

Data Visualization - Tutorial

Stefan Halfpap

www.bifold.berlin

https://github.com/klauck/data_visualization_tutorial

About me





PostDoc in the Database Systems and Information Management (DIMA) group

- Distributed stream processing
- Data management systems

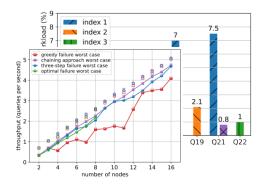


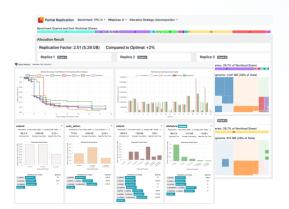


Plattner Institut PhD student in the Enterprise Platform and Integration Concepts (EPIC) group

- Physical database optimization
- Columnar main-memory database systems









Outline



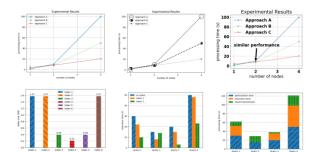
- 1. Motivation
- 2. Types and Tools
- 3. Hands on Examples
- 4. Best Practices
- 5. Summary and Discussion

Outline



- 1. Motivation
- 2. Types and Tools
- 3. Hands on Examples
- 4. Best Practices
- 5. Summary and Discussion

Matplotlib for static graphs

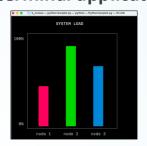


(Intro) Chart.js for interactive web applications

What you will learn

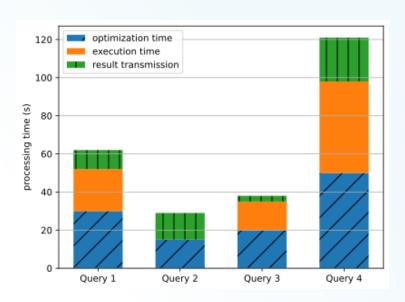


(Intro) Python curses for terminal applications



Motivation





Motivation



• Humans can more easily perceive visual differences than sequences of text or numbers

Data visualizations

Exploring, understanding, and communicating data

Essential research skill

- Early prototypes -> systematic experiments -> final publications
- Papers/theses, presentations, or applications/demonstrations



What to use?

It depends!



What to use?

It depends on..

the purpose, data, and own experience.



Factors that influence suitable data visualizations and tools

Purpose

- Self usage vs. communicate to others vs. products
- Papers/theses, presentations, or applications/demonstrations
- Customization, automation, reproducibility
- Static vs. dynamic

Data

- Size, dimensions, and kind (2D vs. 3D; metric, categorical, maps, ...)
- Your own (programming) experience and available software
 - Beginner
 - Intermediate
 - Advanced



Factors that influence data visualization types and tools

- Purpose
 - Self usage vs. communicate to others vs. products
 - Papers/theses, presentations, or applications/demonstrations
 - Customization, automation, reproducibility
 - Static vs. dynamic
- Data
 - Size, dimensions, and kind (2D vs. 3D; metric, categorical, maps, ...)
- Your own (programming) experience and available software
 - Beginner (OpenOffice Calc or MS Excel, Google Sheets, web applications)
 - Intermediate (Tableau Software or other BI tools)
 - Advanced (visualization libraries, e.g., Matplotlib, chart.js)



Suitable visualization libraries depend on the programming language and your specific needs

Hands on



https://github.com/klauck/data_visualization_tutorial



Matplotlib for static graphs

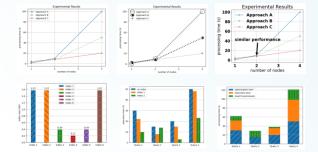
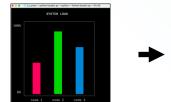


Chart.js for interactive web applications





Python curses for terminal applications





Best Practices



- Use what fits best to you(r needs)
- Organize the data to visualize (structure, version, queryable)
- Configurable, automated, and reproducible visualizations
- Visualization styles
 - Keep it simple
 - Keep it clear/self-contained
 - Use titles, labels (e.g. axes with units), annotations, legends
 - Avoid distortion and misrepresentation (do not mislead the audience)
 - Use appropriate sizes, markers, colors, patterns
 - Use consistent styles, colors, and names
 - Chart type, data selection, scaling (linear vs. logarithmic)

See also: Friends Don't Let Friends Make Bad Graphs https://github.com/cxli233/FriendsDontLetFriends

Data Visualization – Tutorial Summary and Discussion



Suitable Types and Tools

depend on purpose, data, and own experience

Best Practices

- Organize the data to visualize
- Configurable, automated, and reproducible
- Appropriate visualization styles

Hands on Examples



https://github.com/klauck/data_visualization_tutorial

Matplotlib for static graphs

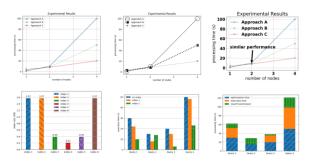


Chart.js for interactive web applications



Python curses for terminal applications

