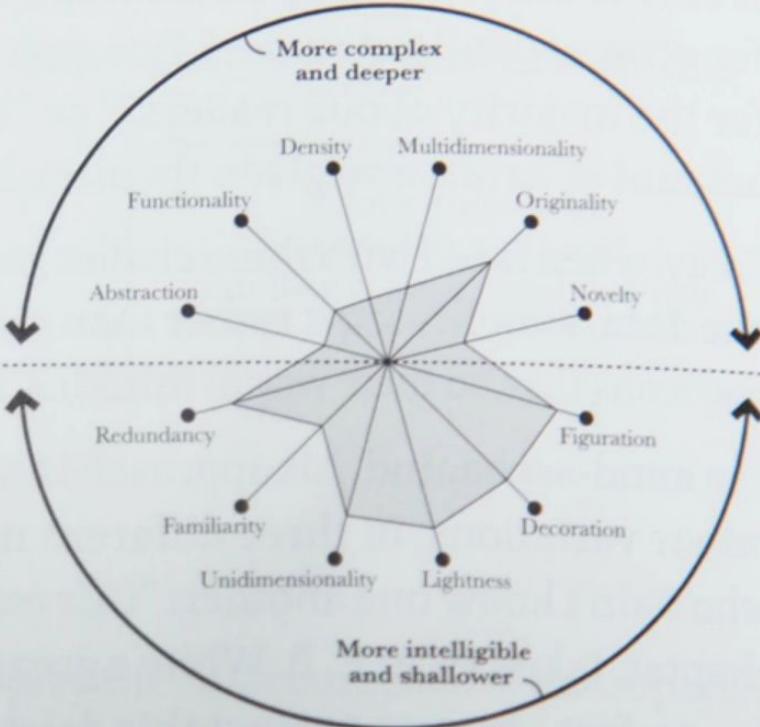
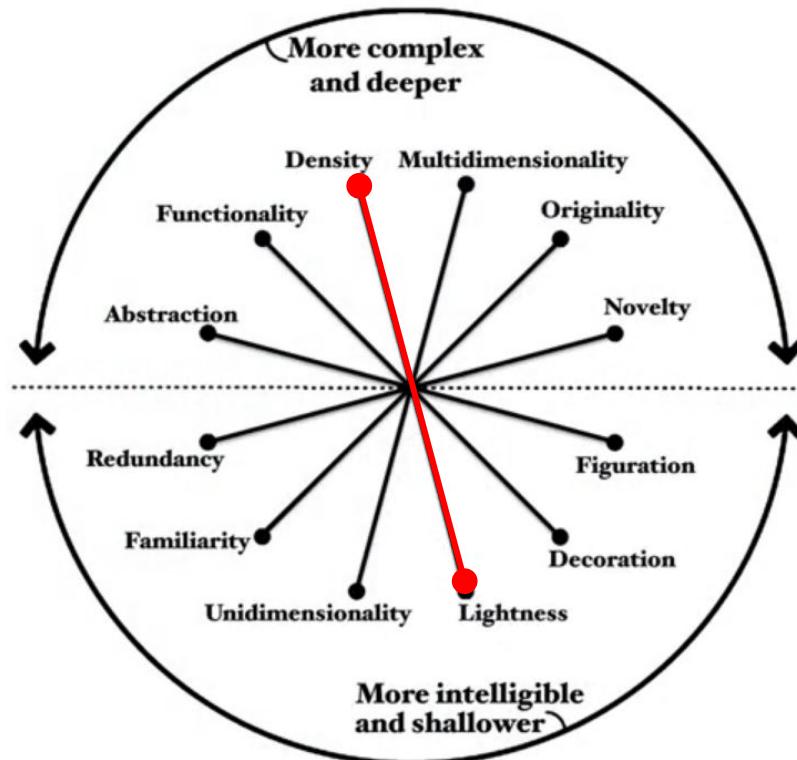


Figure 3.11 Different professional backgrounds, different ways of facing projects.

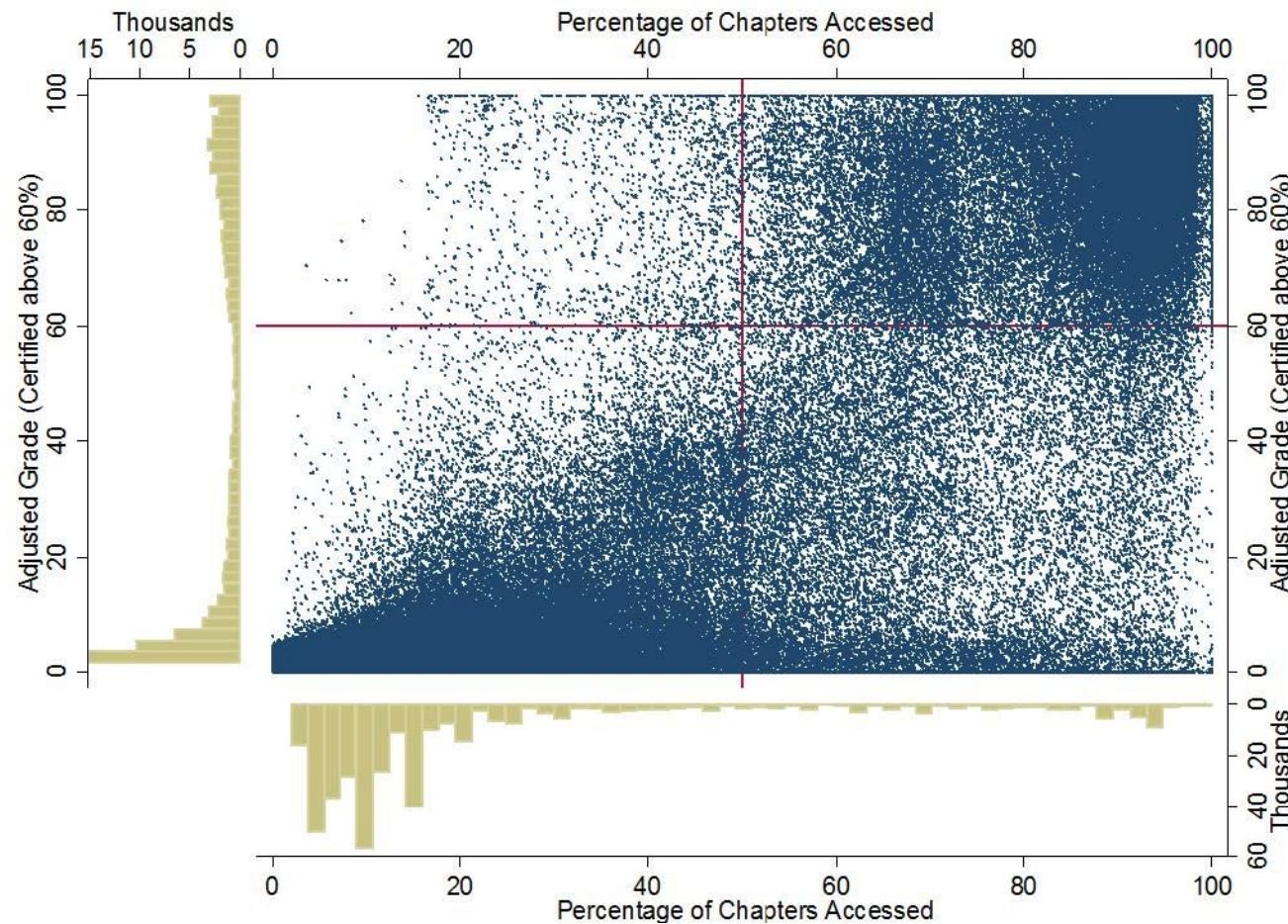
The Visualization Wheel - Tradeoffs

Density <> Lightness: Must be studied in depth (density) or understandable at first sight (lightness)



The Visualization Wheel - Tradeoffs

Density <> Lightness: Must be studied in depth (density) or understandable at first sight (lightness)



The Visualization Wheel - Tradeoffs

Density <> Lightness: Must be studied in depth (density) or understandable at first sight (lightness)

P20

NYT

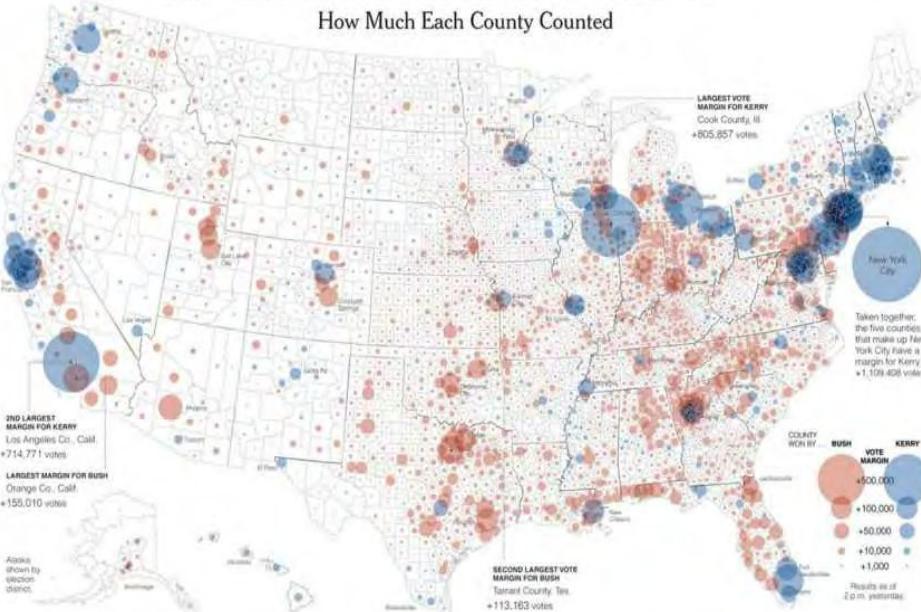
THE NEW YORK TIMES, THURSDAY, NOVEMBER 4, 2004

THE 2004 ELECTIONS
THE ELECTORAL MAP

Red and Blue, the Divided Electorate, in All Its Shades

The simple formula for winning an election is to get more votes than your opponent in as many counties as possible. It worked for President Bush. He won some of those states. However, Mr. Bush's relatively smaller but consistent margins in suburban and rural counties, in much of the South and West, helped him overcome Mr. Kerry's urban-county margins.

How Much Each County Counted



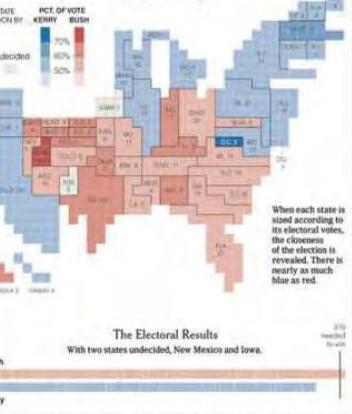
Two Views of the Electoral Vote

By Geography

Using the typical map of the United States, vast swaths are painted red.



By Electoral Weight



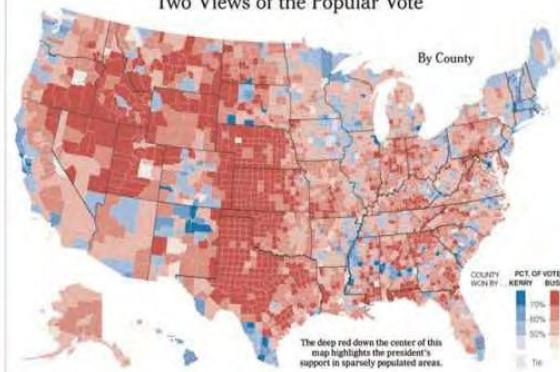
The Electoral Results

With two states undecided, New Mexico and Iowa.



Two Views of the Popular Vote

By County



The deep red down the center of this map highlights the president's support in sparsely populated areas.

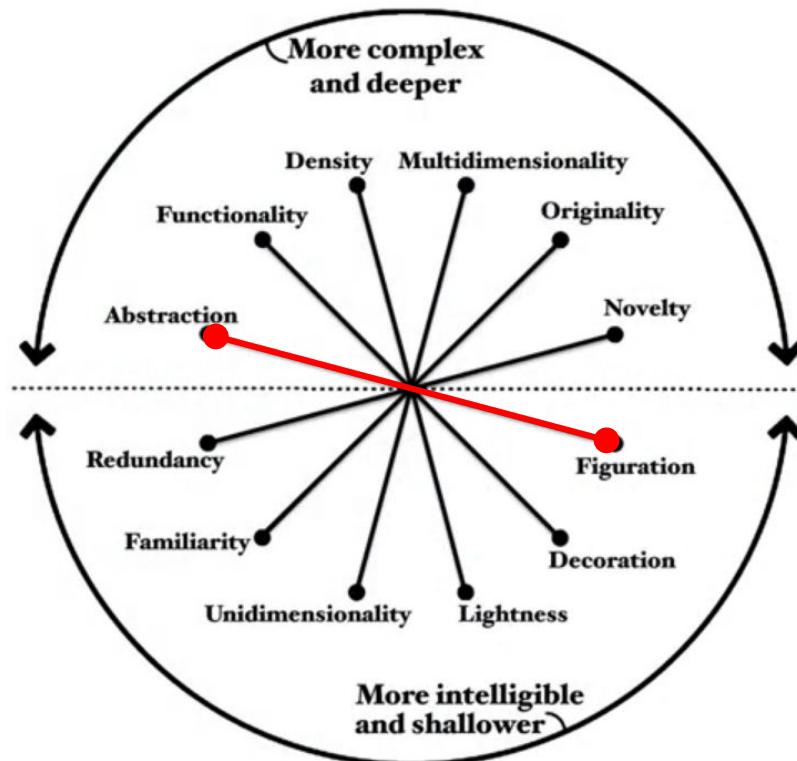
By Population Density



This map removes mostly uninhabited areas, revealing Mr. Bush's suburban and rural support in the East and South.

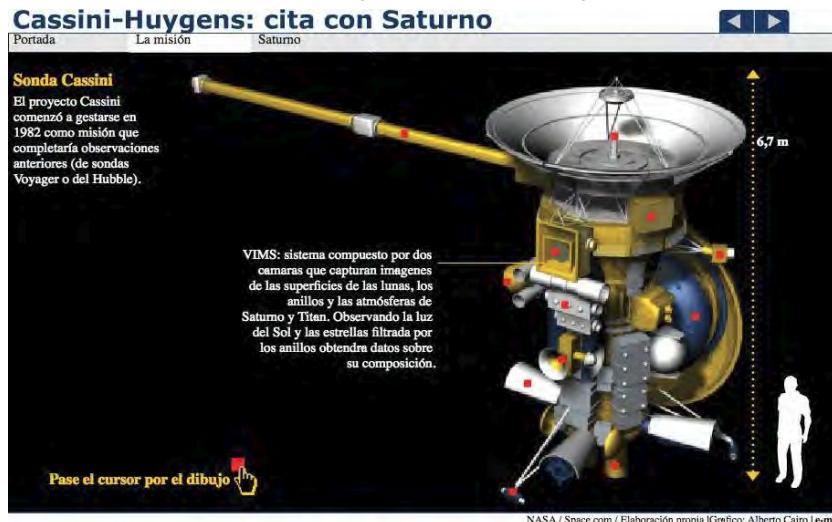
The Visualization Wheel - Tradeoffs

Abstraction <>> Figuration: Boxes and charts (abstractions) or real-world drawings/images (figuration)



The Visualization Wheel - Tradeoffs

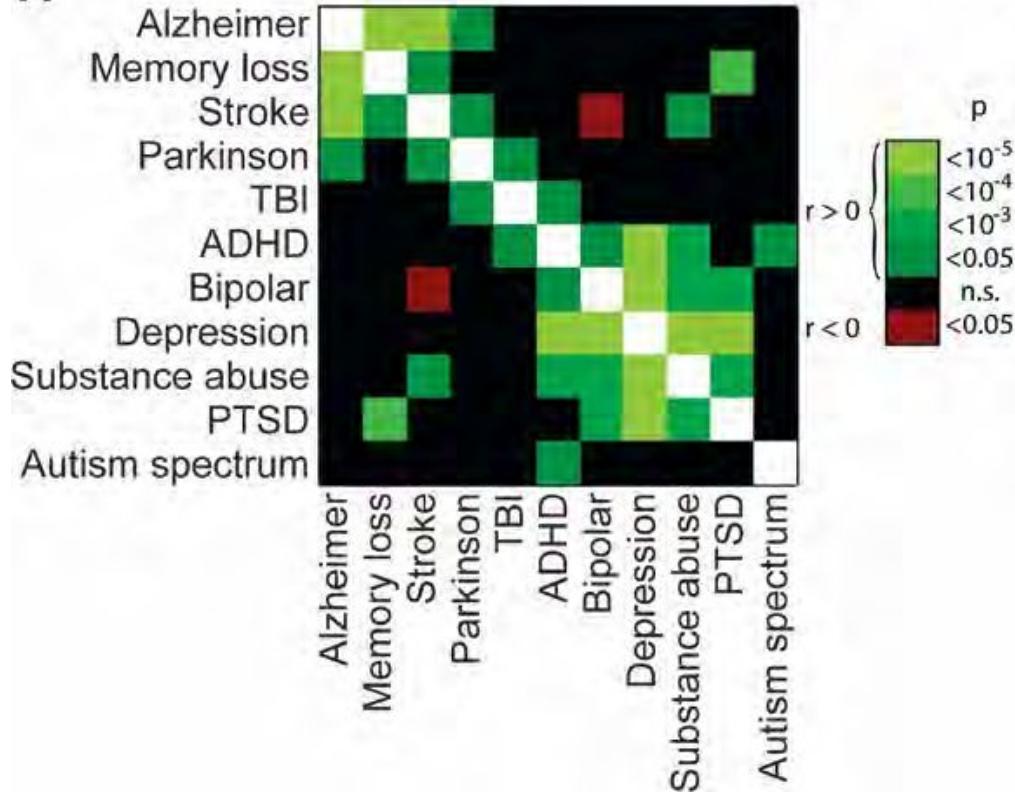
Abstraction <> Figuration: Boxes and charts (abstractions) or real-world drawings/images (figuration)



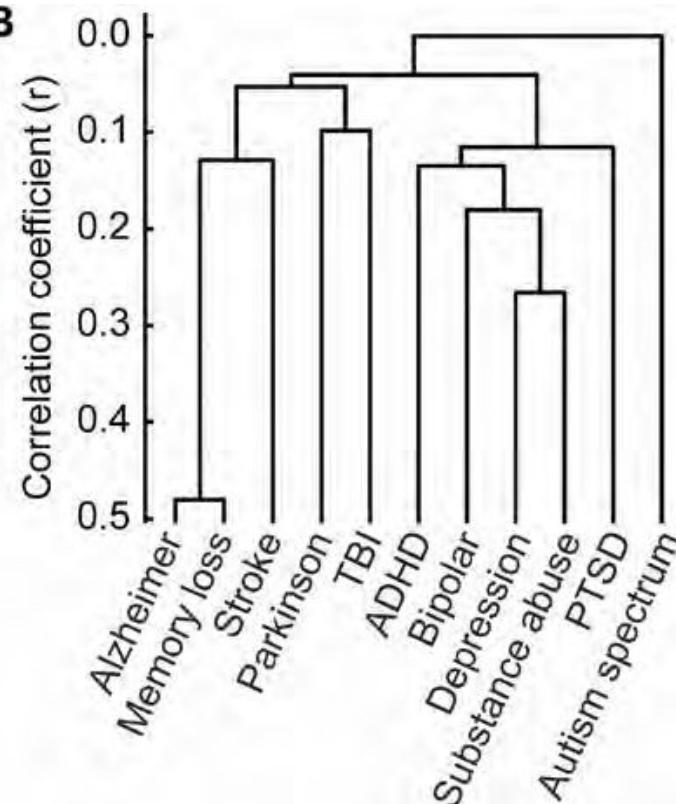
The Visualization Wheel - Tradeoffs

Abstraction <> Figuration: Boxes and charts (abstractions) or real-world drawings/images (figuration)

A

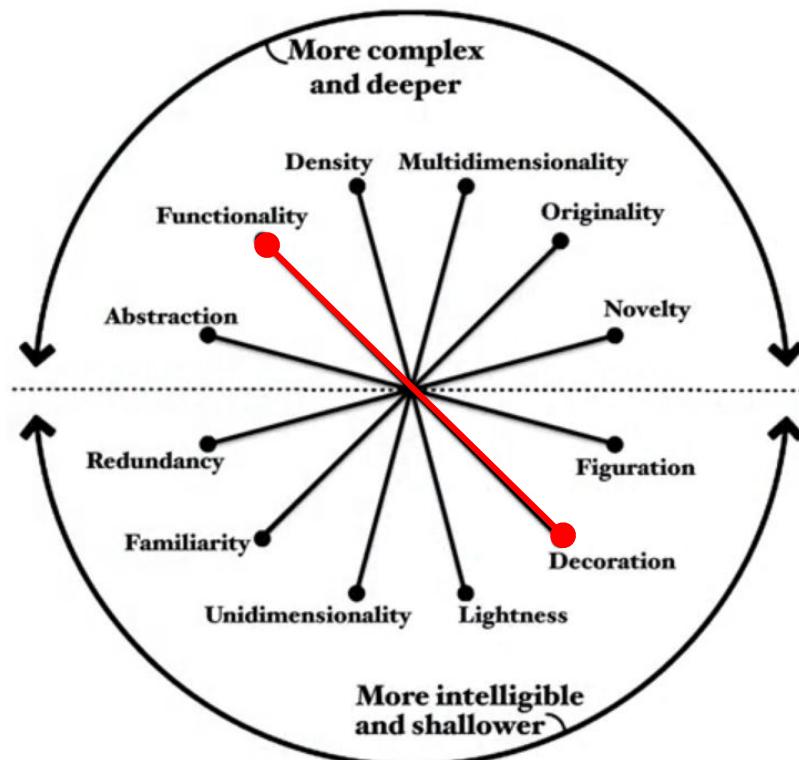


B



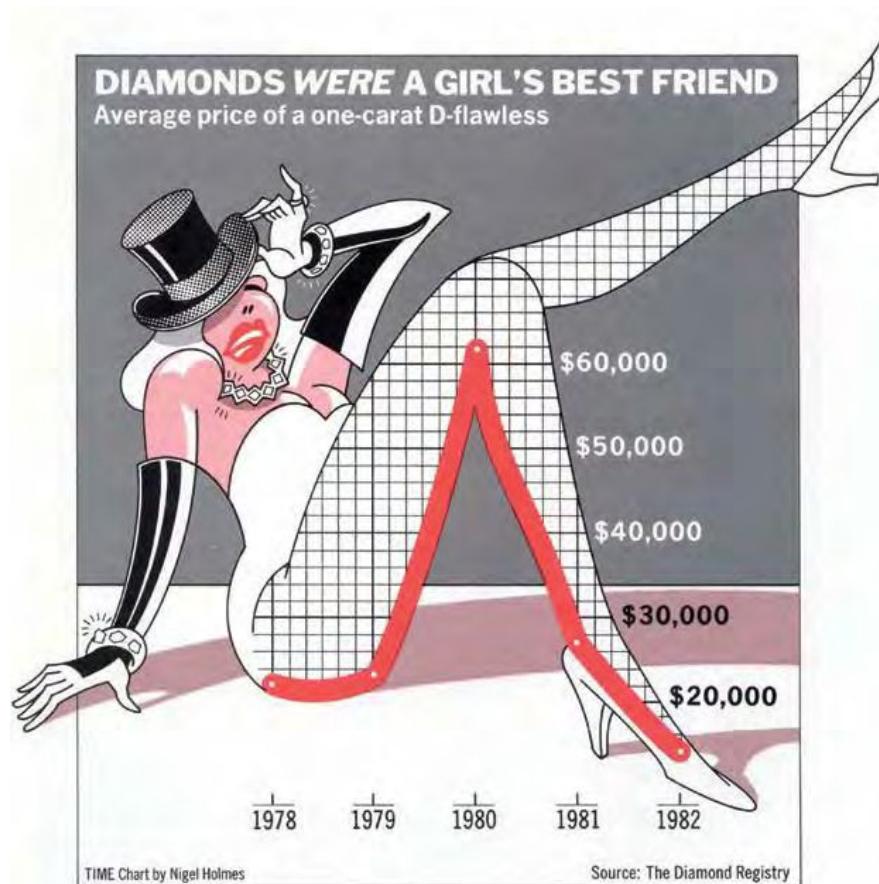
The Visualization Wheel - Tradeoffs

Functionality <> Decoration: No embellishments (focus on functionality) or artistic details (decoration)



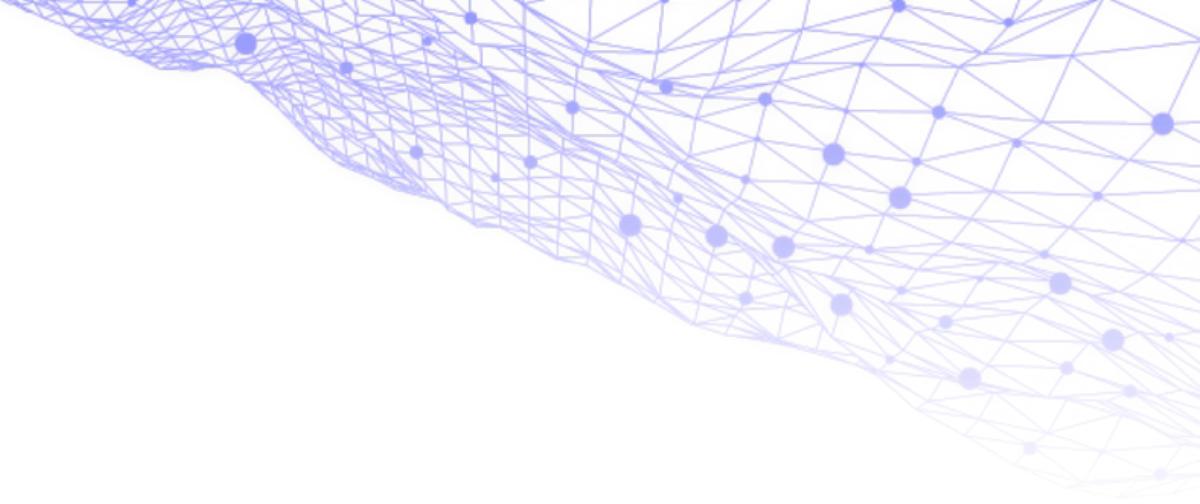
The Visualization Wheel - Tradeoffs

Functionality <> Decoration: No embellishments (focus on functionality) or artistic details (decoration)



The Visualization Wheel - Use it for better communication

- Know your audience, then use the wheel to evaluate your approach
- Know your problem, then use the wheel to evaluate your approach
- Let the wheel help your heuristics (practical decisions)



Graphic Heuristics: The Data-ink Ratio (Edward Tufte)

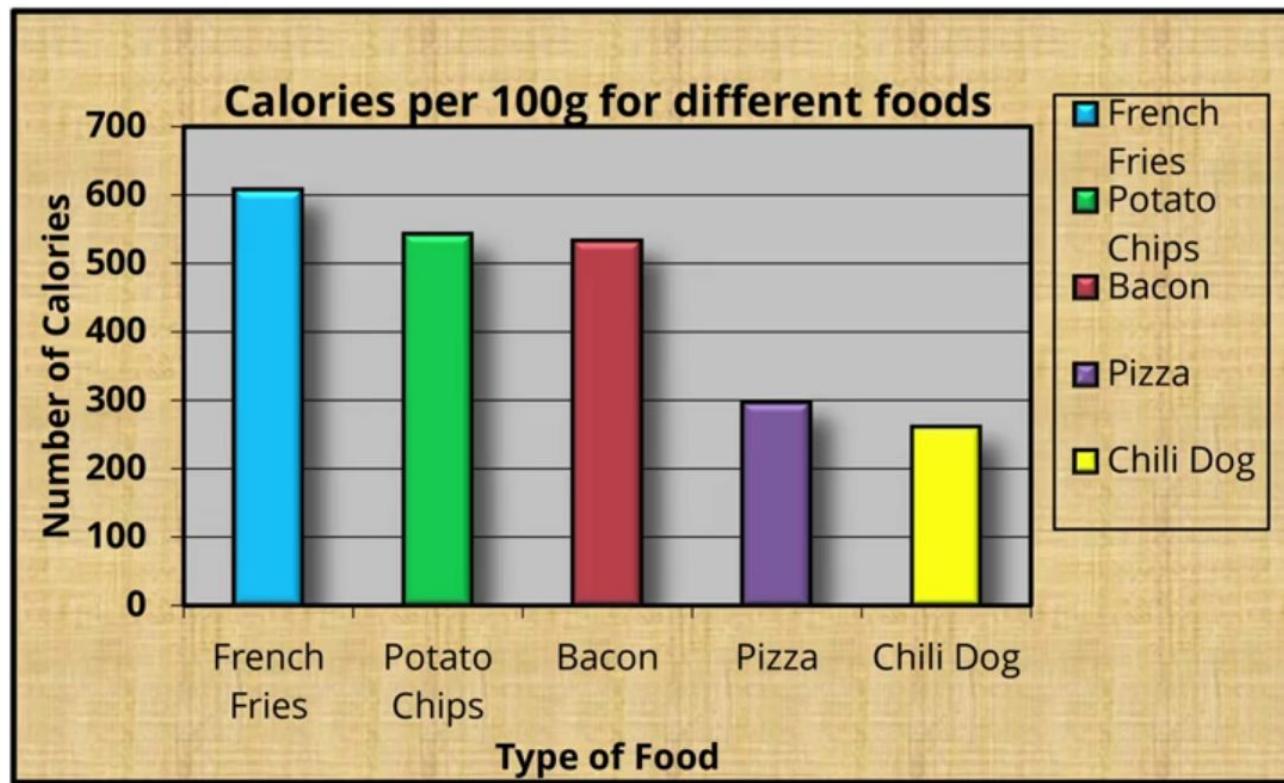
Remove to Improve

Data-ink Ratio

Remove the unnecessary

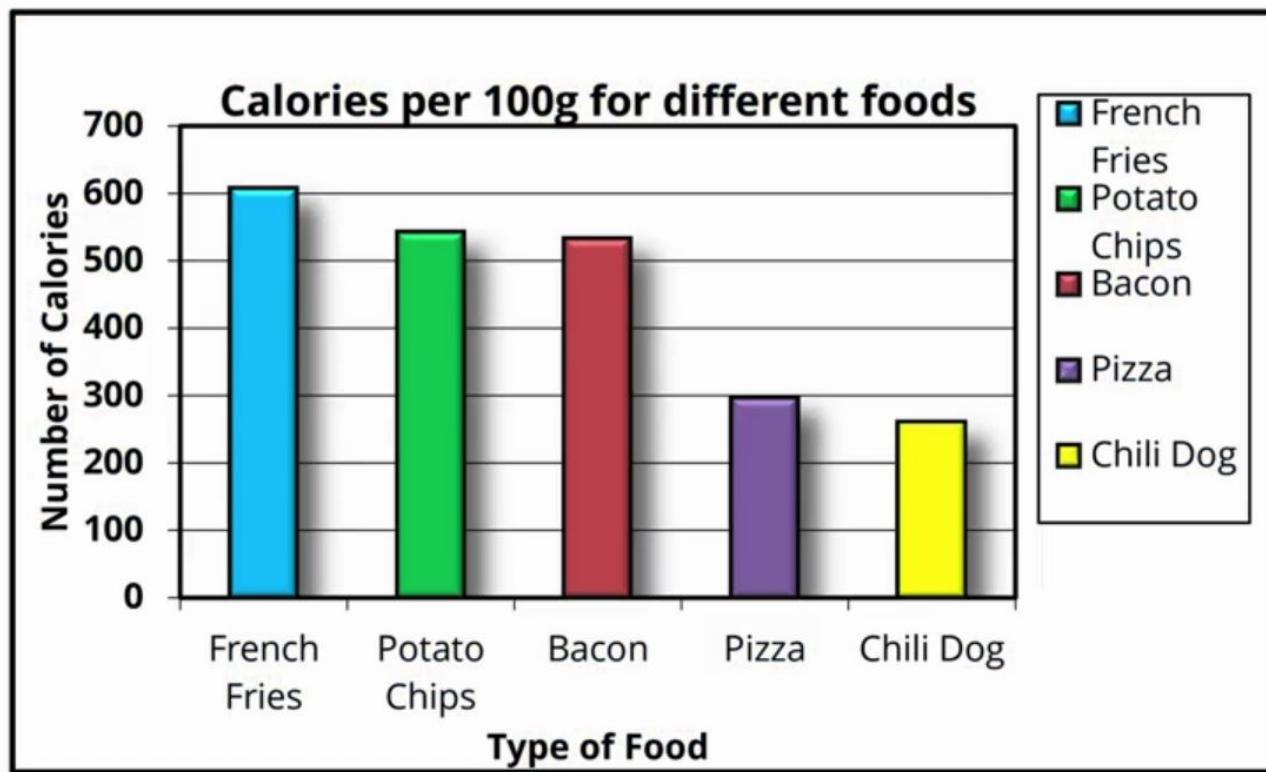
Data-ink Ratio

Remove the unnecessary - Backgrounds



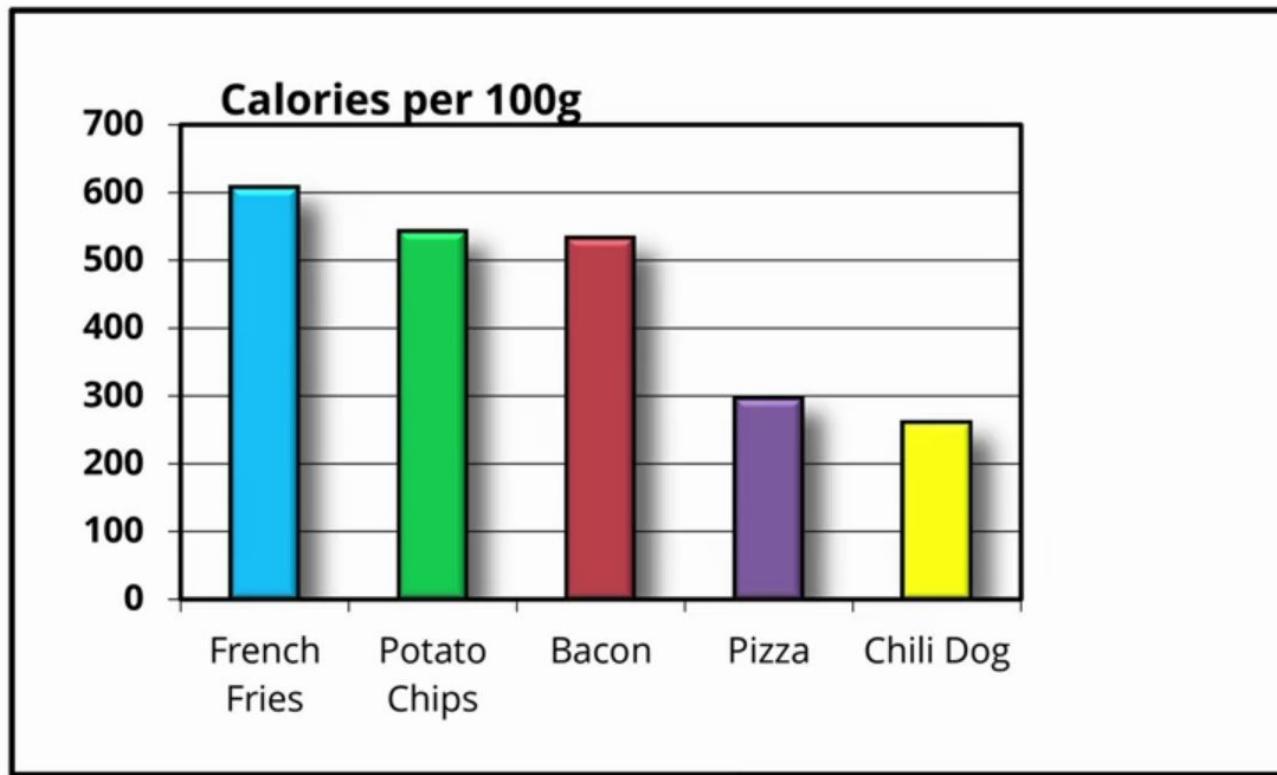
Data-ink Ratio

Remove the unnecessary - Unnecessary/redundant info



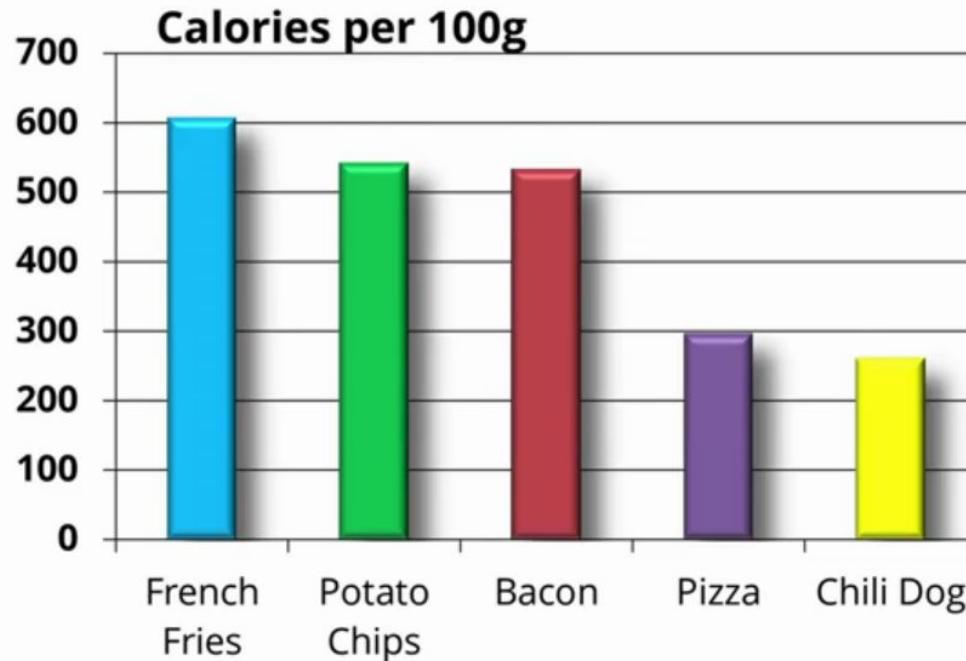
Data-ink Ratio

Remove the unnecessary - Borders



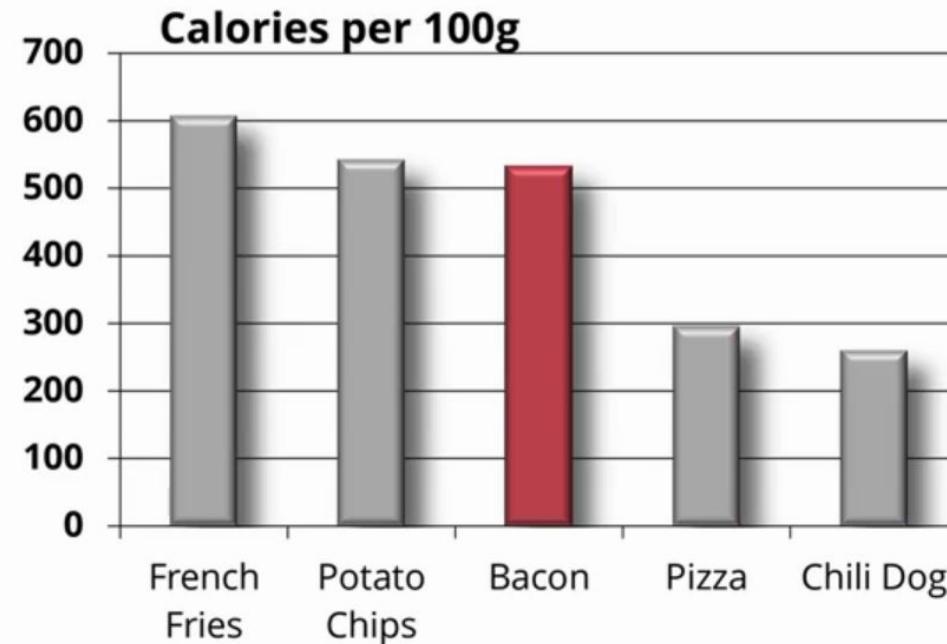
Data-ink Ratio

Remove the unnecessary - Colors



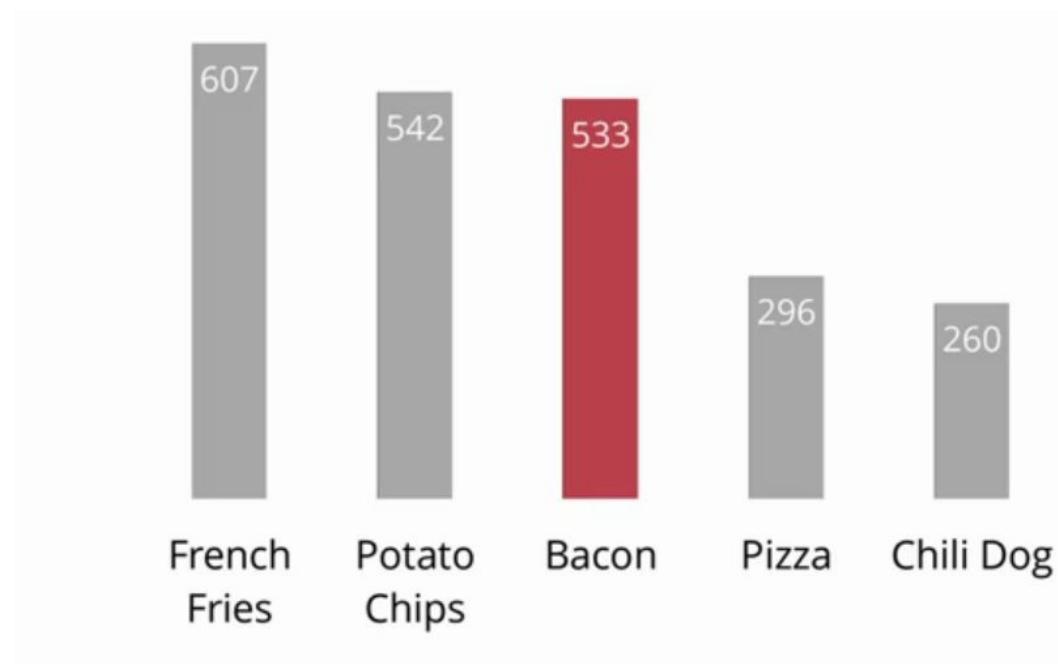
Data-ink Ratio

Remove the unnecessary - Special Effects



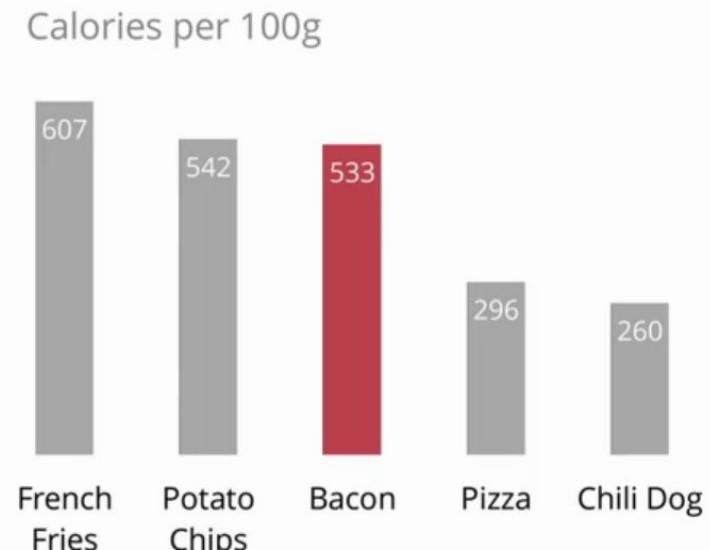
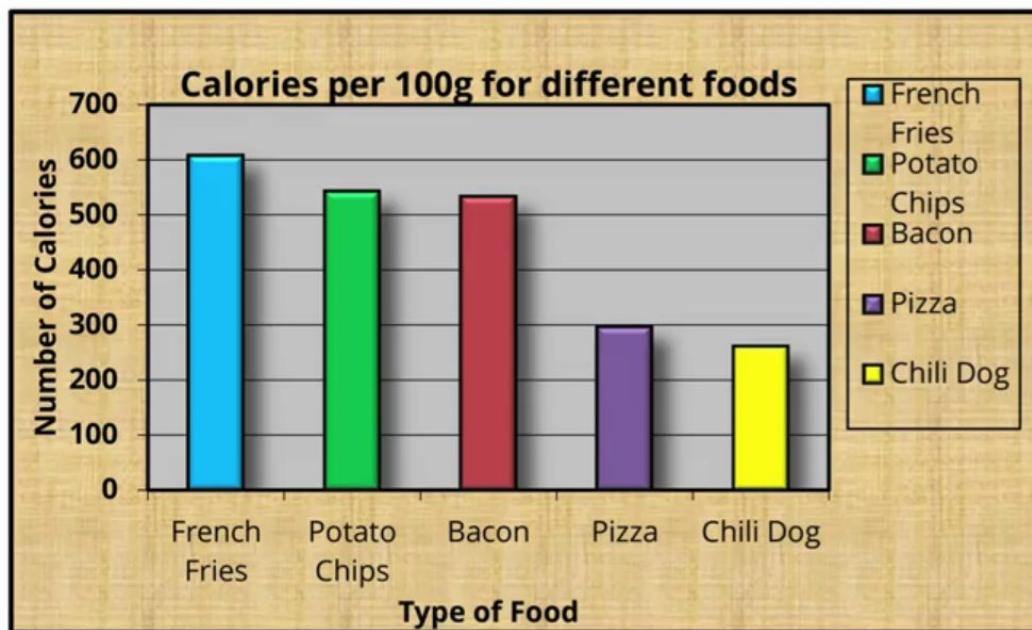
Data-ink Ratio

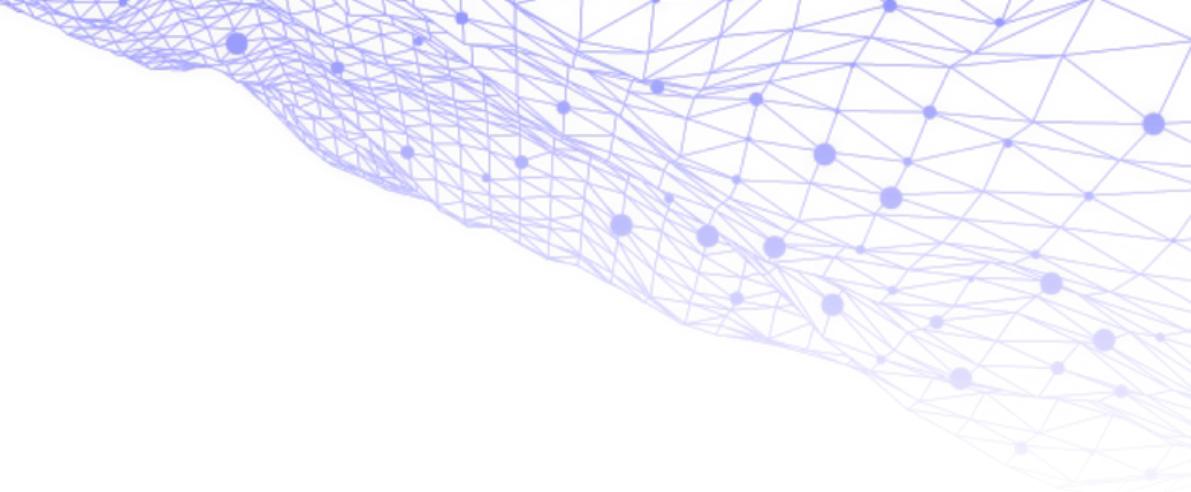
Remove the unnecessary - **Hidden Data**



Data-ink Ratio - Effective Communication

Remove the unnecessary





Lies and Honest Mistakes

Improving your criticism skills

Lies and Honest Mistakes

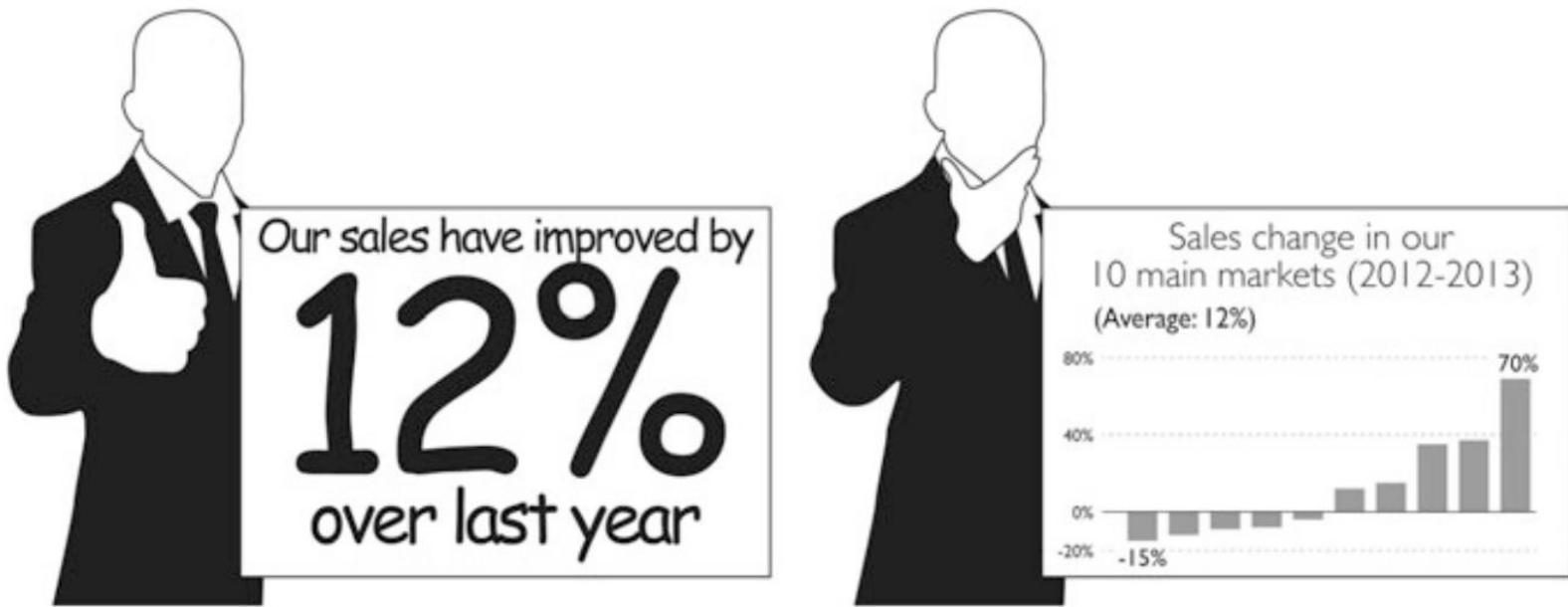


Fig. 5.1 Two versions of a fictional graphic. The first one displays just the average change, hiding the fact that the ten markets in which this company operates behaved very differently between 2012 and 2013. That reality is shown on the second version of the graphic. The average of all the values encoded in the bar graph is indeed 12 %

Lies and Honest Mistakes

PRESIDENTIAL ELECTIONS, 2013

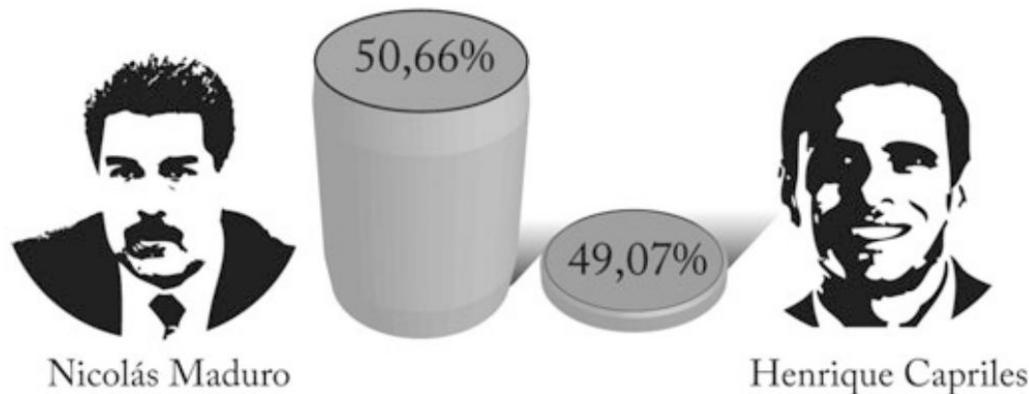


Fig. 5.3 Presidential election results in Venezuela, based on a graphic by Venezonala de Televisión. Notice the truncated Y-axis which greatly distorts the difference between the percentages of vote

Lies and Honest Mistakes

PRESIDENTIAL ELECTIONS, 2013



Fig. 5.3 Presidential election results in Venezuela, based on a graphic by Venezonala de Televisión. Notice the truncated Y-axis which greatly distorts the difference between the percentages of vote

PRESIDENTIAL ELECTIONS, 2013

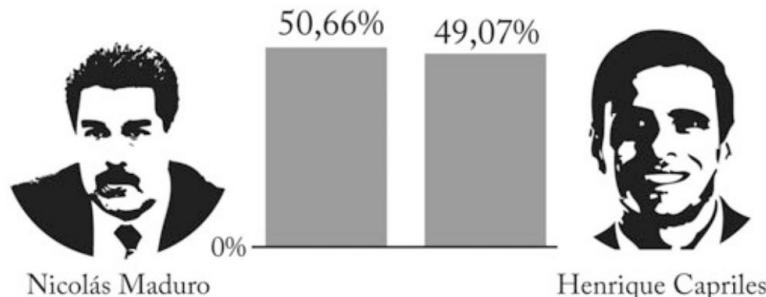


Fig. 5.4 An alternative version of the previous graphic in which a 0-baseline has been added, and the 3D effect has been removed

Lies and Honest Mistakes

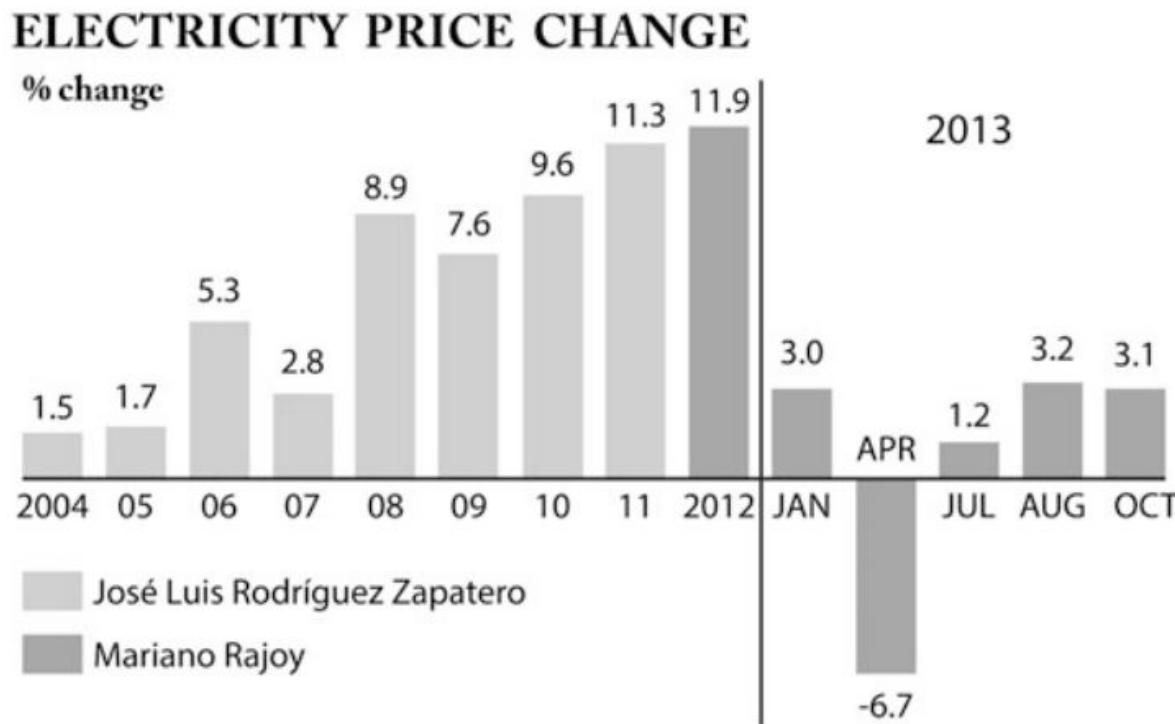


Fig. 5.7 Electricity price change, based on a graph by Spain's Partido Popular. Notice the inconsistent intervals on the X-axis

Conclusions

Qualities of a great visualization

- **Truthful:** do not mislead yourself nor your audience;
- **Functionality:** do not waste your time nor your audience's time;
- **Insightful:** empower multiple understandings with less.

Qualities and skills of great data scientists regarding dataviz:

- **Skepticism:** people, and therefore charts, lie / make mistakes;
- **Purpose-driven:** if there is no purpose on plotting or incrementing a plot, do not do it;
- **Know the tools:** delivering a message often depends on the dataviz tools you have.

References

- [1] <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>
- [2] [http://www.scirp.org/\(S\(vtj3fa45qm1ean45vvffcz55\)\)/reference/ReferencesPapers.aspx?ReferenceID=1592780](http://www.scirp.org/(S(vtj3fa45qm1ean45vvffcz55))/reference/ReferencesPapers.aspx?ReferenceID=1592780)
- [3] <https://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>
- [4] <https://www.merriam-webster.com/dictionary/science>