

Role of Visualization

Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

- Define data visualization
- Describe the value of data visualization for data science

“The ability to **take data**—to be able to **understand it**, to process it, to extract value from it. to **visualize it**, to communicate it—that’s going to be a **hugely important skill** in the next decades... Because now we really do have essentially free and ubiquitous data.” (emphasis mine)

- Hal Varian*, Google’s Chief Economist

* Interview with James Manyika. Hal Varian on how the Web challenges managers. McKinsey&Company. Oct. 2008.

Defining Visualization

“The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.” [Card et al., 1999]

“The representation and presentation of data to facilitate understanding.” [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Defining Visualization

“The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.” [Card et al., 1999]

“The representation and presentation of data to facilitate understanding.” [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Defining Visualization

“The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.” [Card et al., 1999]

“The representation and presentation of data to facilitate understanding.” [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Example 1 – raw data

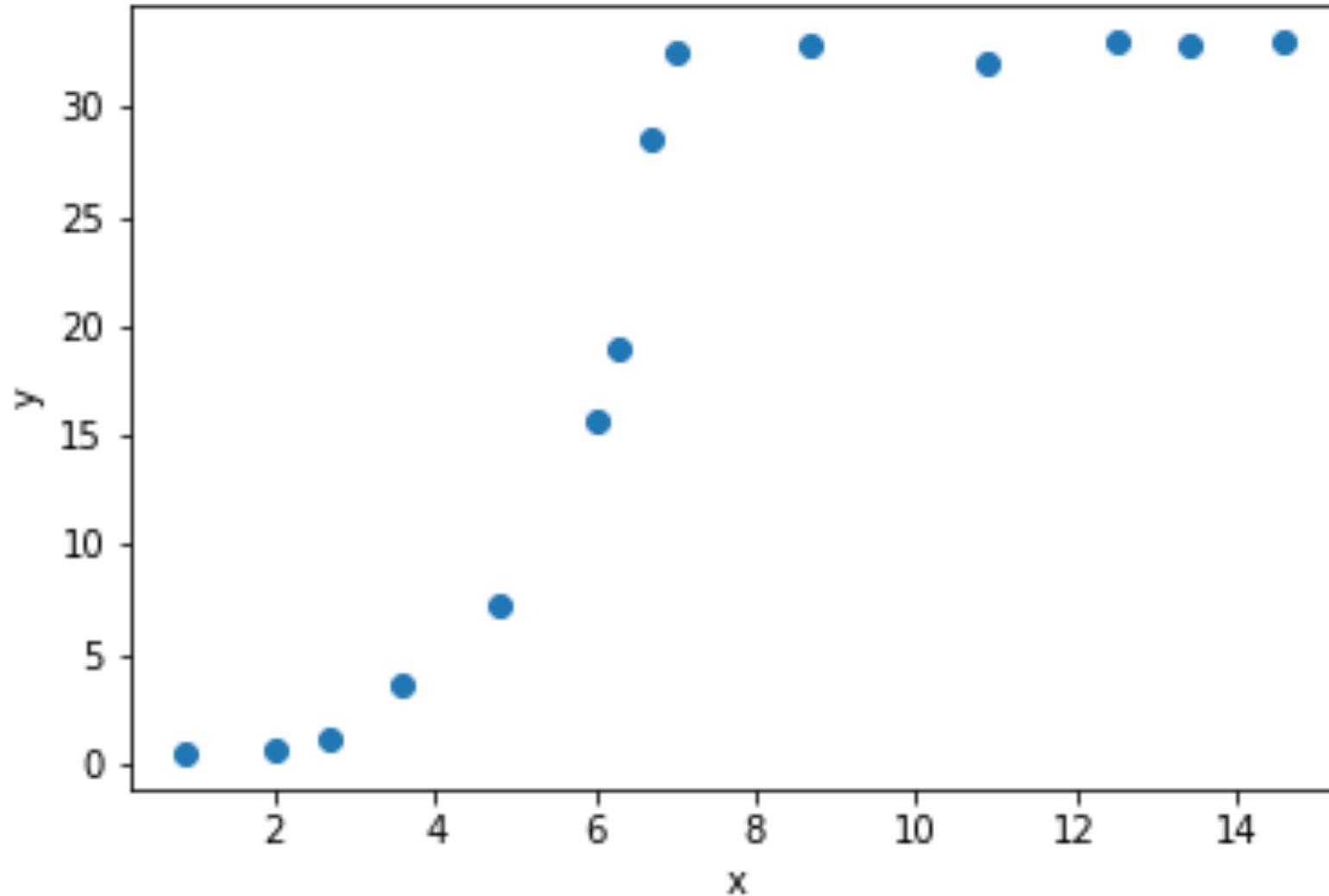
X	Y
0.9	0.5
2.7	1.1
6.7	28.6
10.9	32.8
6.0	15.7
6.3	19
7.0	32.6

X	Y
8.7	32.3
4.8	7.3
12.5	33.1
13.4	32.9
2.0	0.75
3.6	3.6
14.6	33

	X	Y
Median	6.5	23.8
Mean	7.2	19.5
STD.DEV	4.2	13.6

Correlation = 0.88

Example 1 – Visualized

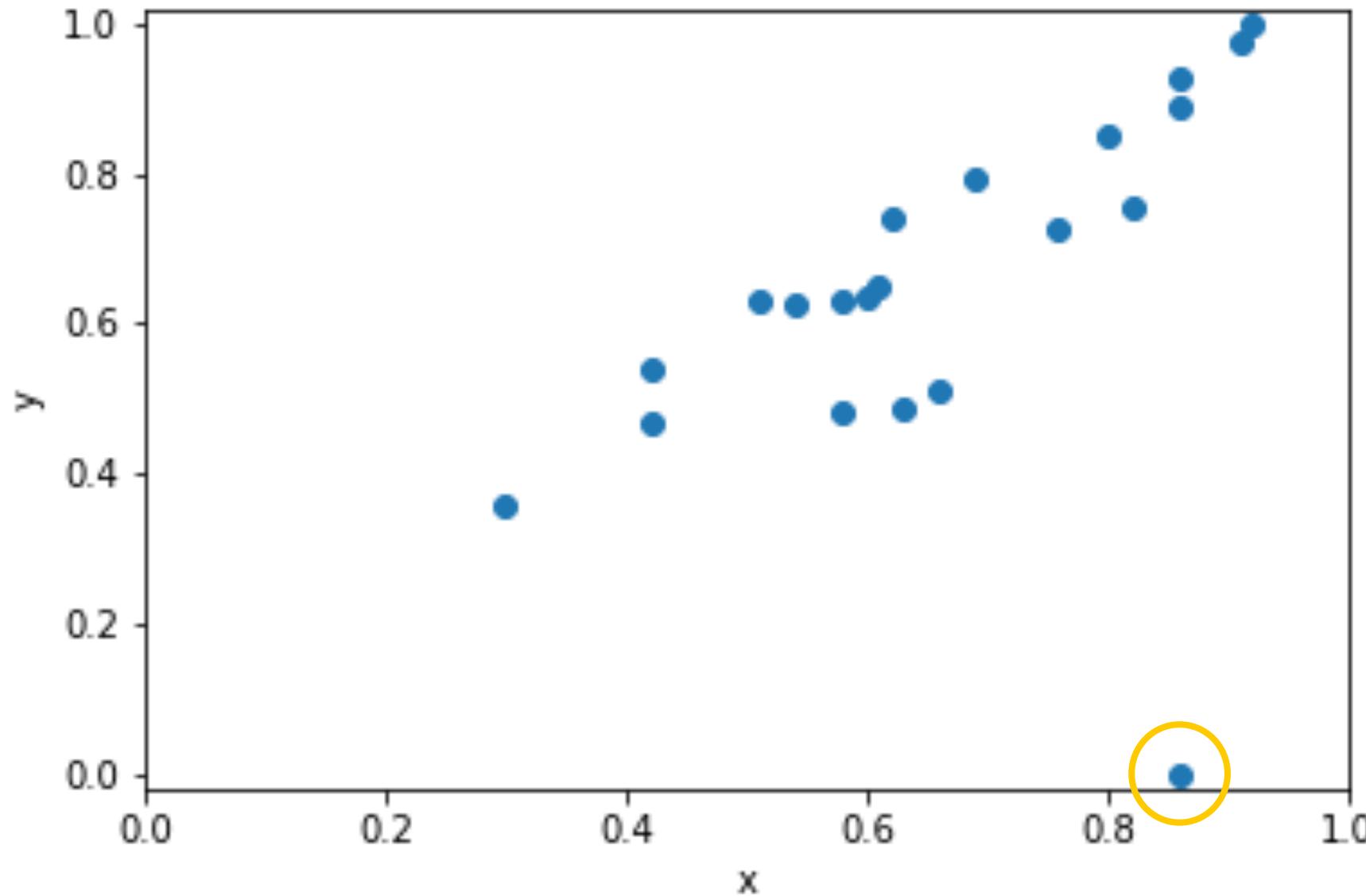


Example 2 – raw data

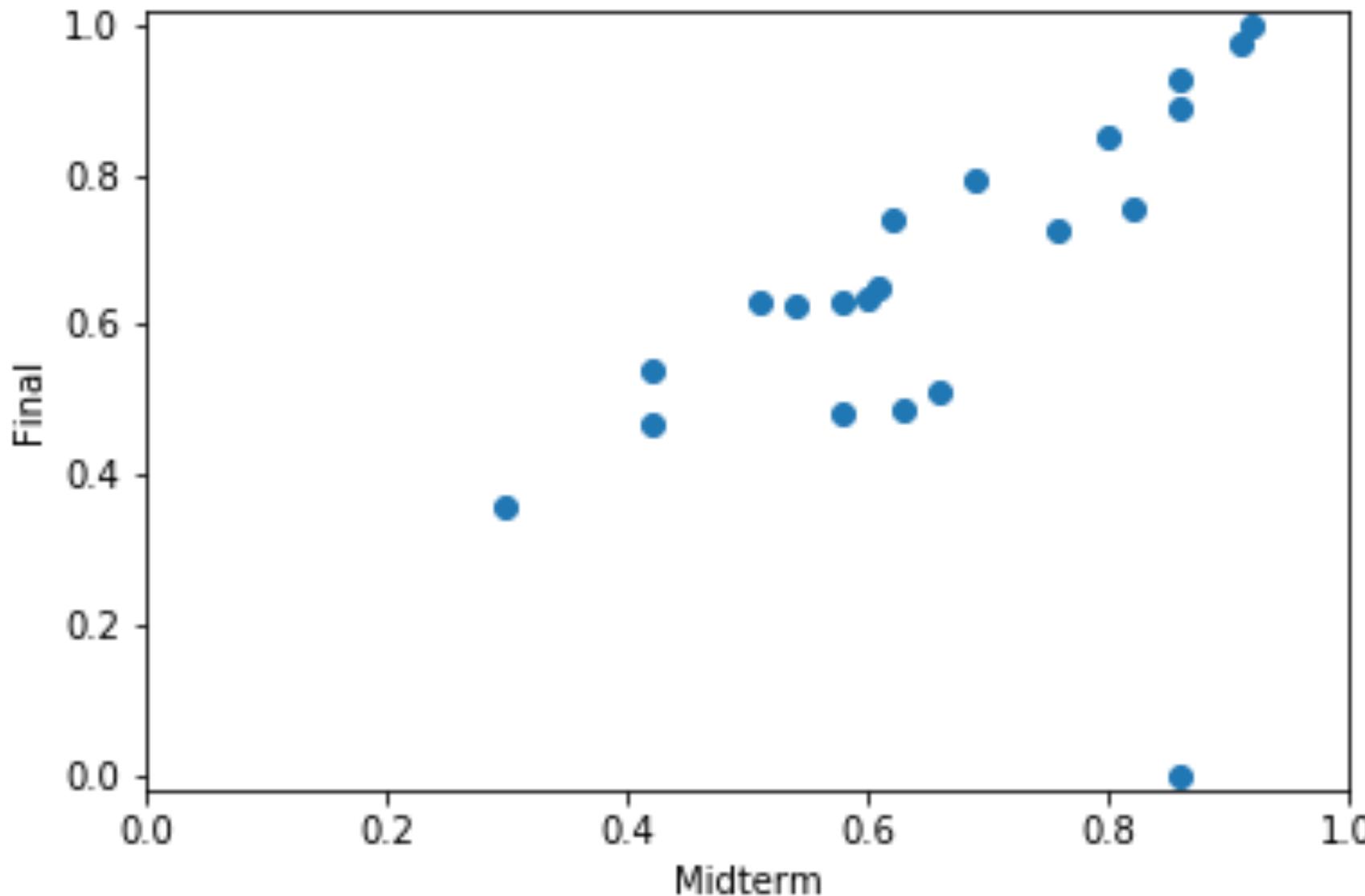
X	Y
0.42	0.46750
0.54	0.62500
0.42	0.53750
0.86	0.92750
0.60	0.63750
0.51	0.63125
0.30	0.35625
0.61	0.65000
0.58	0.63125
0.76	0.72500
0.58	0.48125

X	Y
0.66	0.51250
0.63	0.48750
0.92	1.00000
0.86	0.88750
0.91	0.97500
0.82	0.75625
0.86	0.00000
0.80	0.85000
0.69	0.79375
0.62	0.74000

Example 2 – Visualized



Example 2 – Visualized



Defining Visualization

“The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.” [Card et al., 1999]

“The representation and presentation of data to facilitate understanding.” [Kirk, 2016]

- Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

BACKUP

Big Data

“According to computer giant IBM,
2.5 exabytes - that's 2.5 billion
gigabytes (GB) - of data was
generated every day in 2012.”

- Laurie Miles*

*In the article by Matthew Wall. Big Data: Are you ready for blast-off. BBC News. March 2014.

Big Data

2.5 exabytes = 1000 petabytes

1 petabyte = 1000 terabytes

2.5 exabytes = 2.5 Million hard disks

Types of Visualizations

Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

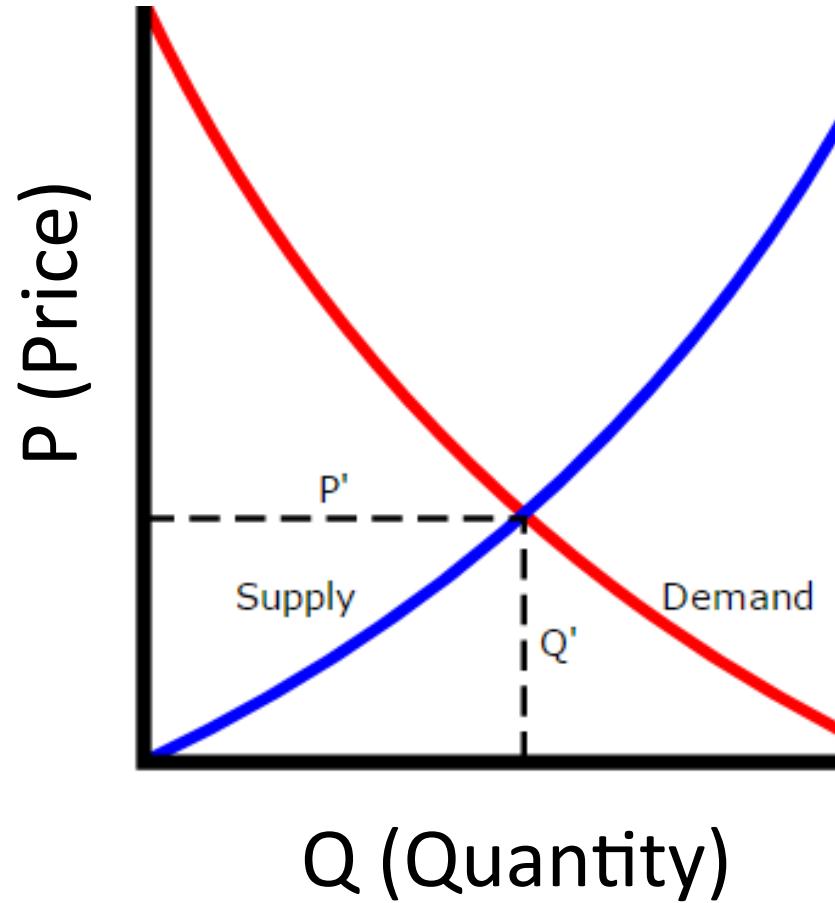
- Understand the different ways data visualization is used in data science

Two key categories*

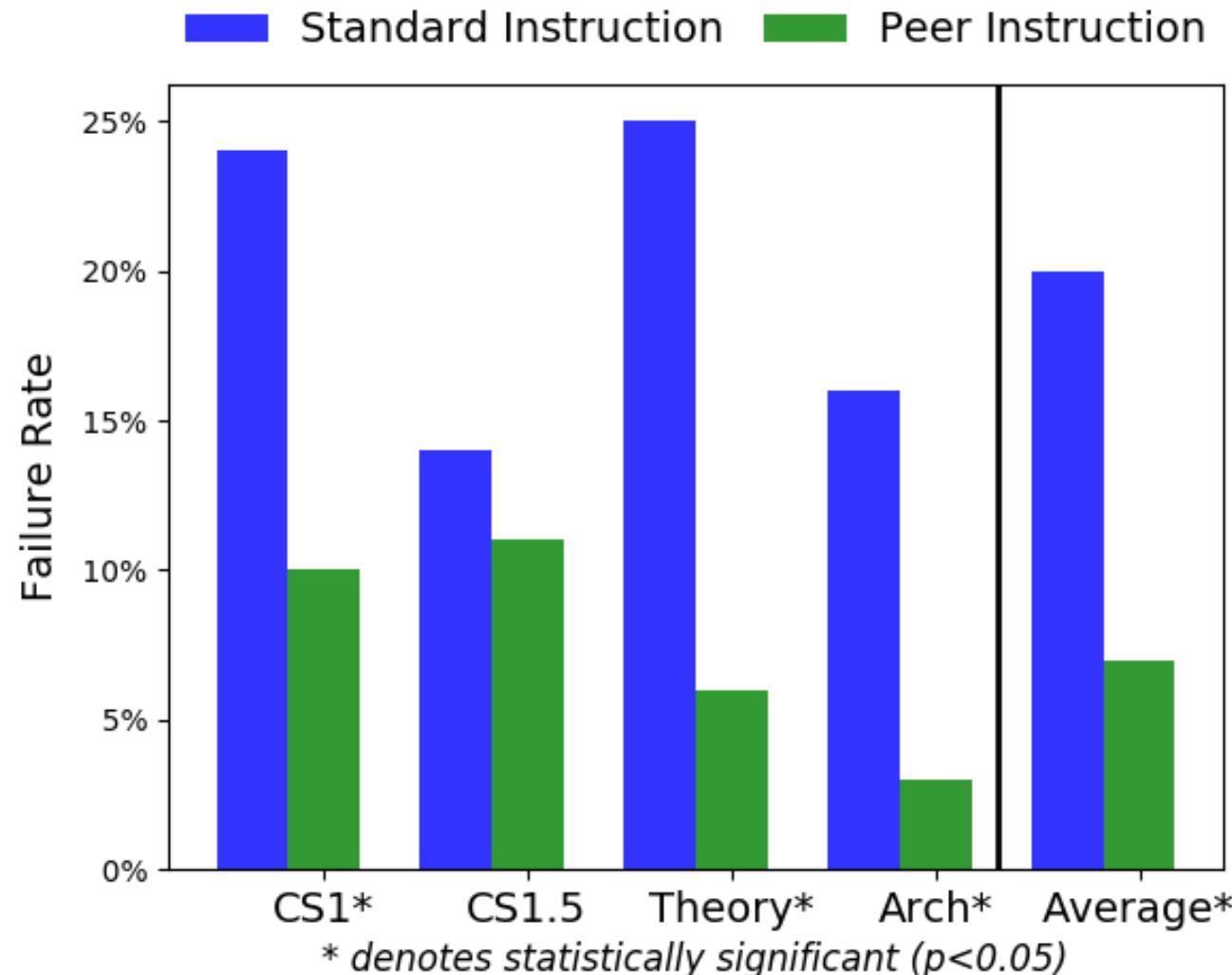
Conceptual or data-driven

Declarative or exploratory

Conceptual: Declarative

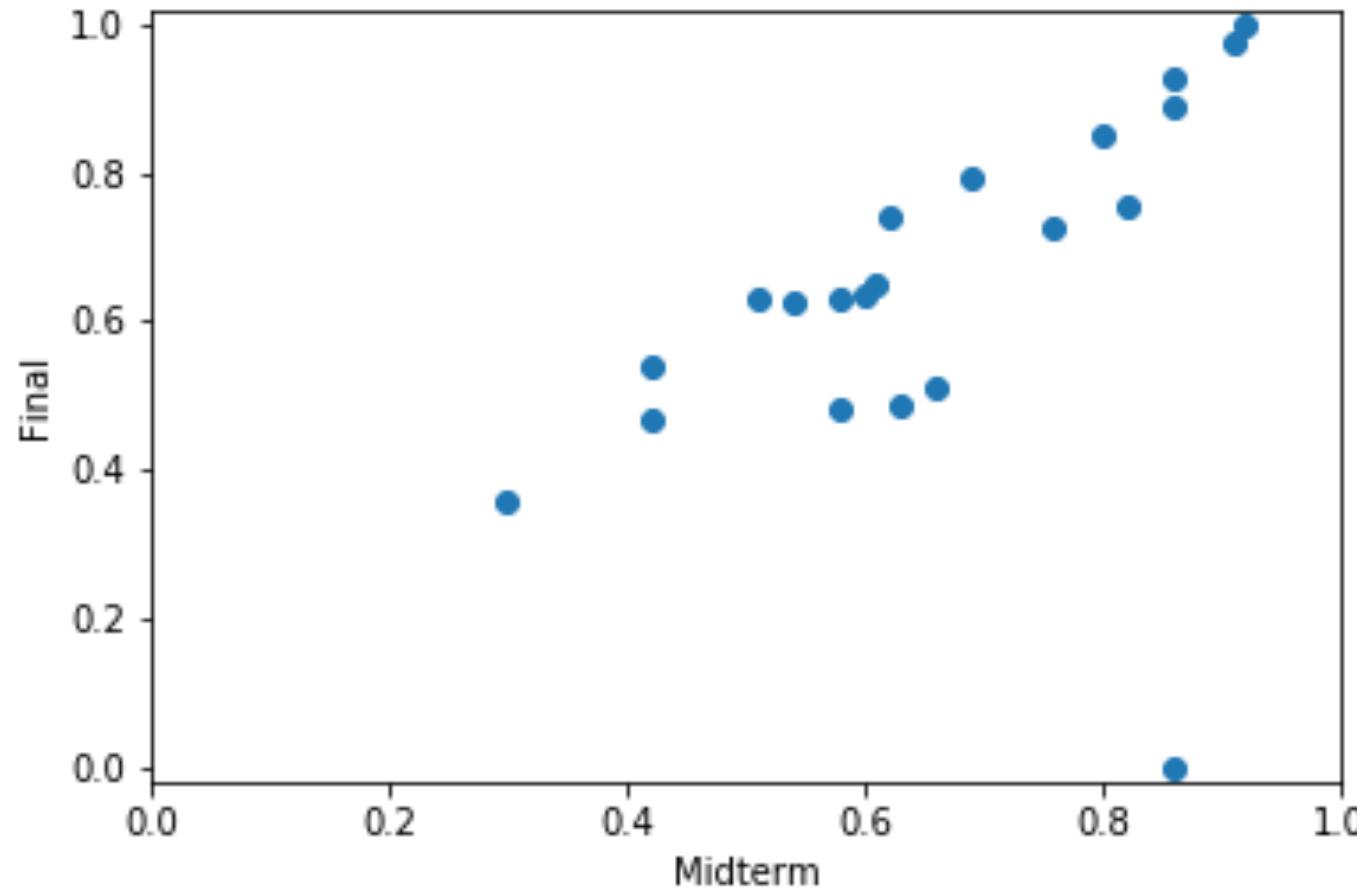


Data-Driven: Declarative



Porter, L., Bailey Lee, C. and Simon, B. Halving fail rates using peer instruction: a study of four computer science courses. In *Proceeding of the 44th ACM technical symposium on Computer science education*. March, 2013.

Data-Driven: Explorative



Visualization: Key design principles

Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

- Recognize qualities of good data visualizations

Principles of Good Design

"Good data visualization is:

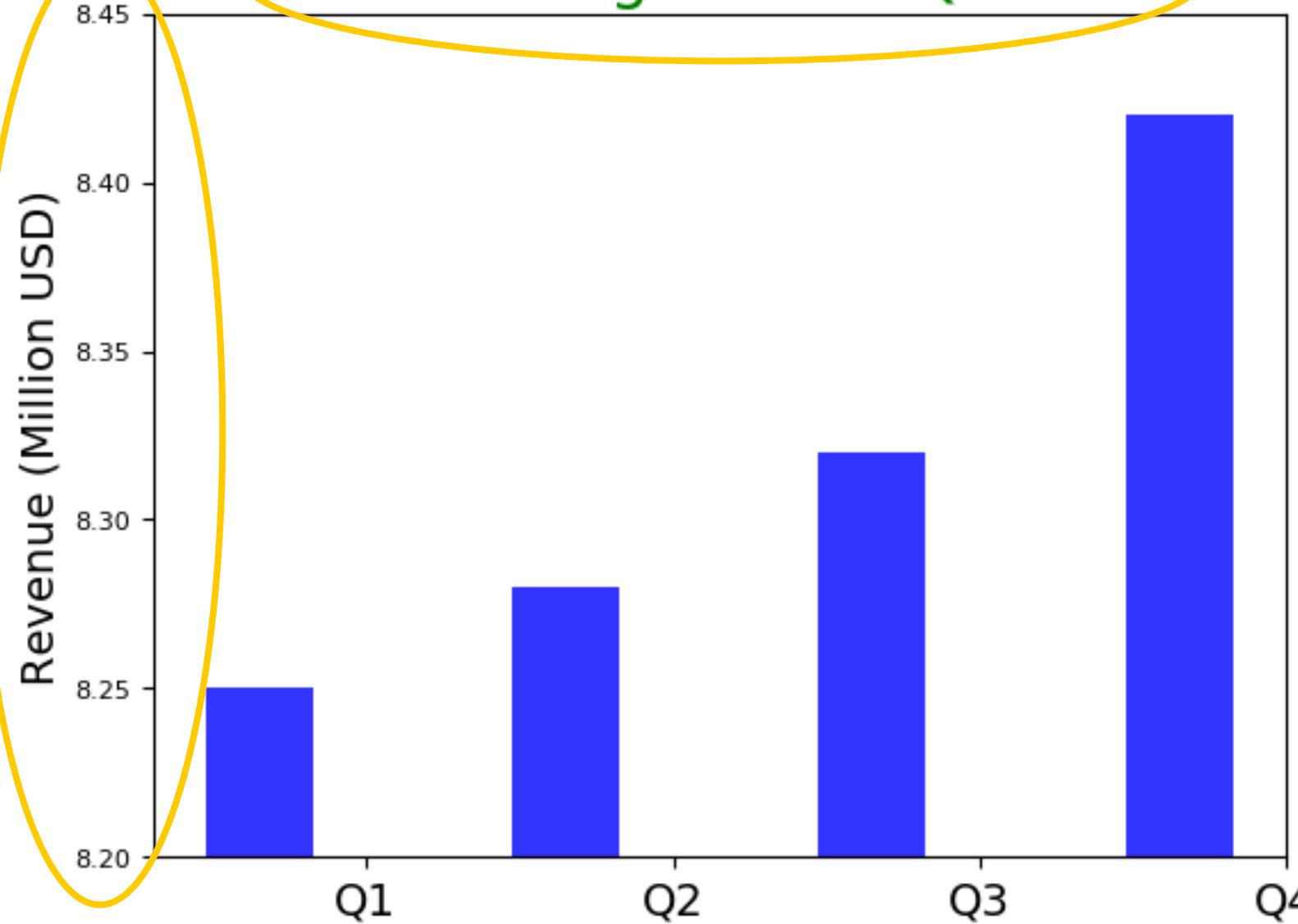
1. Trustworthy
 2. Accessible
 3. Elegant"
- Andy Kirk*

* Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Trustworthy

Trustworthy

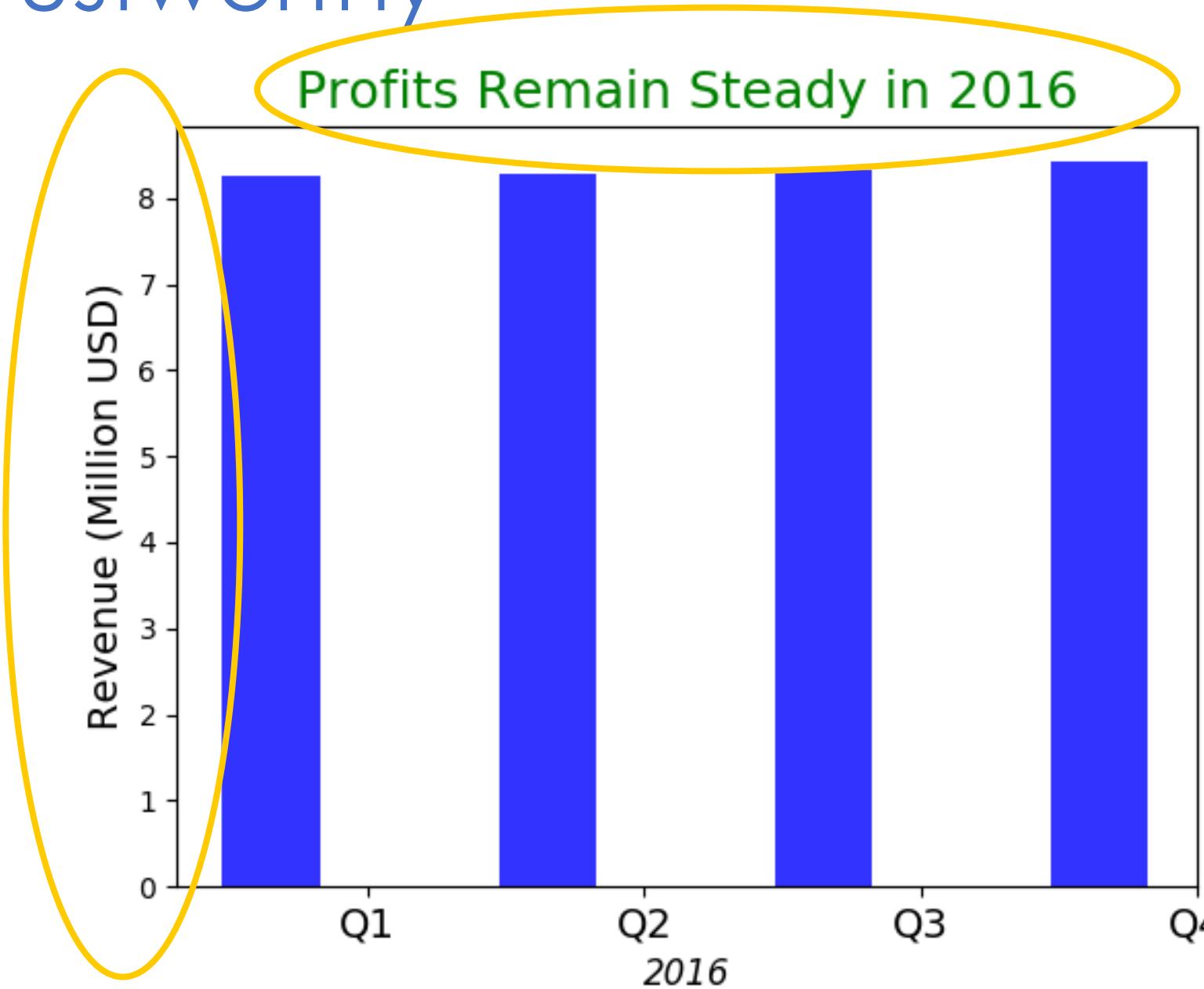
Python for Data Science



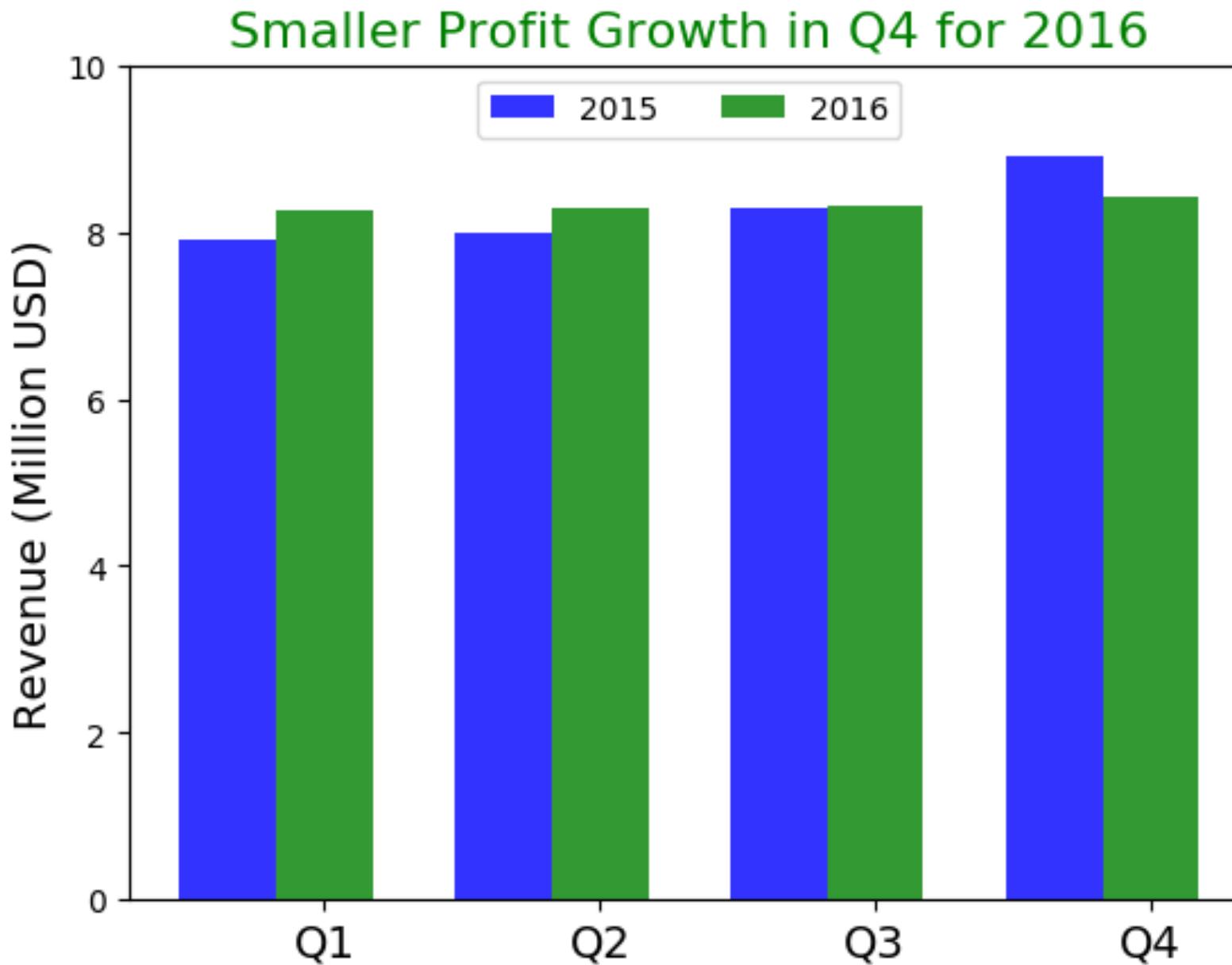
Profits Surge in 4th Quarter

Python for Data Science

Trustworthy



Trustworthy



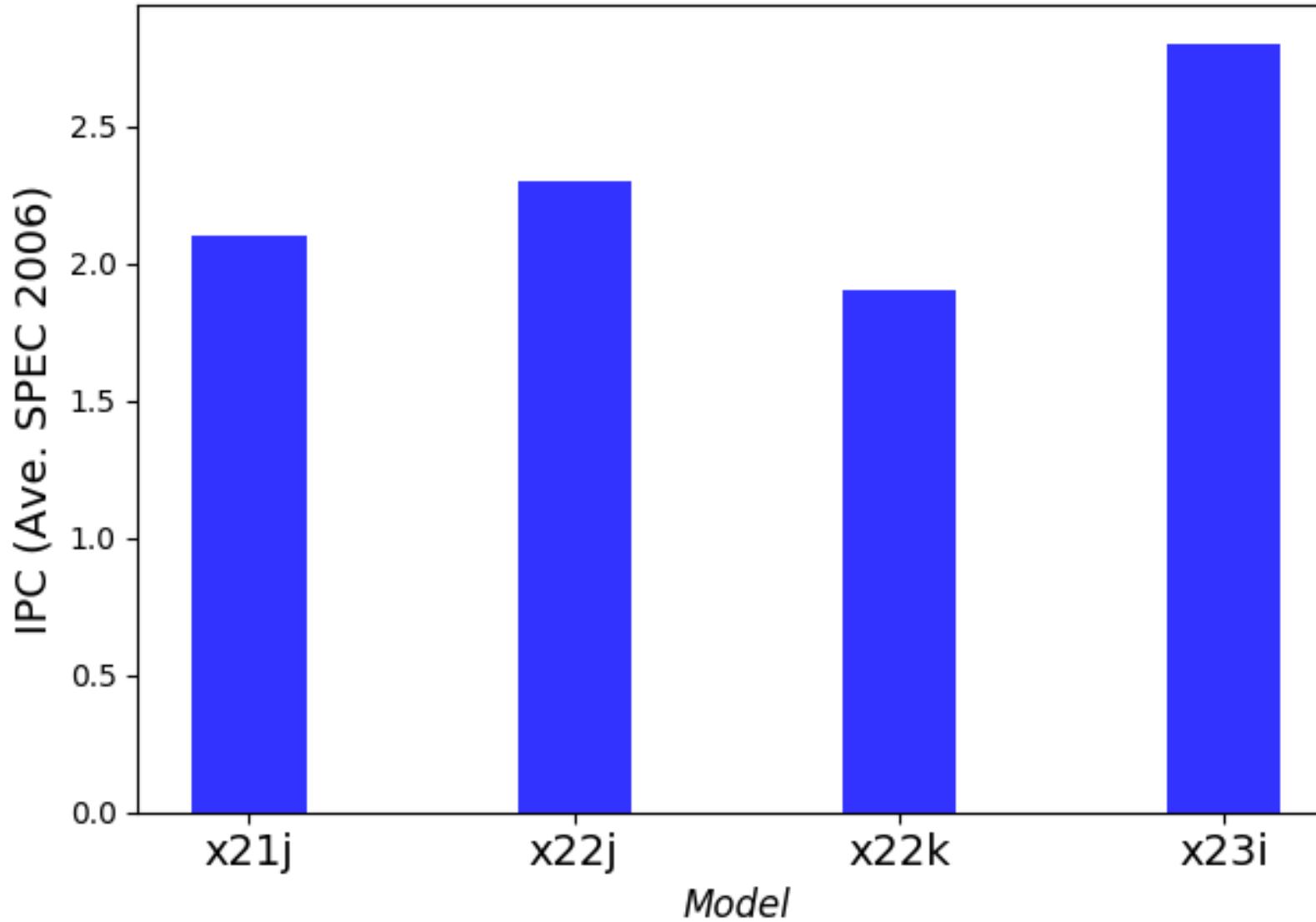
Trustworthy

- Trust is hard to earn, easy to lose.
- Honesty and integrity should be everywhere in the data science process

Accessible

Accessible

Performance by CPU Model

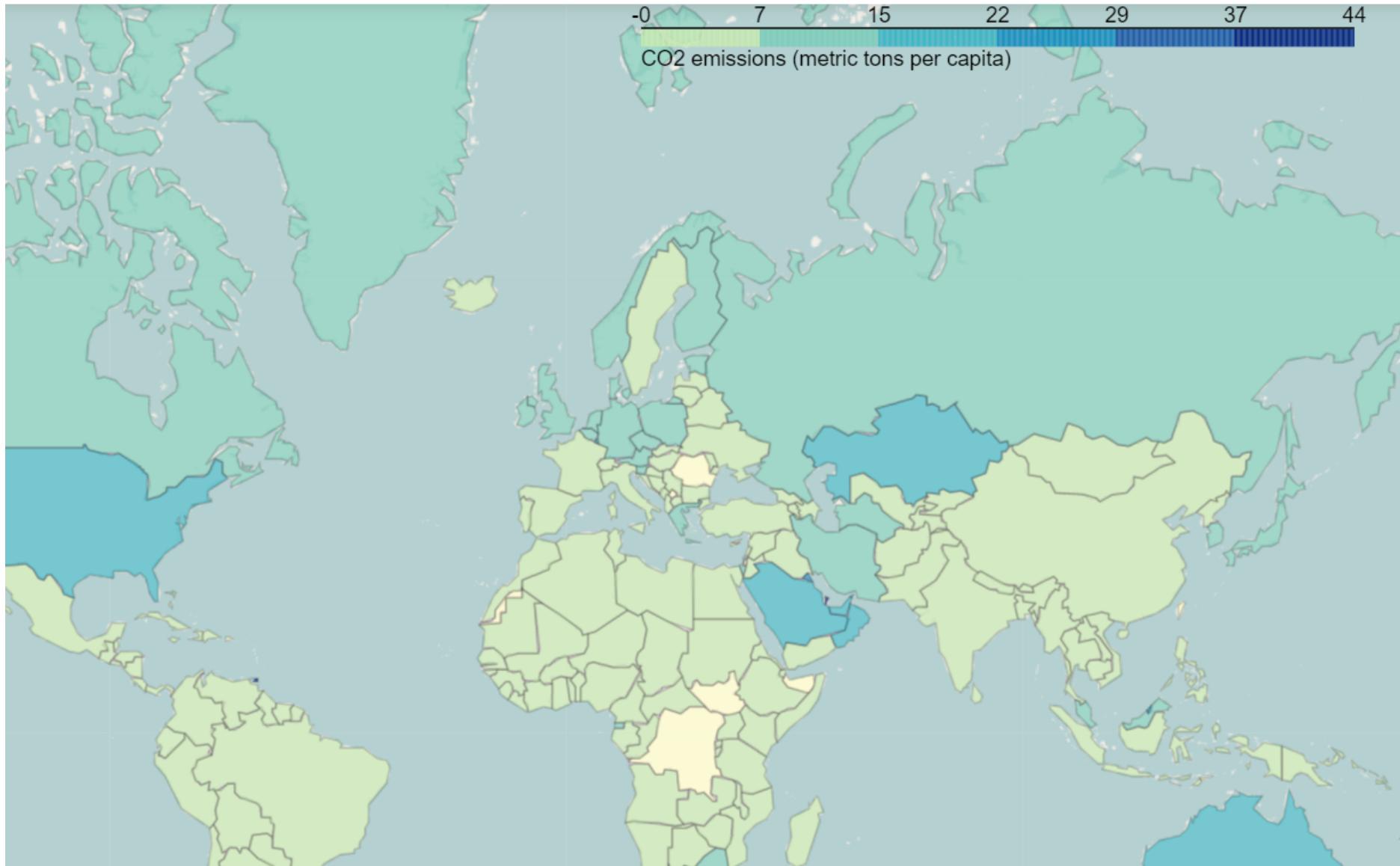


Accessible

- Know your audience
- Understand the purpose of the visualization

Elegant

Elegant



Elegant

- Focus on the relevant
- Be stylish if possible
- Think about decorations

Principles of Good Design

"Good data visualization is:

1. Trustworthy
 2. Accessible
 3. Elegant"
- Andy Kirk*

* Kirk, A. Data Visualisation: A handbook for Data Driven Design. SAGE publications, 2016.

Matplotlib

Dr. Ilkay Altintas and Dr. Leo Porter

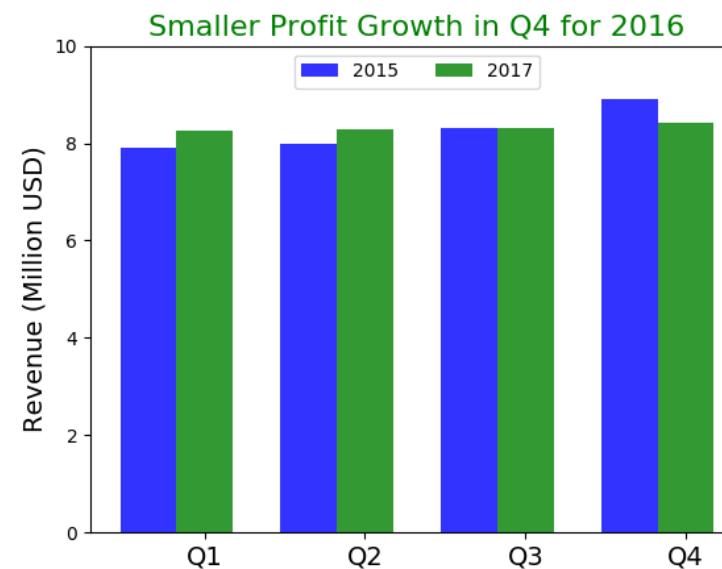
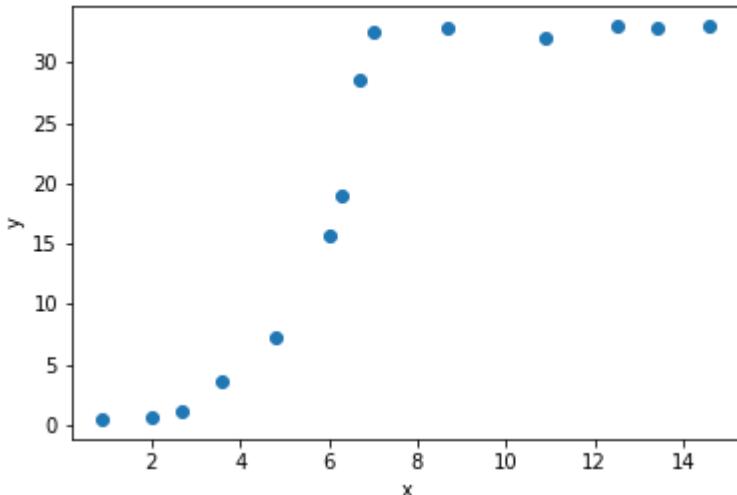
Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

- Explain the role of Matplotlib in python

What is matplotlib

- Plotting Library for Python



Why matplotlib

"Matplotlib tries to make easy things
easy and hard things possible."

<https://matplotlib.org/>

Why matplotlib

"Matplotlib tries to **make easy things easy** and hard things possible."

<https://matplotlib.org/>

Why matplotlib

"Matplotlib tries to make easy things
easy and **hard things possible.**"

<https://matplotlib.org/>

Other libraries?

- Seaborn
- ggplot
- Altair
- Bokeh
- Plotly
- Folium

Visualization: Basic Plotting in Matplotlib Part 1

Dr. Ilkay Altintas and Dr. Leo Porter

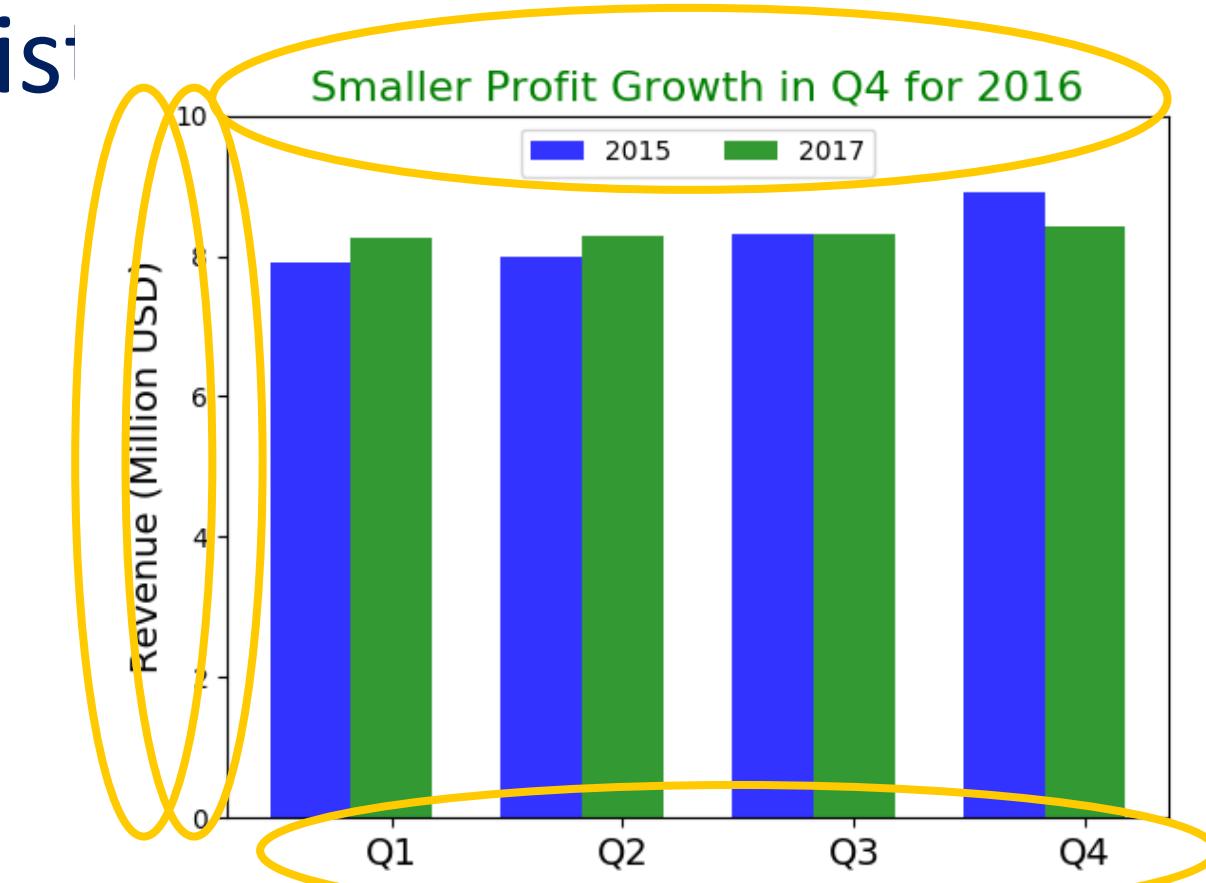
Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

- Create bar charts, line charts, and histograms using matplotlib
- Recognize the common components in a matplotlib figure

Common Components

- Chart type (bar, line, histogram)
- Axes data ranges
- Axes labels
- Figure labels
- Legend
- Aesthetics
- Annotations.



Visualization: Matplotlib Additional Examples

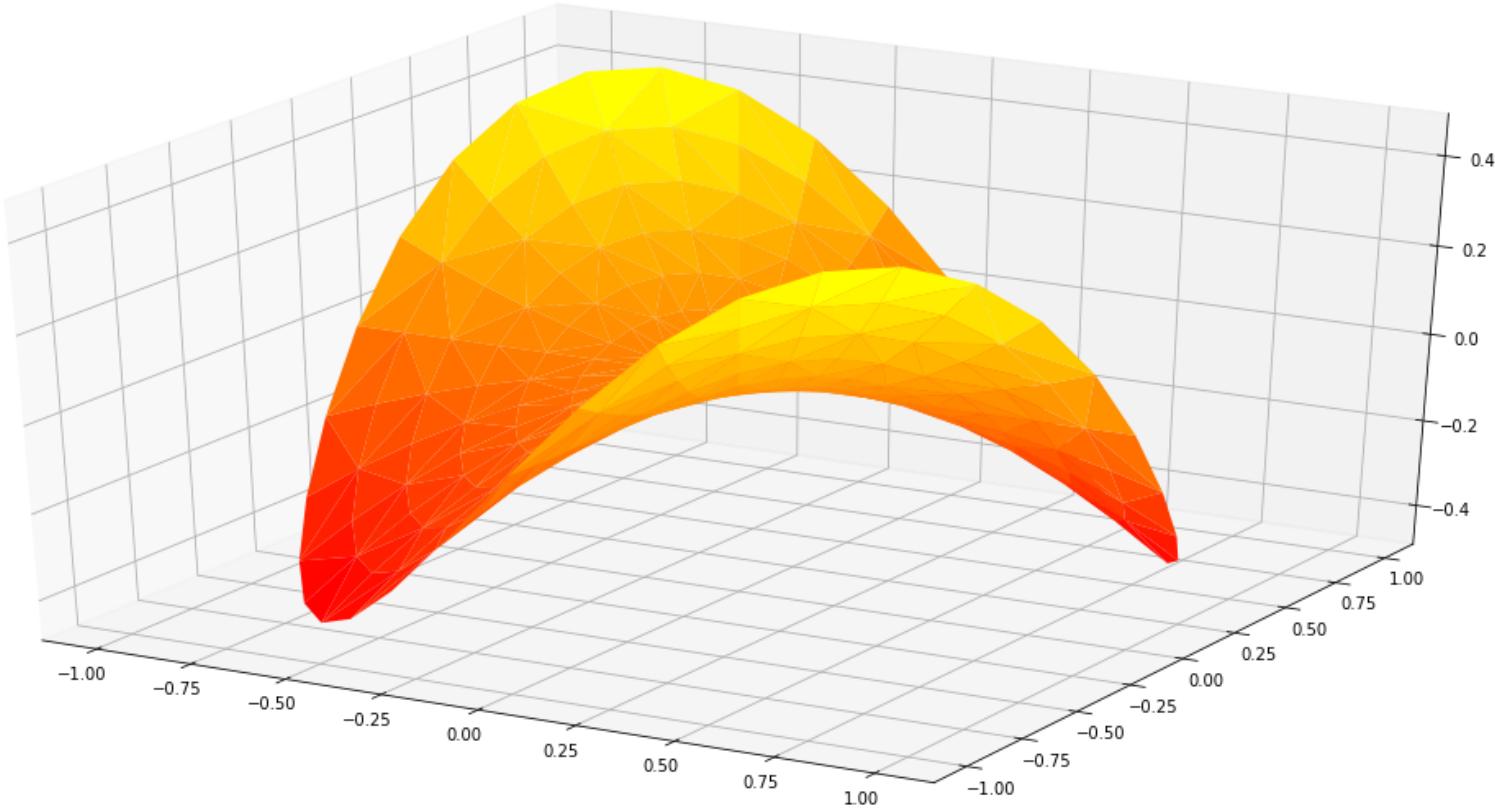
Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

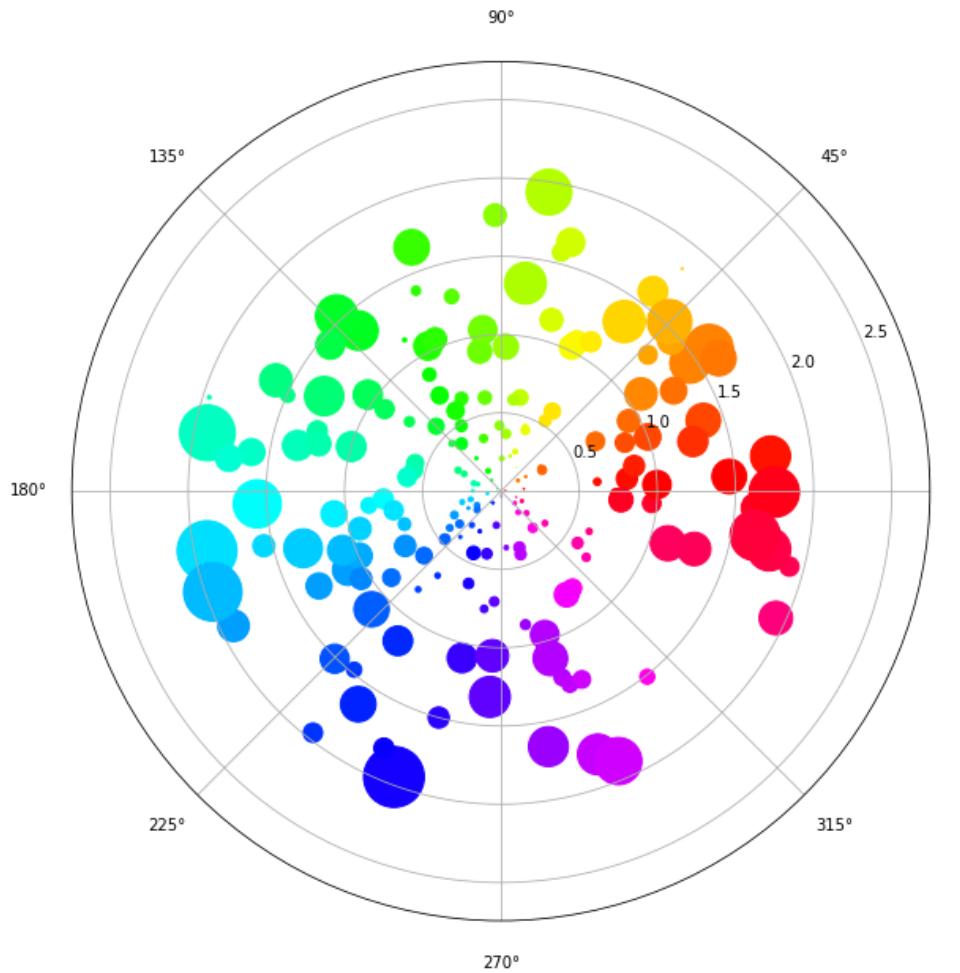
By the end of this video, you should be able to:

- Use the additional Jupyter notebook for examples

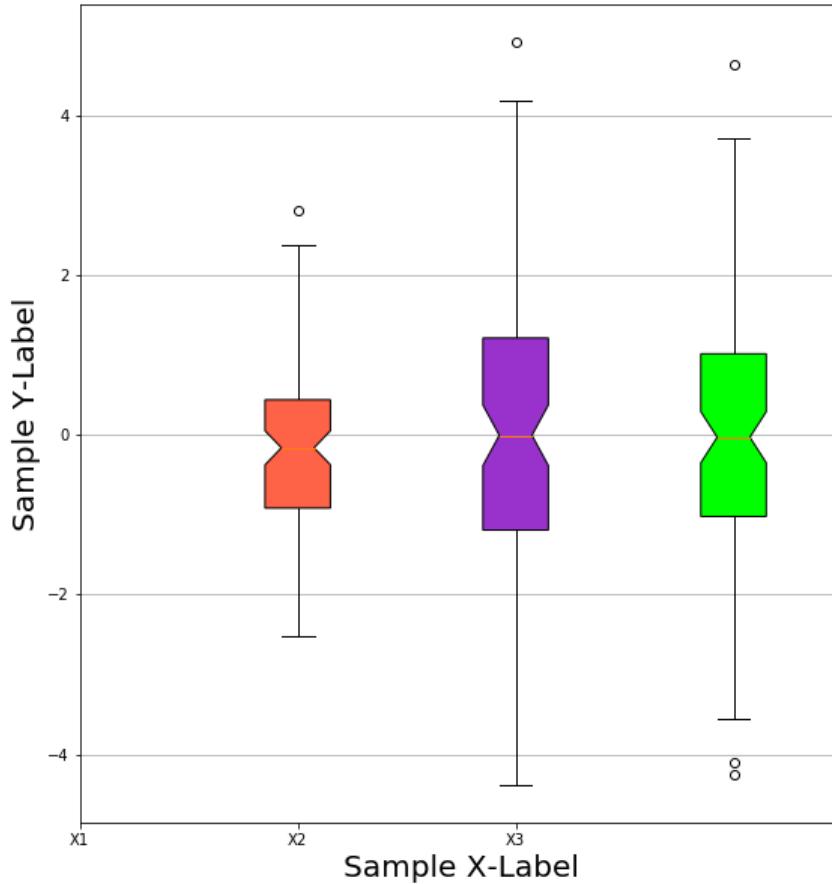
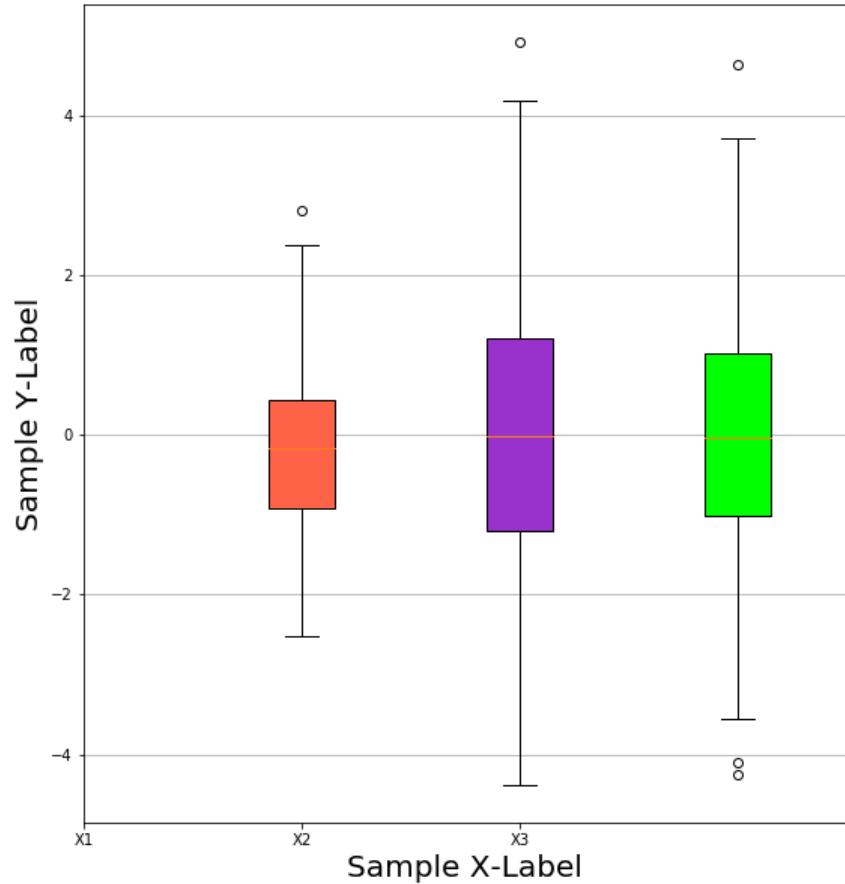
Additional Examples:



Additional Examples:



Additional Examples:



Visualization: Cholera Case Study

Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

By the end of this video, you should be able to:

- Describe the role played by data visualization in understanding how cholera spreads

Cholera

- Bacterial infection
- Causes severe diarrhea, possibly leading to death by dehydration
- Remains public health threat
 - 1.3-4.0 million cases worldwide
 - 21,000-143,000 deaths worldwide

How Cholera Spreads



Photo by Ashley Wheaton, April 2009.

How they thought Cholera spreads



"Doktor Schnabel von Rom"
Artwork by Paulus Fürst, 1656

The outbreak



The Cholera Map



Python for Data Science





Visualization: The Russian Campaign of 1812 Case study

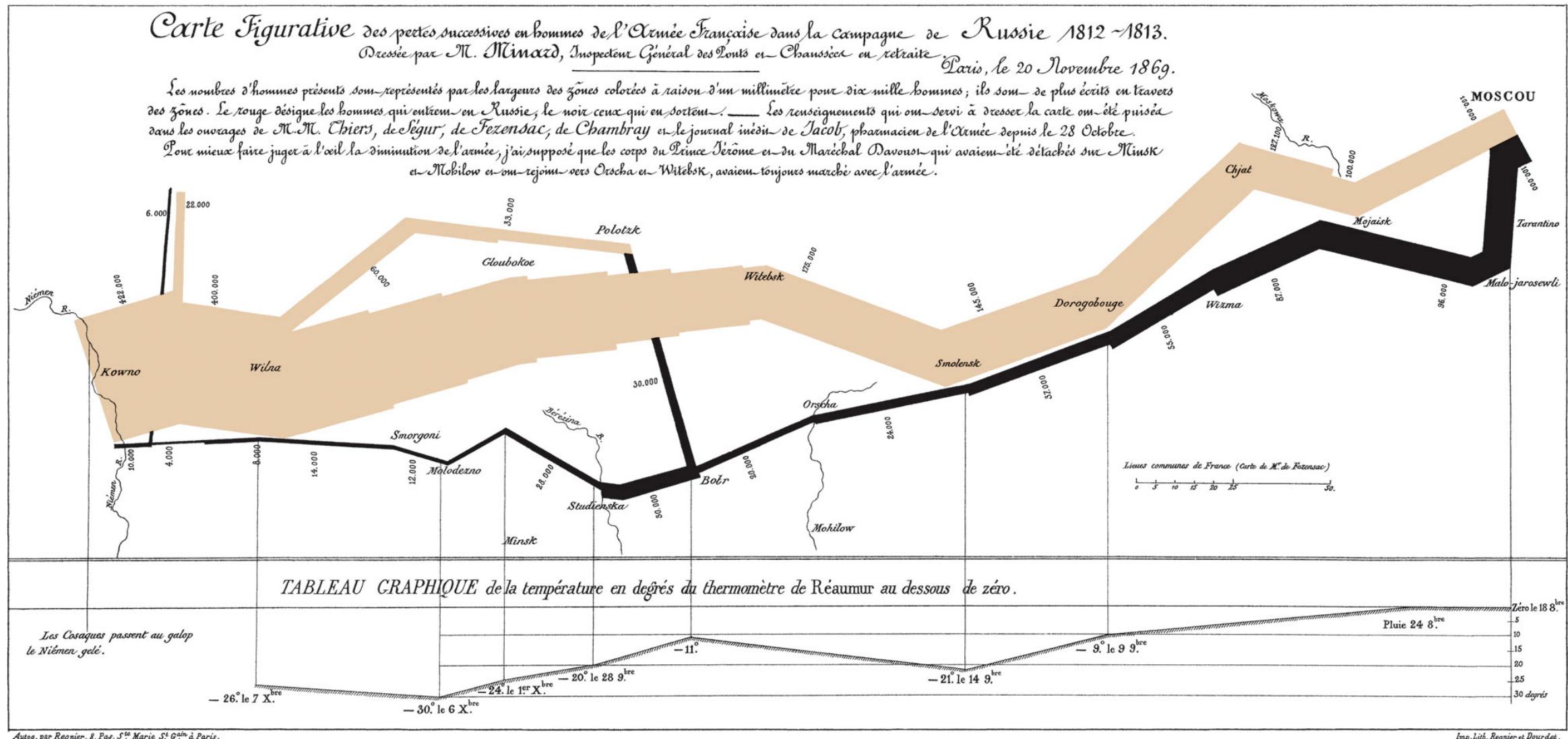
Dr. Ilkay Altintas and Dr. Leo Porter

Twitter: #UCSDpython4DS

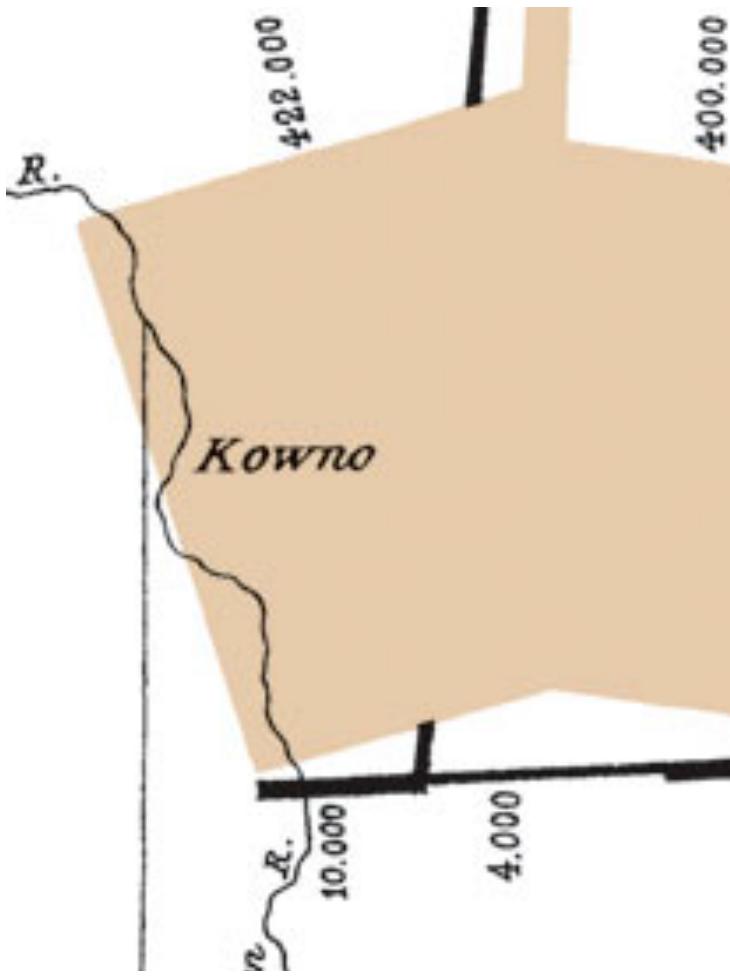
By the end of this video, you should be able to:

- Explain why the data visualization of Napoleon's Russian campaign in 1812 is particularly effective

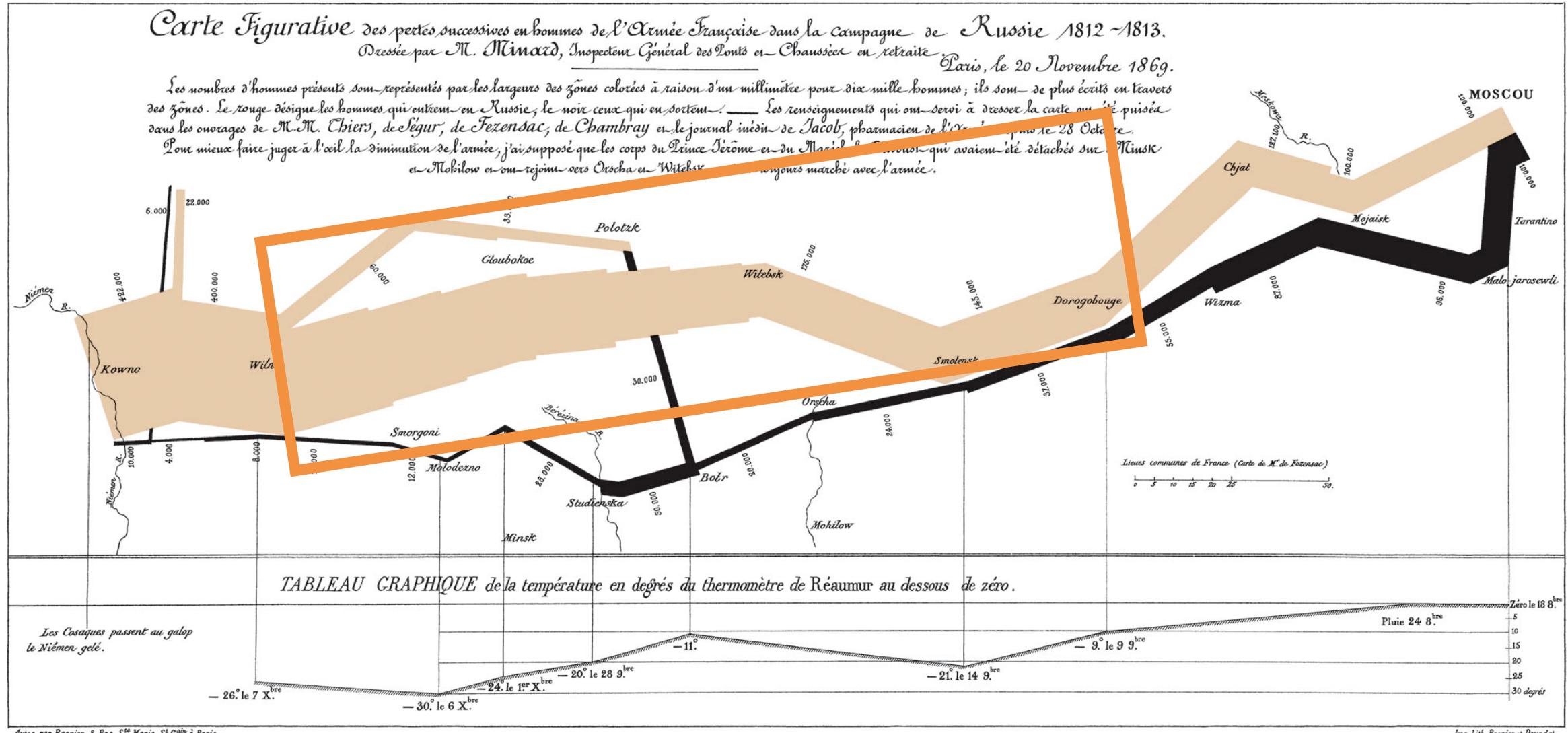
The Russian Campaign of 1812



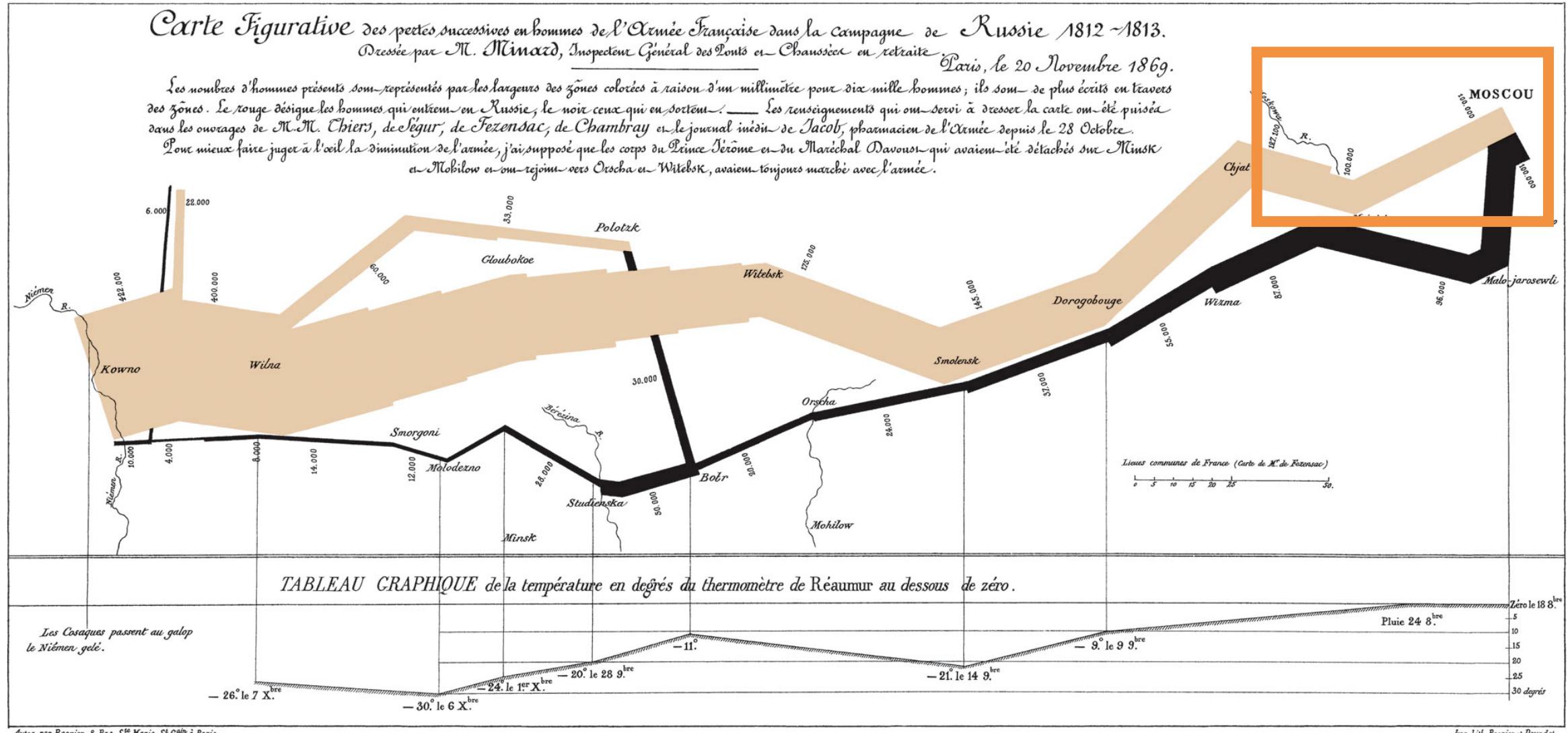
The Russian Campaign of 1812



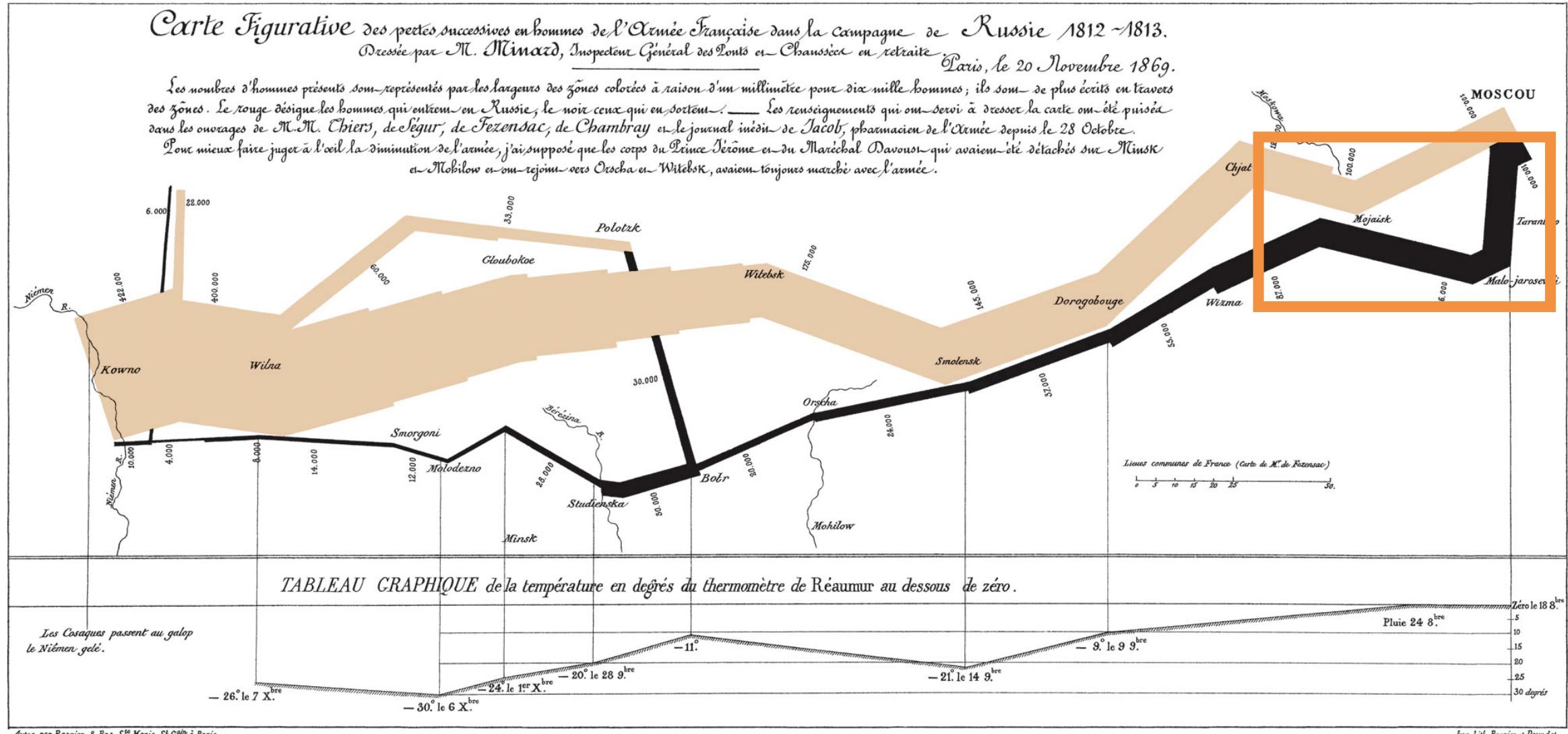
The Russian Campaign of 1812



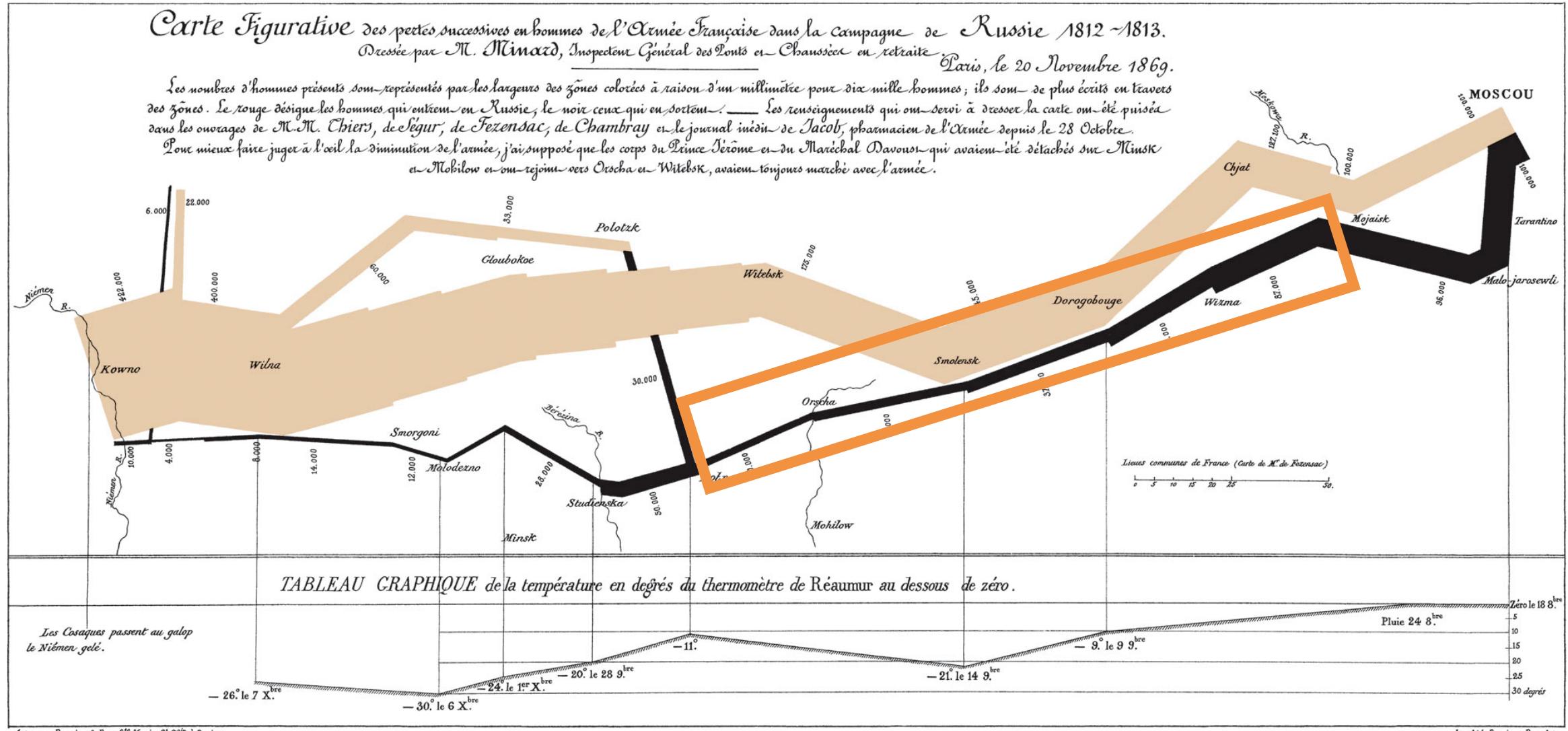
The Russian Campaign of 1812



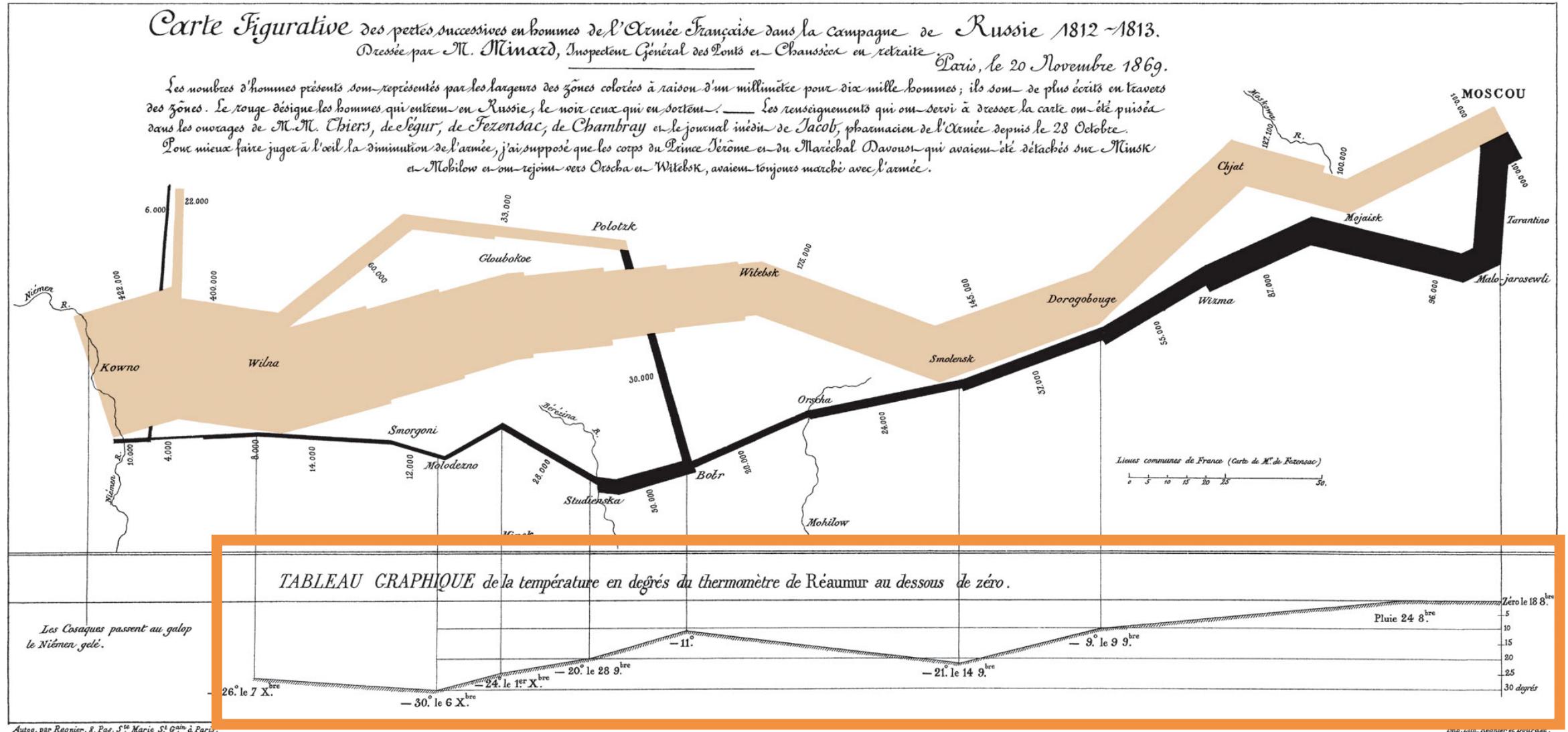
The Russian Campaign of 1812



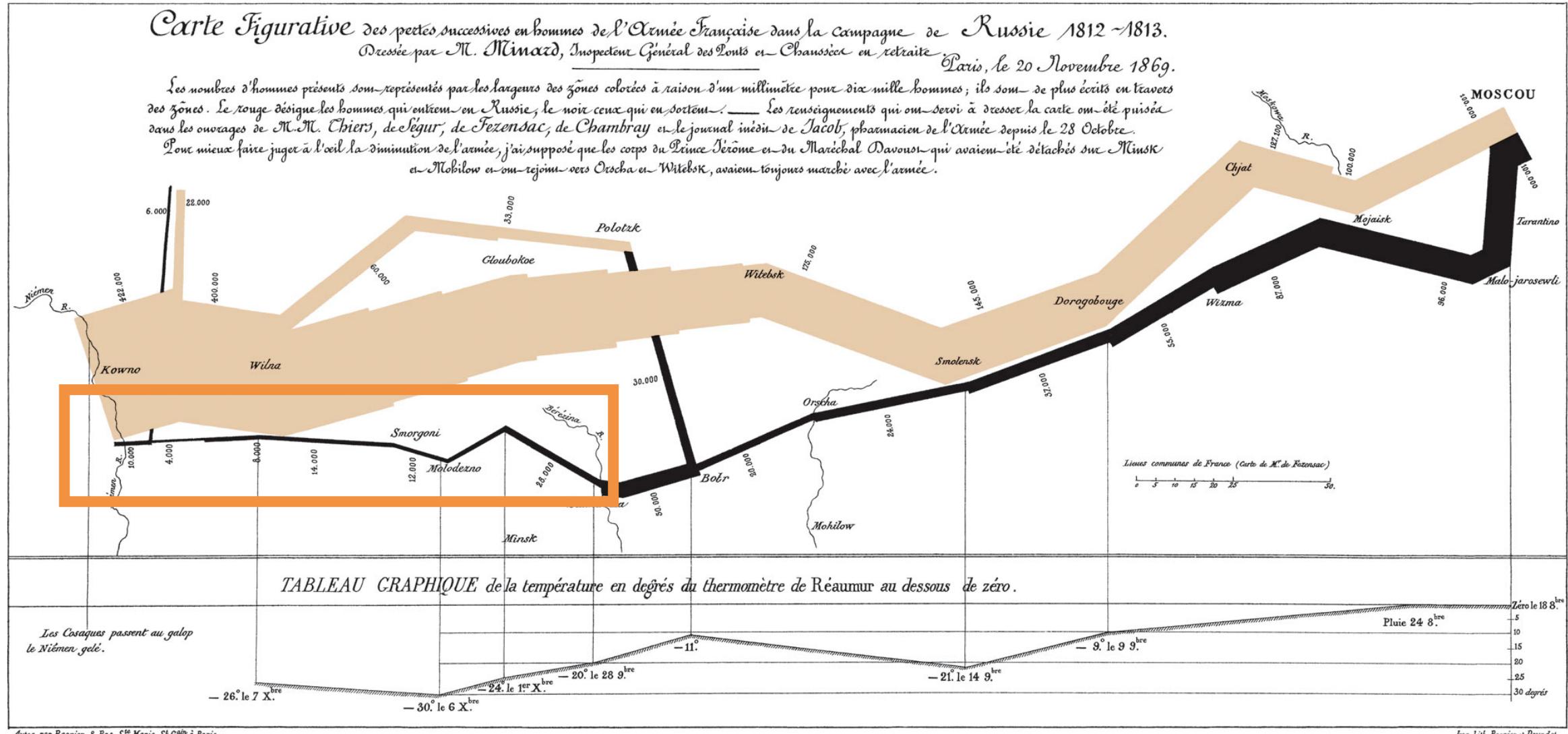
The Russian Campaign of 1812



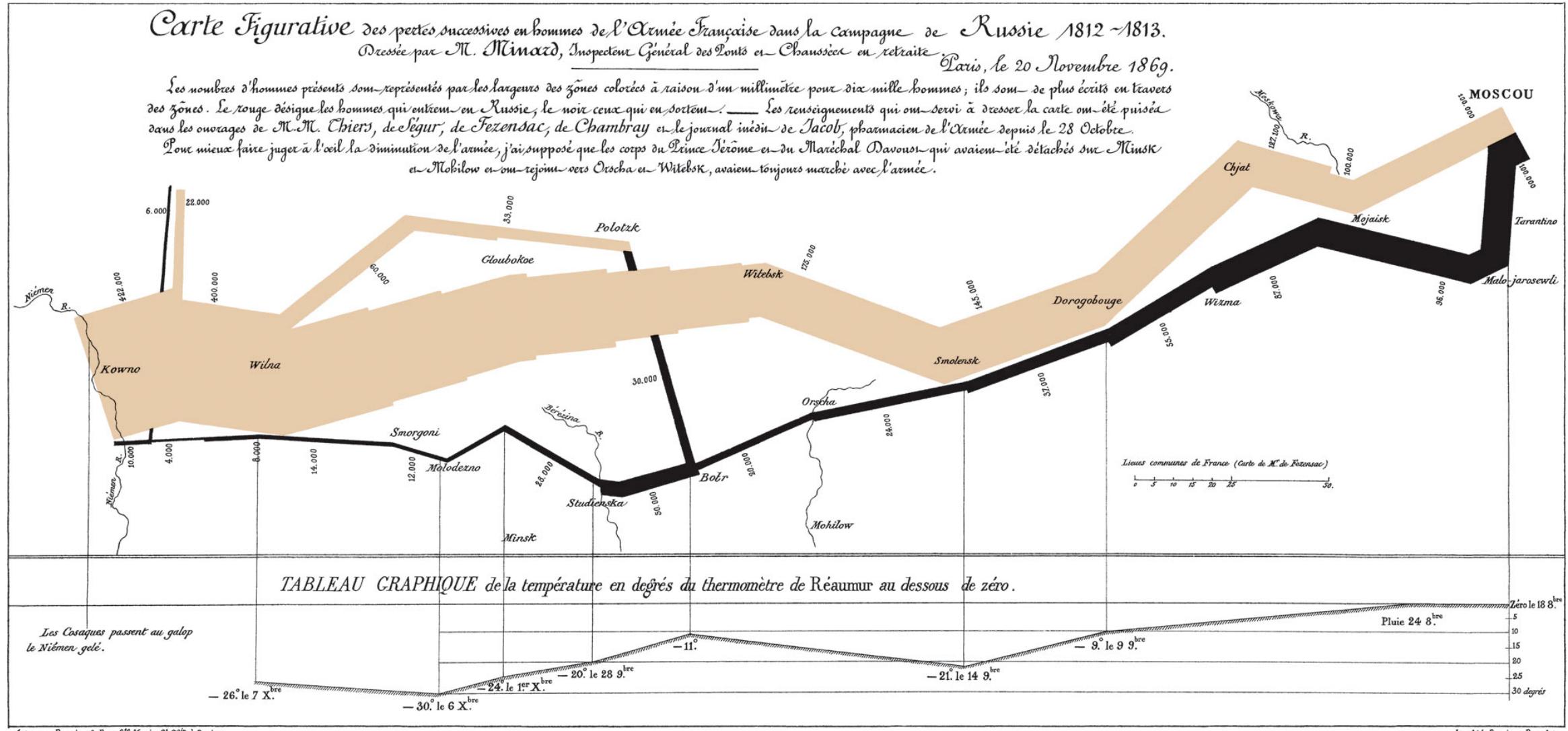
The Russian Campaign of 1812



The Russian Campaign of 1812



The Russian Campaign of 1812



Sources

- Jesse Greenspan. Napoleon's Disastrous Invasion of Russia. History.com, June 2012.

<http://www.history.com/news/napoleons-disastrous-invasion-of-russia-200-years-ago>

- G.D. Sankey or Harness. The Economist. June 2011.

<http://www.economist.com/blogs/dailychart/2011/07/data-visualisation>

- The Russian Campaign, 1812. PBS.org.

http://www.pbs.org/empires/napoleon/n_war/campaign/page_12.html