Eötvös Loránd University (ELTE)
Faculty of Informatics (IK)
Pázmány Péter sétány 1/c
1117 Budapest, Hungary

VISUALIZATION SOLUTIONS

Open-source Technologies for Real-Time Data Analytics

Imre Lendák, PhD, Associate Professor

2019
Budapest, Hungary

Introduction

- Visualization types
  - Scientific viz
  - Information viz
  - Visual analytics
- Graph types
  - Line, bar, stacked bar, pie
  - Choropleth, scatter, heat
- Interactive visualization
- Viz tools: Tableau, Plotly, Datawrapper, Kibana, etc.

Data Visualization Process

Goals and data

- What is the goal of the visualization?
- What data do you have available?
- What level of detail does it go down to?
- How can you use other data to supplement your data?

Audience

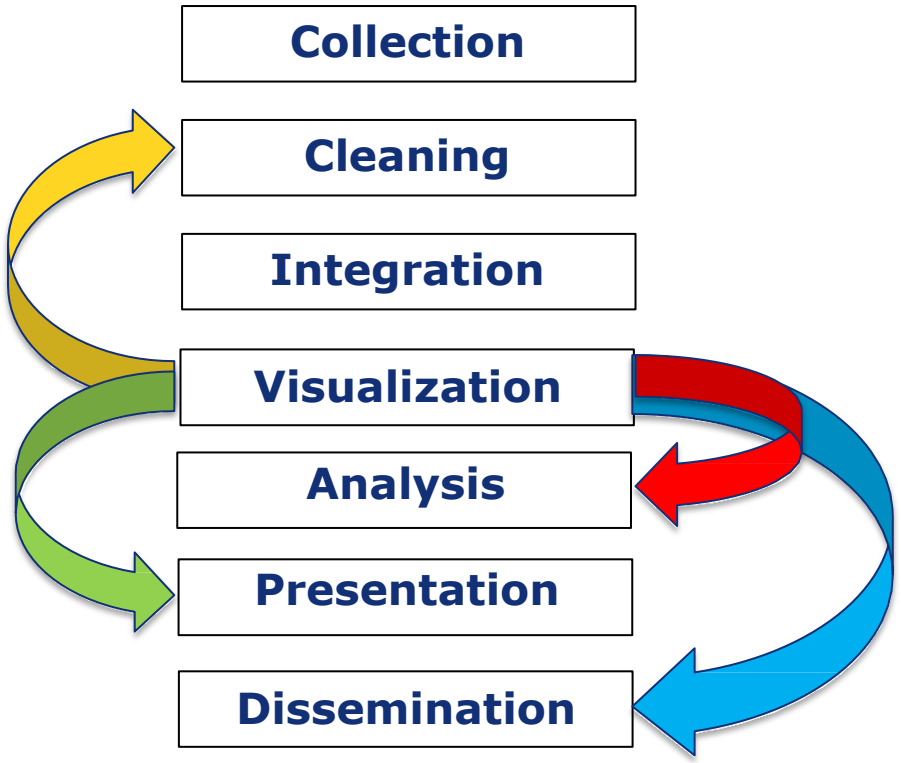
- How detailed do they want to see the data?
- Do they have a technical background?
- How will the visualization(s) be viewed? (desktop, mobile, print)

Data visualization goals (selection!)

- Reporting automation
- Executive reporting and presentation
- Customer reporting
- Self-service BI and visual data analysis
- Visual status monitoring
- Geo data visualization
- Visual data preparation
- Data journalism
- Math visualization
- Visual social media

Viz building blocks

- Collection → data acquisition
- Cleaning → data pre-processing
- Integration → merge data from different sources
- Visualization → create visual representations
- Analysis
- Presentation → create reports
- Dissemination → communicate



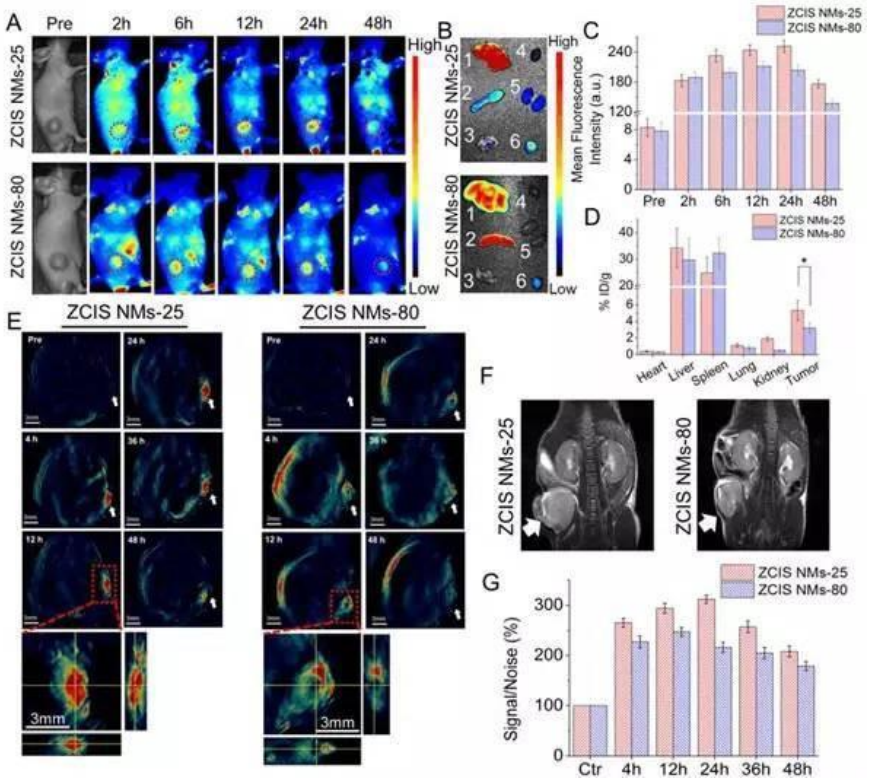
Pre-history of viz

- Selected figures
  - William Playfair (1821) – line, bar charts, etc.
  - Charles Joseph Minard (1869) – Napoleon's march, etc.
  - Jacques Bertin (1967) – "semiology of graphics"
  - John Tukey (1977) – "exploratory data analysis"
  - Edward Tufte (1983) – statistical graphics standards/practices
- 1985 NSF Workshop on Scientific Visualization
- 1990: S.K.Card, et al. Readings in Information Visualization: Using Vision to Think

VISUALIZATION TYPES

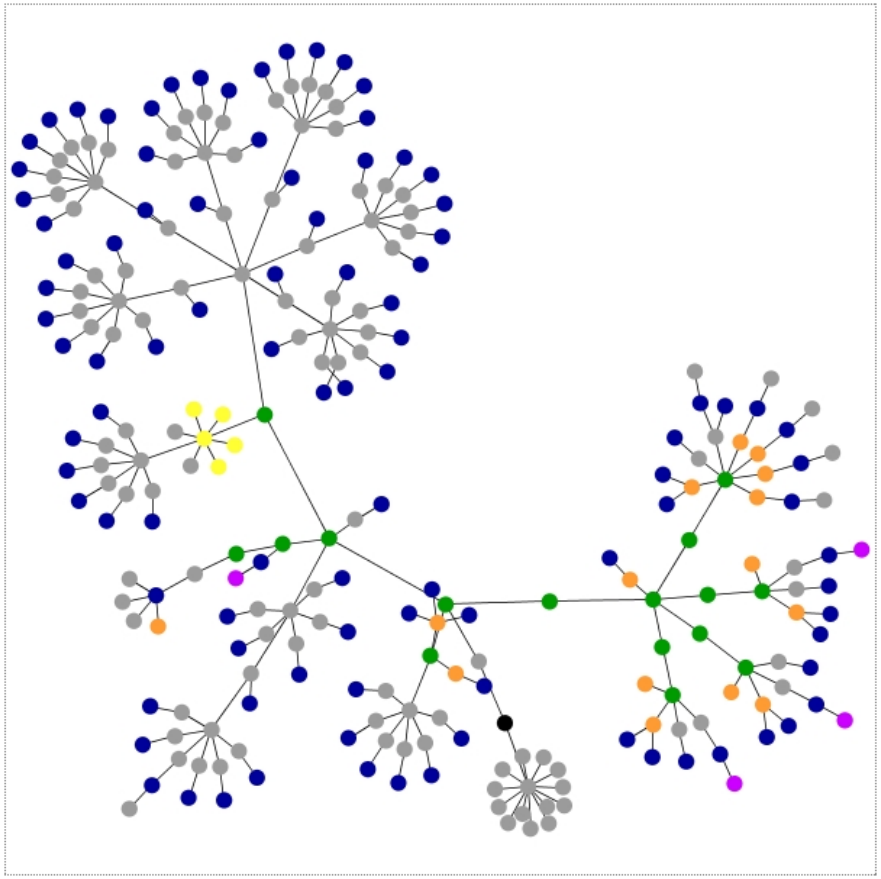
Type #1: Scientific visualization

- **DEF:** Scientific visualization focuses on the 2D or 3D visualization of scientific data.
- Used in:
  - architecture
  - meteorology
  - medicine
  - biological systems
  - …



Type #2: Information Visualization

- **DEF:** Information visualization is the study of interactive visual representations of abstract data to enhance human cognition.
- Transforms abstracts concepts into visually consumable information.
- Includes: histograms, trend graphs, flow charts, and tree diagrams



https://towardsdatascience.com/9-data-visualization-tools-that-you-cannot-miss-in-2019-3ff23222a927

Type #3: Visual Analytics

- DEF: Visual analytics solutions allow analytical reasoning (usually about data) through an interactive visual interface (aka dashboard)

Dashboards in viz analytics

Introduction

- **DEF:** A data dashboard is an information management tool that visually **tracks, analyzes and displays** key performance indicators (**KPI**), metrics and key data points to monitor the health of a business, department or specific process
- A key goal of dashboards in general is to control performance, especially in a business environment, i.e. company
- Features:
  - Customizable to meet the specific needs of a department and company

Functionalities

- **Strategic planning**, e.g. impacts of new business line opened
- **Monitor efficiency**, e.g. lines of source code produced
- **Identify bottlenecks**, e.g. 3rd party supplier always late in delivery
- **Identify negative trends**, e.g. lower sales volume
- **Monitor efficiency of changes made**, e.g. new management installed

3 dashboard types

Strategic

- Develop, view and align company or institutional **strategy**
- Used by business developers and top managers

Tactical

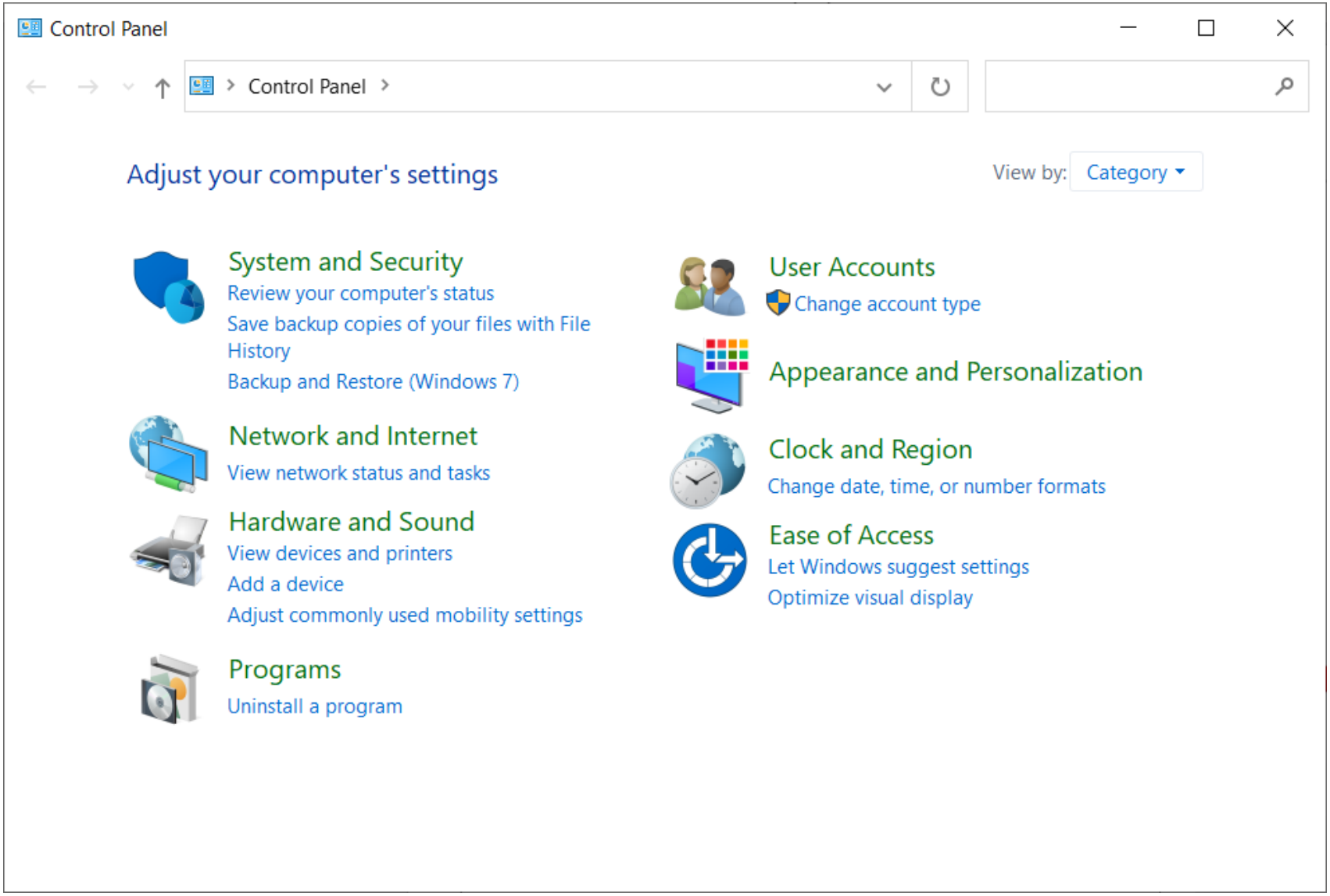
- Measure the progress of **important** projects
- Used by project managers and mid- to top management

Operational

- Detailed monitoring of activities in (near) real-time
- Used by data analysts and up to mid-management, e.g. security analyst
- E.g. SIEM in a SOC

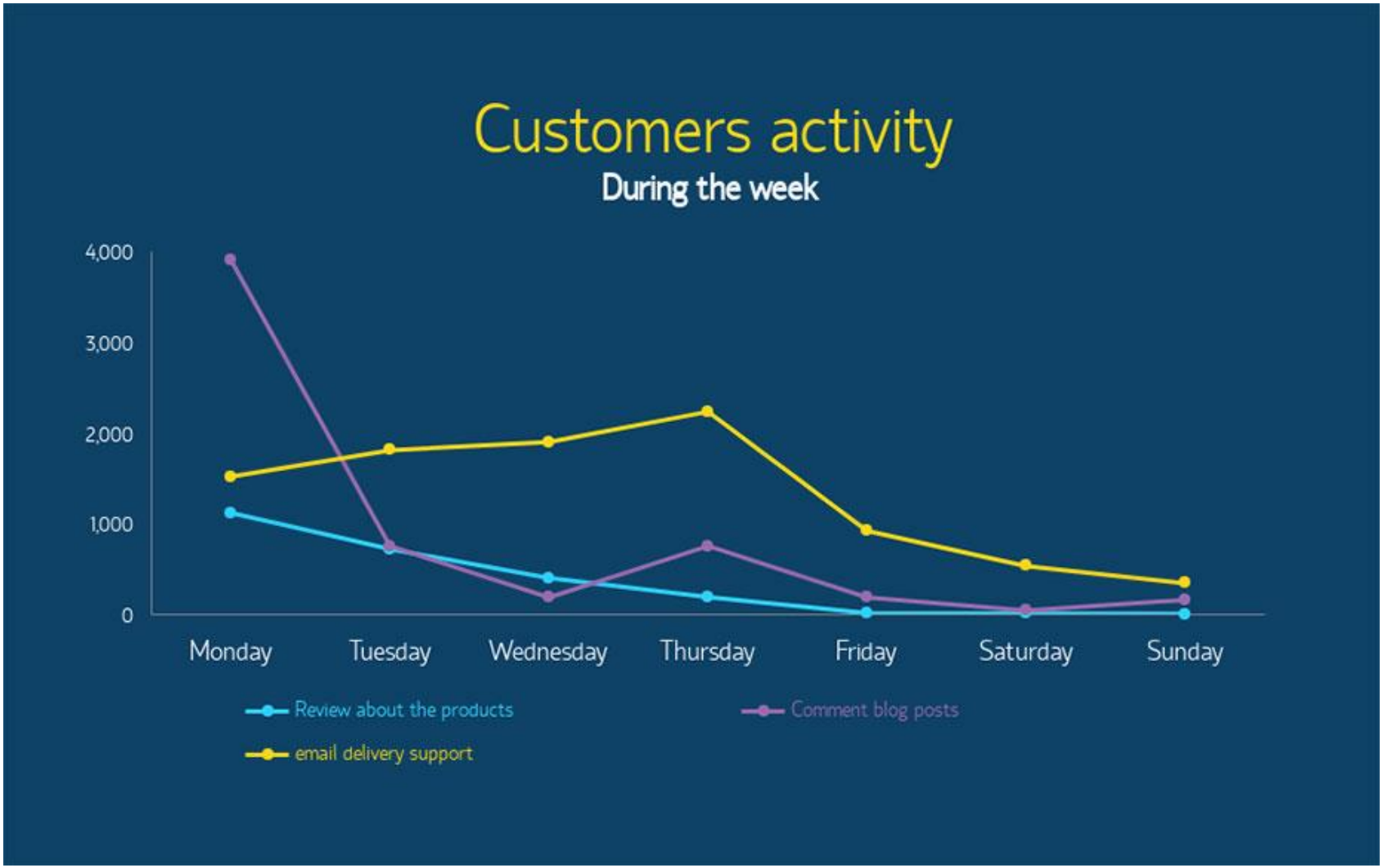
https://www.bidashboard.org/types.html

Dashboard in Windows OS



VISUALIZATION GRAPH TYPES

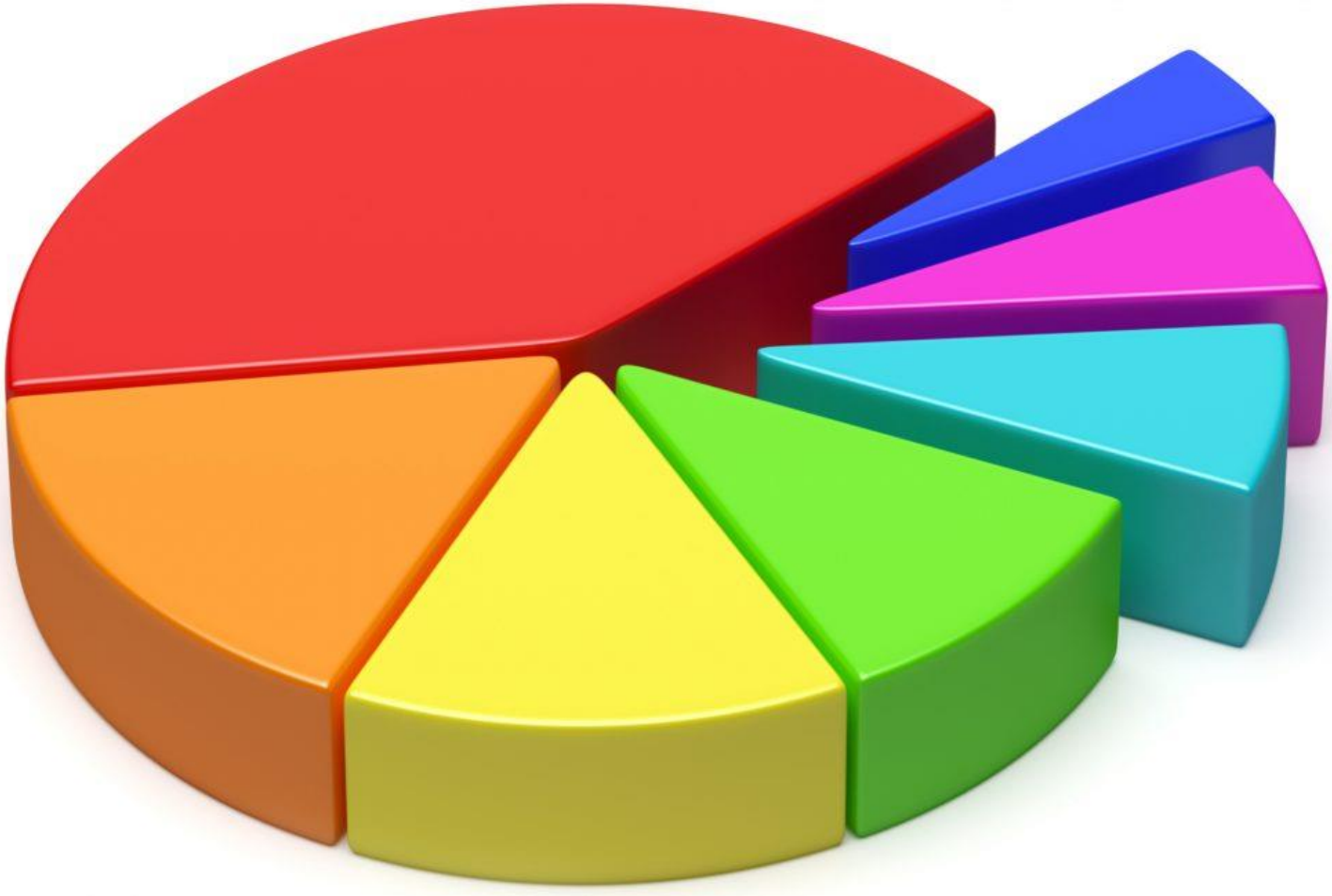
Line Graph


Customers activity
During the week
4,000
3,000
2,000
1,000
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Review about the products
Comment blog posts
email delivery support

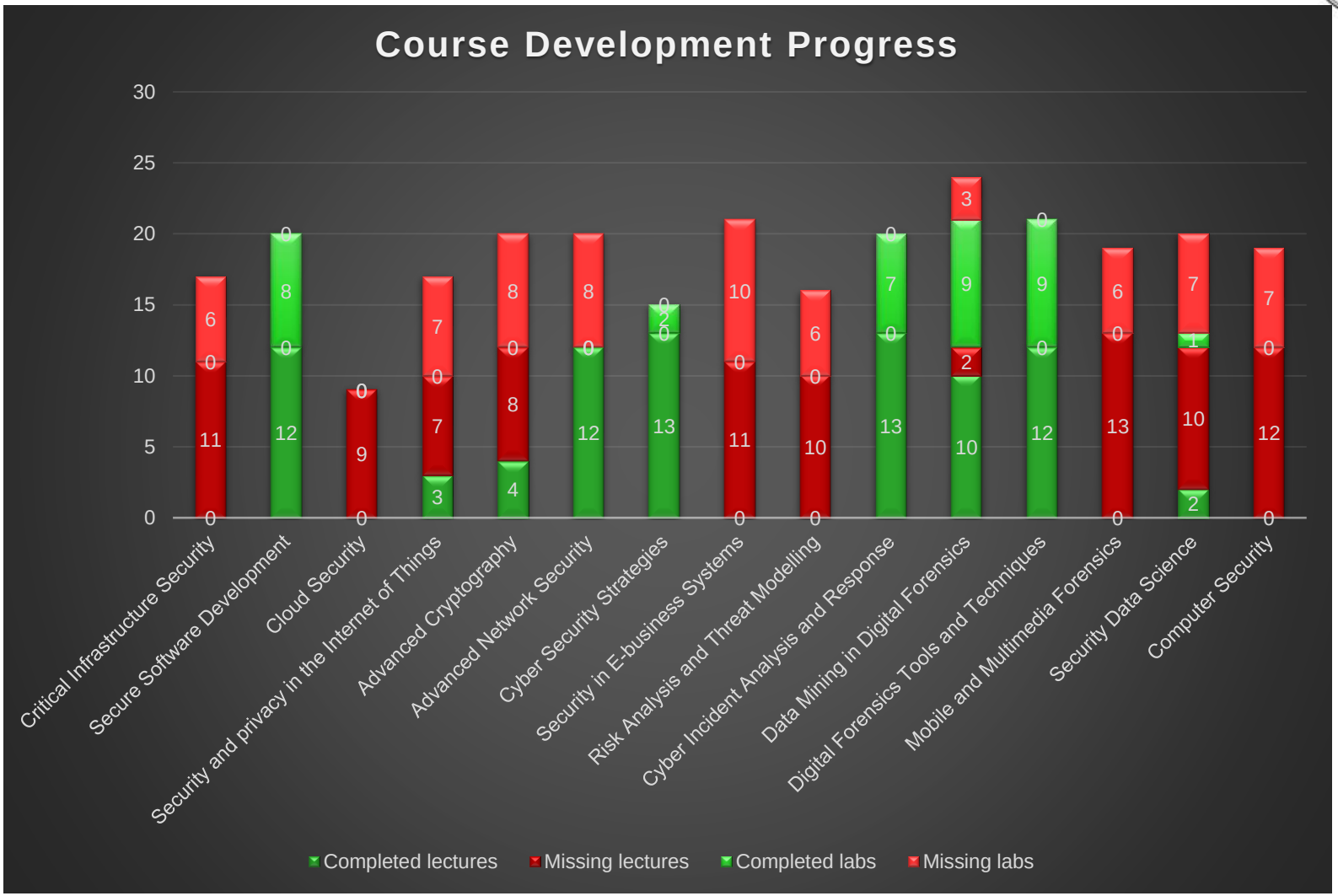
Bar Graph


Usefulness of Weapons to Fight Zombies
Take to Bed
Kick Ass
Don't Share
Not Bad
Keep Looking
Toss
Baseball Bat
Machete
Chainsaw
Chainsaw w/o Gas
Shotgun
Flamethrower
Pillow
Pillowcase Full of Batteries
Atomic Bomb

Pie Chart

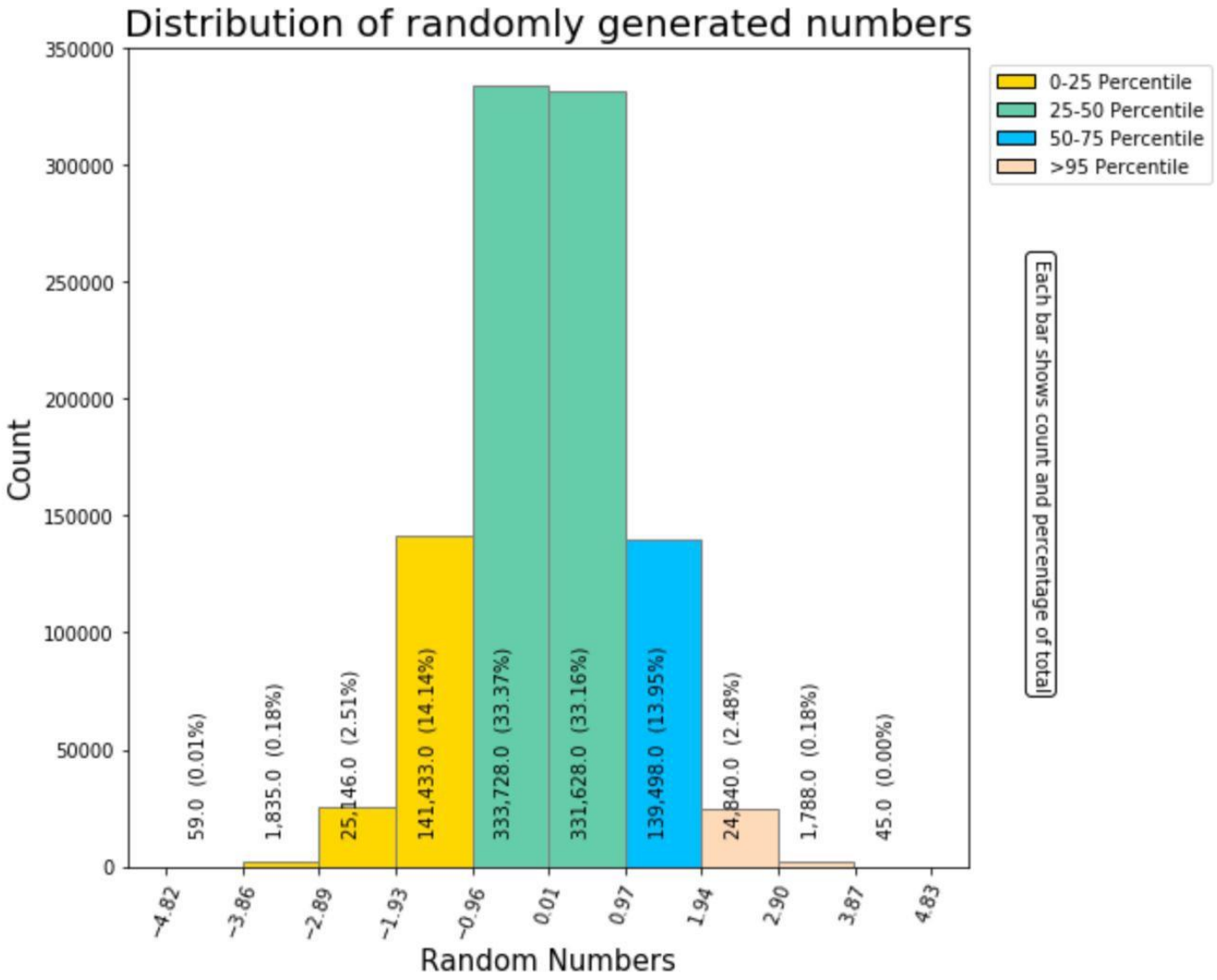


Stacked Bar Graph

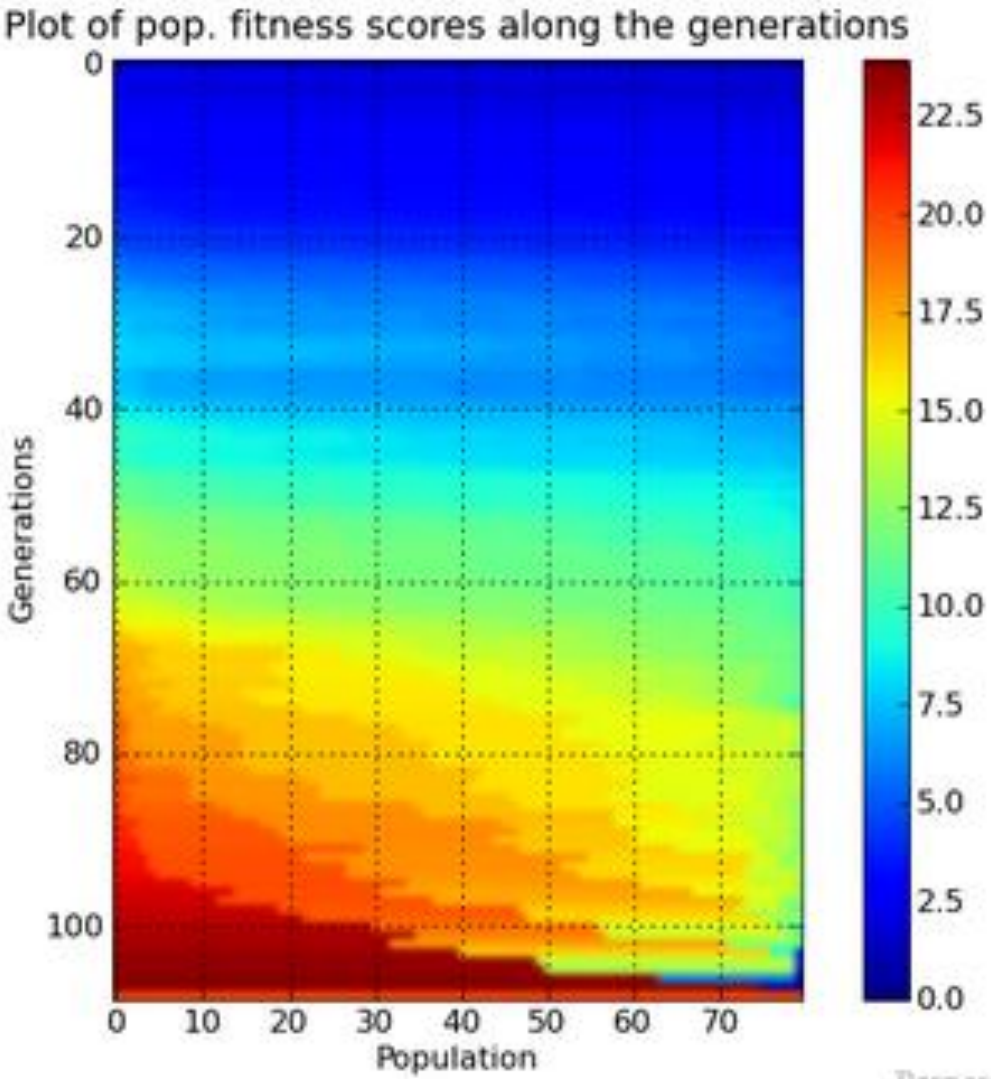


* Information Security Services Education in Serbia, Erasmus+ CBHE Project, www.isses.etf.bg.ac.rs

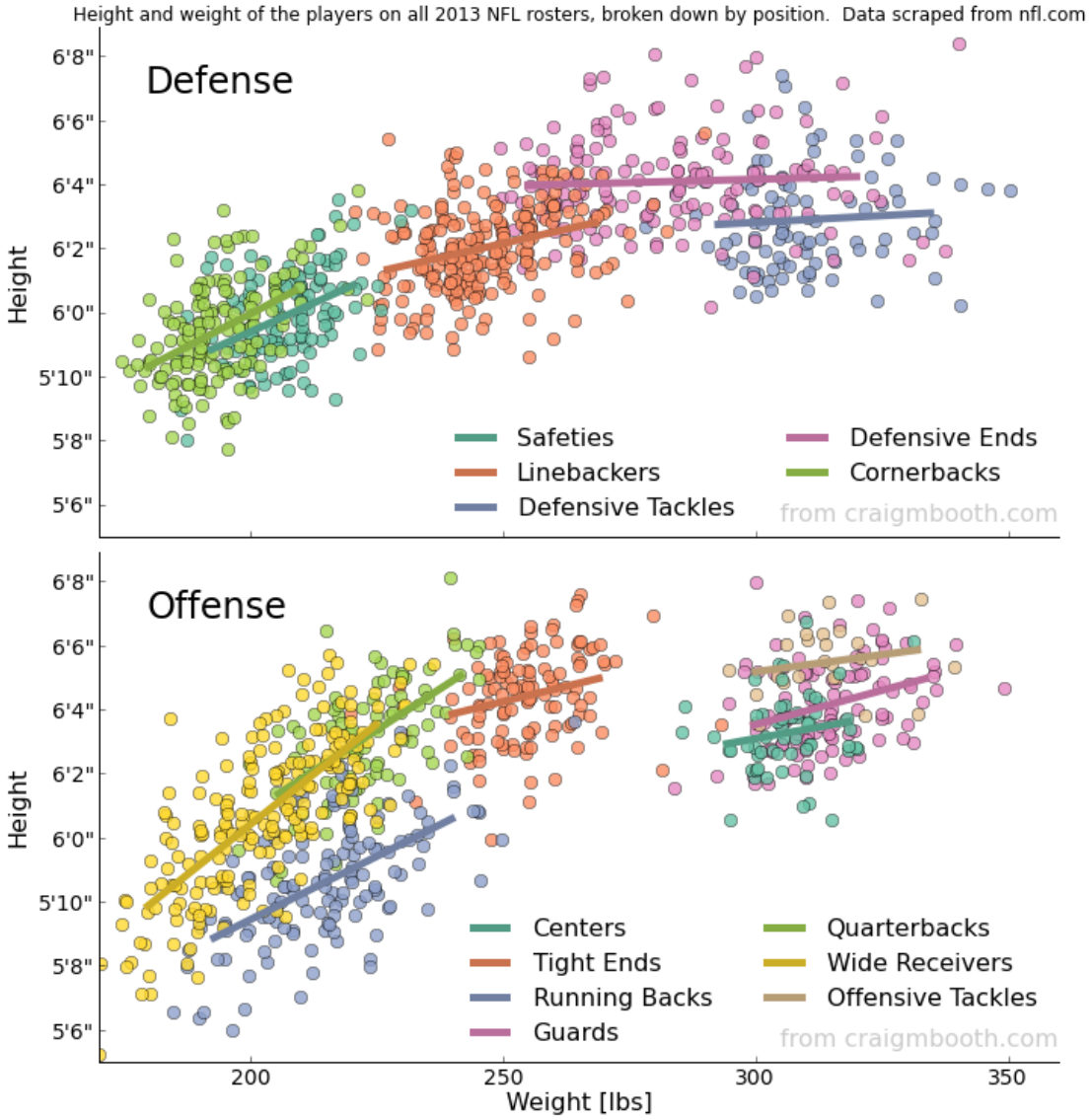
Histogram


Distribution of randomly generated numbers
Count
Random Numbers
0-25 Percentile
25-50 Percentile
50-75 Percentile
>95 Percentile
Each bar shows count and percentage of total
59.0 (0.01%)
1,835.0 (0.18%)
25,146.0 (2.51%)
141,433.0 (14.14%)
333,728.0 (33.37%)
331,628.0 (33.16%)
139,498.0 (13.95%)
24,840.0 (2.48%)
1,788.0 (0.18%)
45.0 (0.00%)
-4.82
-3.86
-2.89
-1.93
-0.96
0.01
0.97
1.94
2.90
3.87
4.83

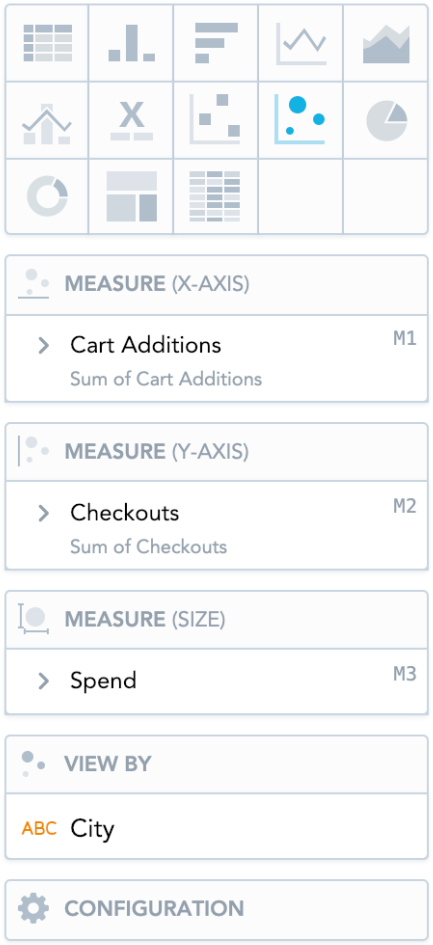
Heat Map

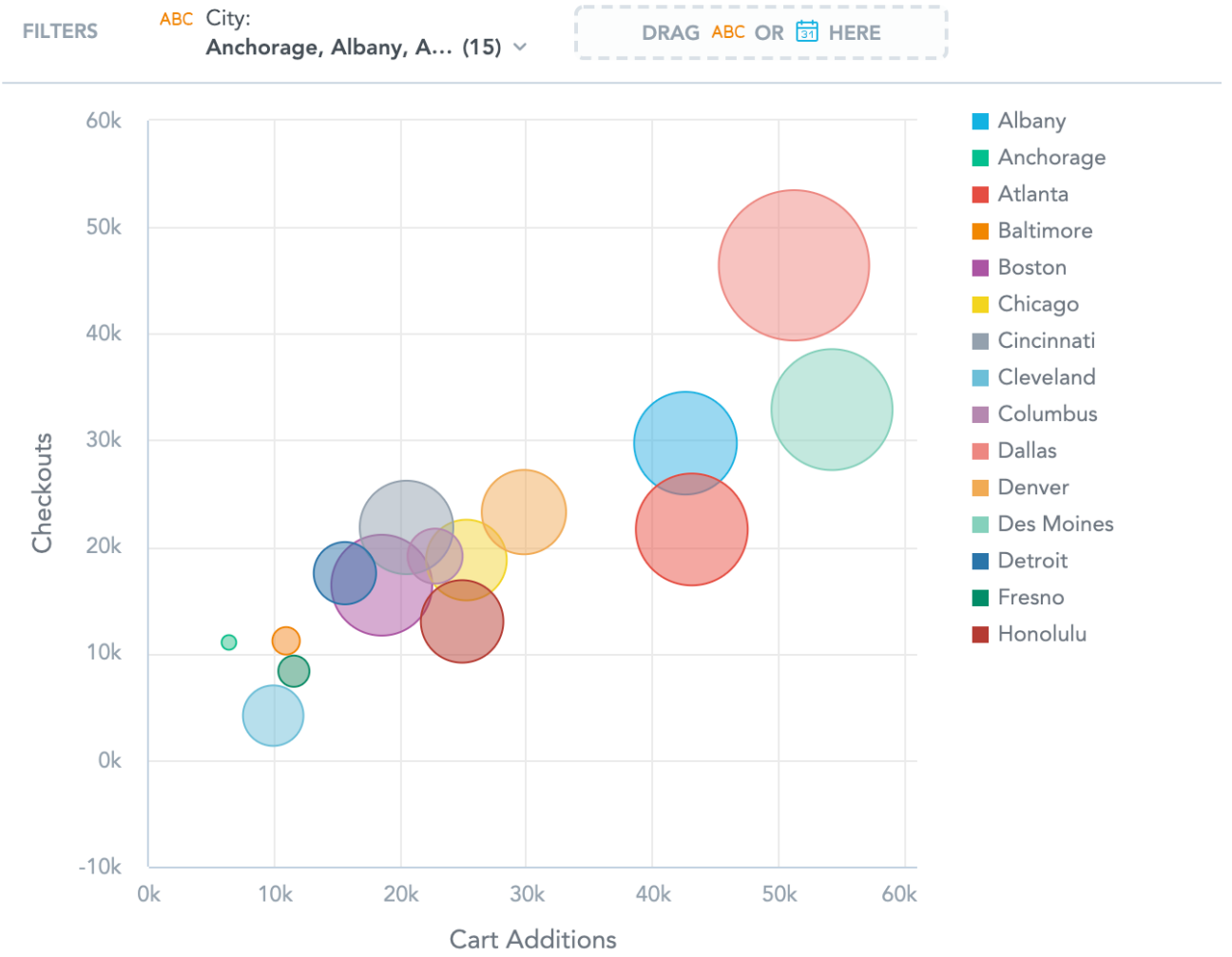

Plot of pop. fitness scores along the generations
Generations
0
20
40
60
80
100
0
10
20
30
40
50
60
70
Population
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Pyevolve 0.5

Scatterplot

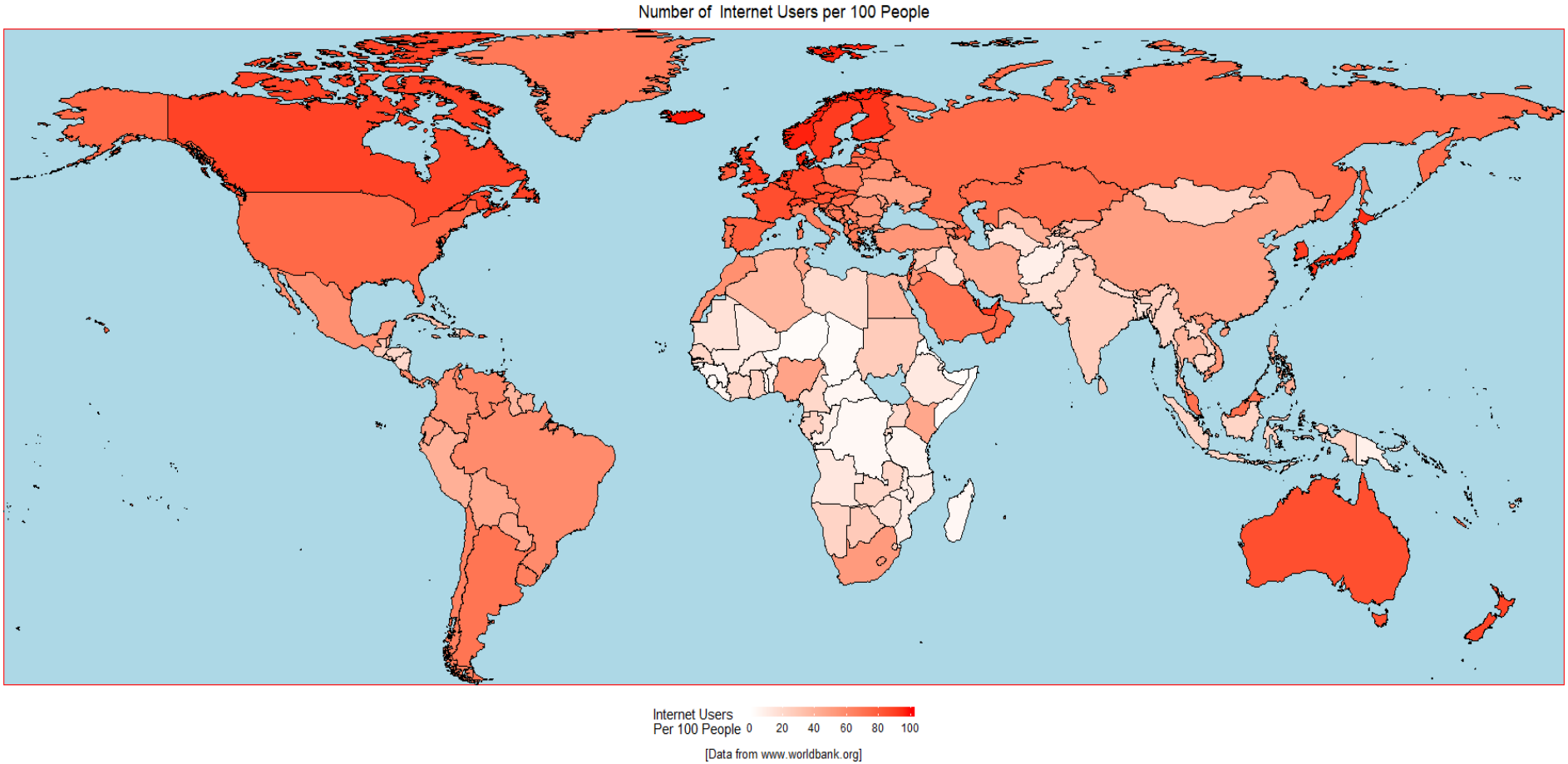


Bubble chart





Choropleth Map



https://www.reddit.com/r/dataisbeautiful/comments/6q811t/choropleth_world_map_of_internet_users100/

Graph functions, i.e. use cases

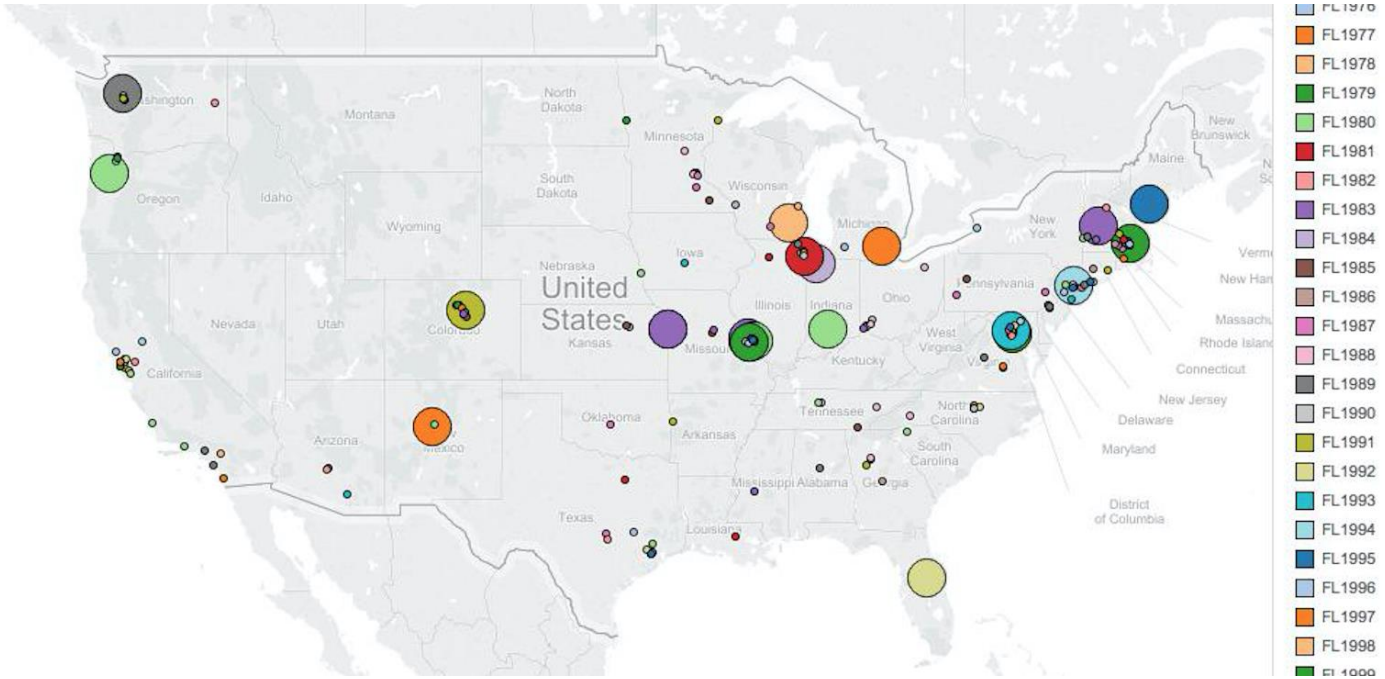
- Line → view trends (over time)
- Bar → compare categorical or time series points
- Pie → compare parts to a whole (up to 4-5 classes!)
- Stacked bar → pie chart alternative, supports more classes
- Histogram → view frequency/distribution
- Heat Map → color-coded frequency
- Scatterplot → relation of (at least) two variables
- Bubble → compare or rank
- Choropleth Map → shade/color on a geo map

Custom visualizations

Eötvös Loránd University (ELTE)
Faculty of Informatics (IK)
Pázmány Péter sétány 1/c
1117 Budapest, Hungary

Olin Fellowship Alumnae in Law

Cecily Stewart-Hawksworth



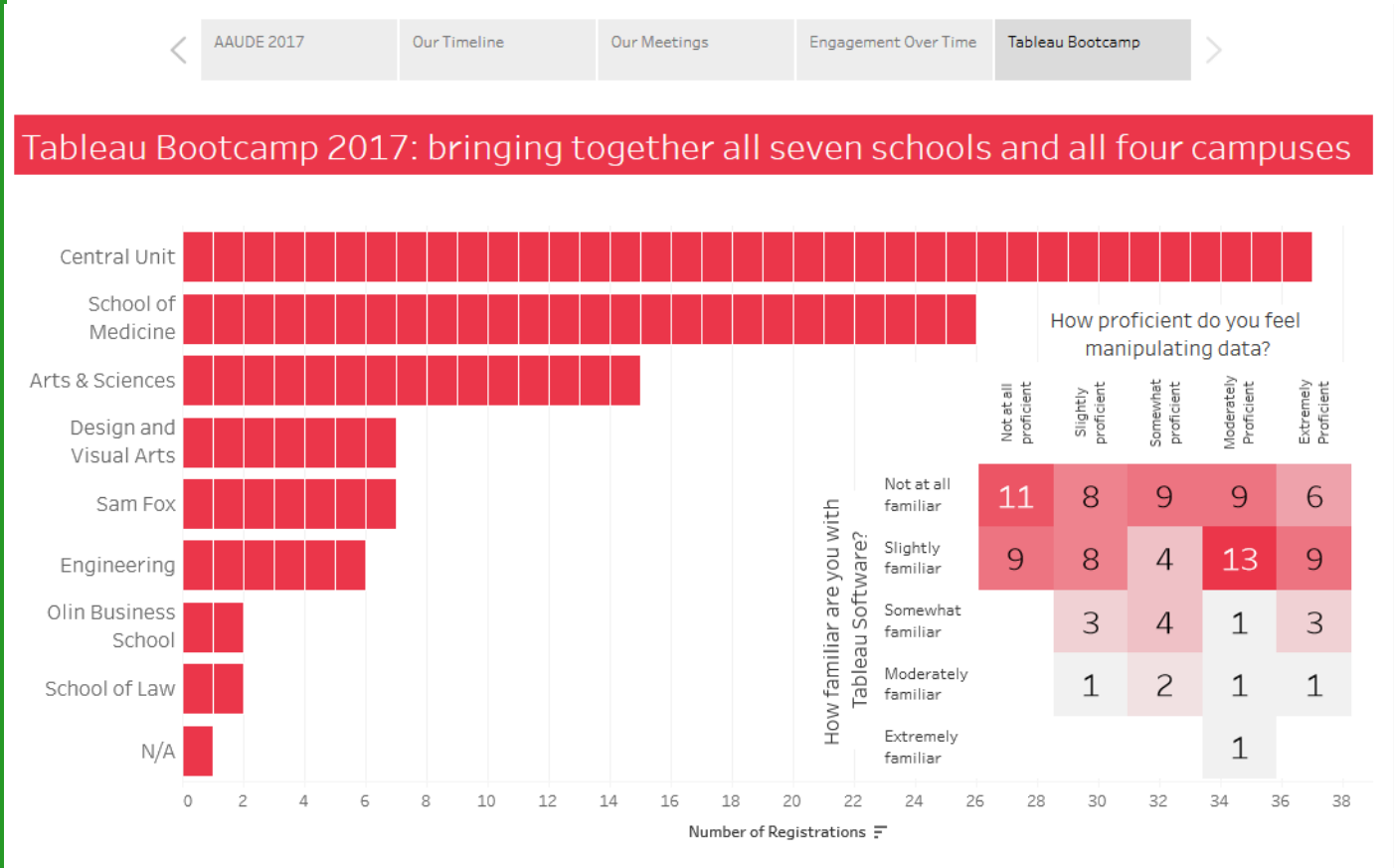
2019
Budapest, Hungary

Eötvös Loránd University (ELTE)
Faculty of Informatics (IK)
Pázmány Péter sétány 1/c
1117 Budapest, Hungary

Survey and Attendance Data

Erin Daugherty

2019
Budapest, Hungary



Interactive visualization

Interactive Visual Analytics

Data preprocessing through visual approaches

- Data mining
- Machine learning
- Statistical methods

- Bring out meaningful:
- patterns
- outliers
- clusters
- gaps

Interactive visualization

- Browse
- search
- monitor

- Discover the most interesting
- relationships among data
- Investigate what-if scenarios
- Verify the presence of biases
- Simulate changes impact

Dissemination tools

- Show the data

- Enlighten the sense of data
- *Tell stories* about them

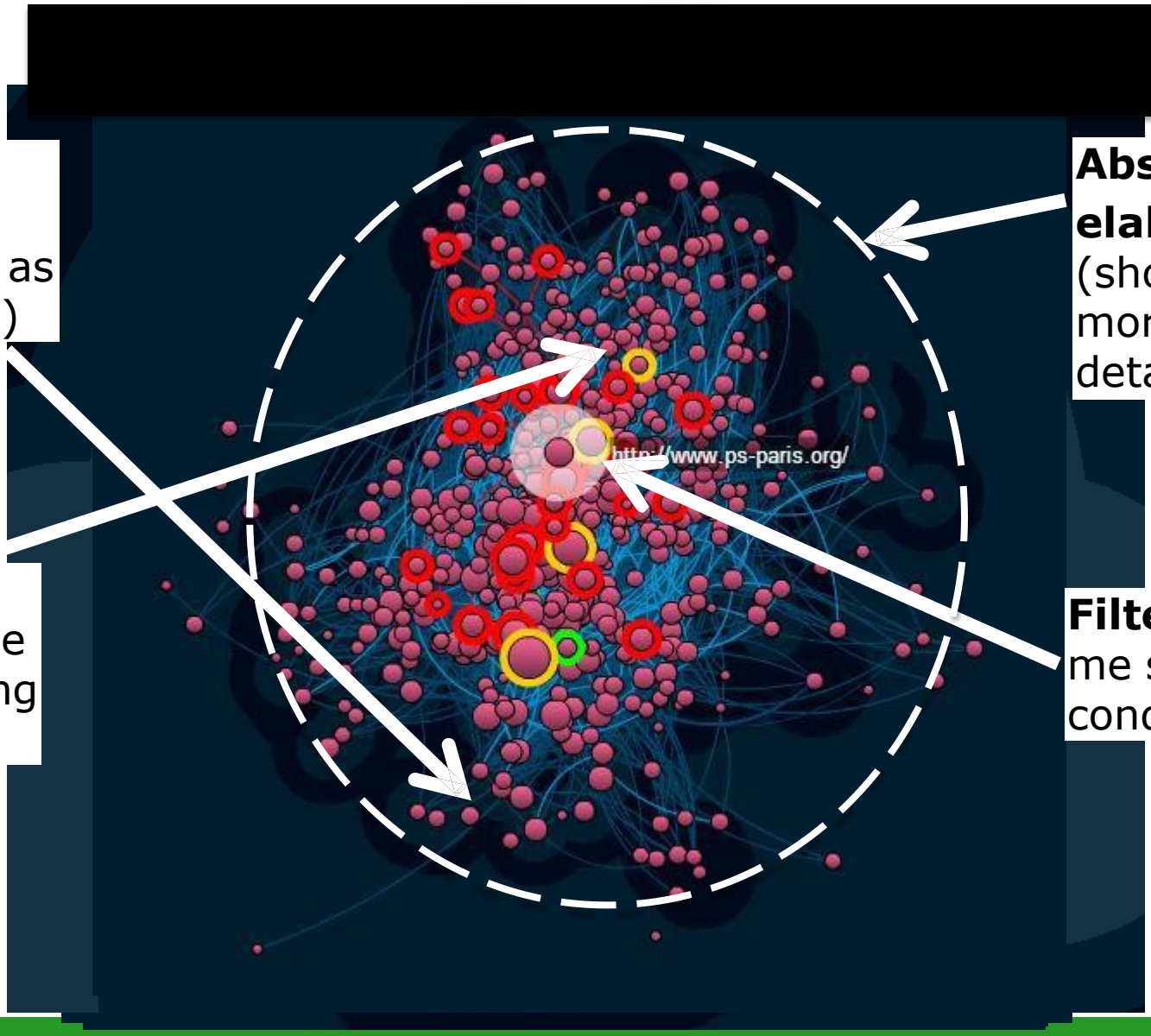
Interactive visualization

- Select (mark something as interesting)
- Explore (show me something else)
- Reconfigure (show me a different arrangement)
- Encode (show me a different representation)
- Abstract/elaborate (show me more or less detail)
- Filter (show me something conditionally)
- Connect (show me related items)

http://www.cs.tufts.edu/comp/250VA/papers/yi2007toward.pdf

Interactive visualization

Select	Ability to mark data items of interest to highlight them	Outlier values
Explore	Enabling users to examine the different subsets in which the data can be divided	Panning across the data
Reconfigure	Provide users with different data perspectives	• Revelation of hidden patterns • visual rearrangements of a series
Encode	Capability of a visualization system to handle and transform the basic elements of human vision	Pre-attentive processing, colours, shapes, dimensions
Abstract/ elaborate	Capability of reduce or increase the details of the visualization	
Filter	Highlight some visual elements that are compliant with specific conditions defined by users	
Connect	Enables users to better emphasize relationships and associations already known or discover the hidden patterns of the data	


Select (mark something as interesting)
Abstract/ elaborate (show me more or less detail)
Explore (show me something else)
Filter (show me something conditionally)
http://www.ps-paris.org/

Viz Tools

Tableau

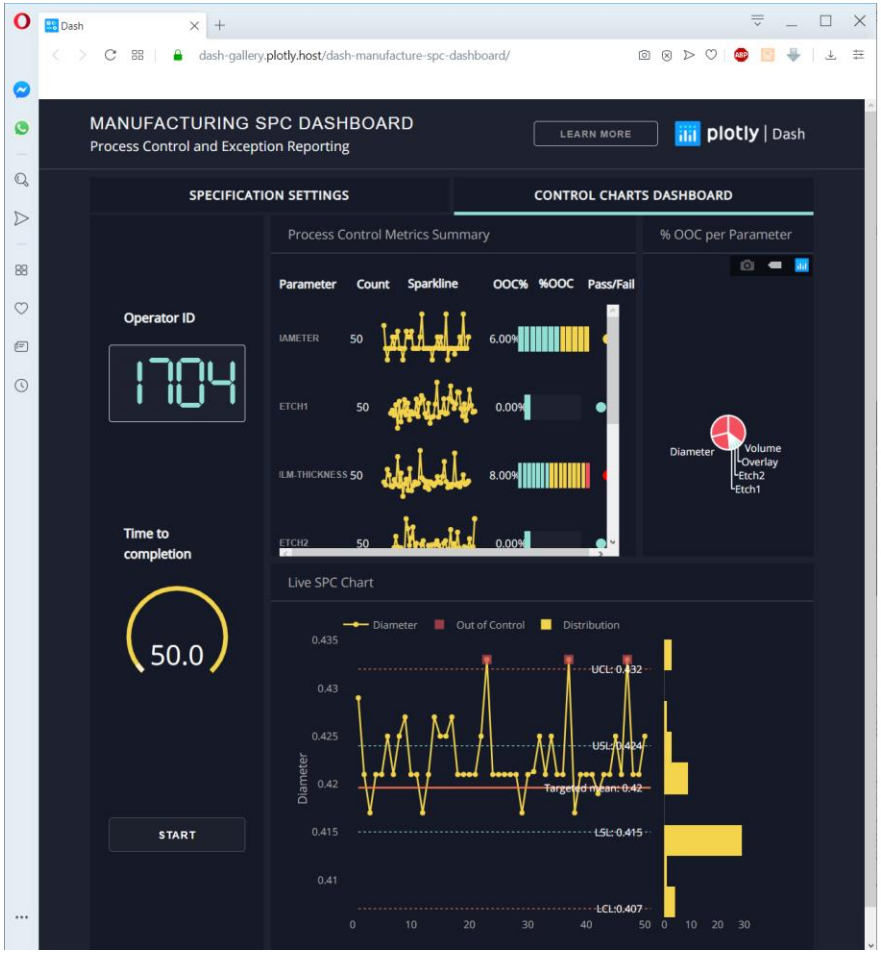
Introduction

- **DEF:** Data visualization software that allows developers to build interactive dashboards that are easily updated with new data and can be shared with a wider audience
  - Developer: Tableau Software Inc, California, USA
  - License: commercial, available for academic use
  - Link: https://www.tableau.com
  - Good: usable for data analysists with minimum programming experience
  - Bad: some licenses paid

Licensing

- Instructors and Researchers
  - Free Desktop license for a year (renewable)
  - Some caveats apply
  - https://www.tableau.com/academic/teaching/course-licenses
- Students
  - Free Desktop license for a year (renewable)
  - https://www.tableau.com/academic/students

Plotly

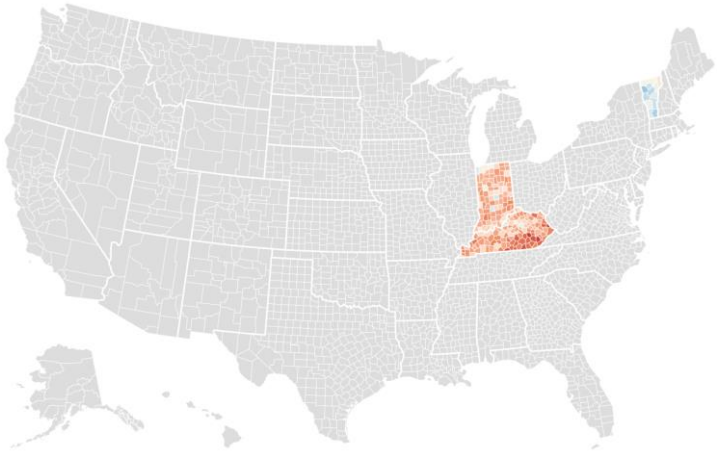


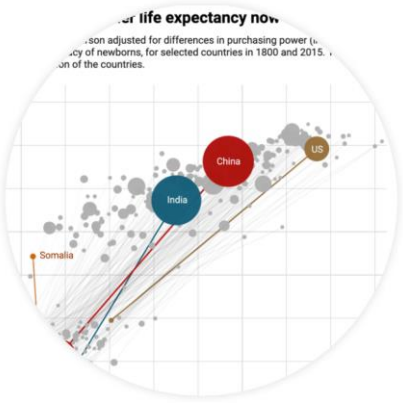
https://dash-gallery.plotly.host/Portal/

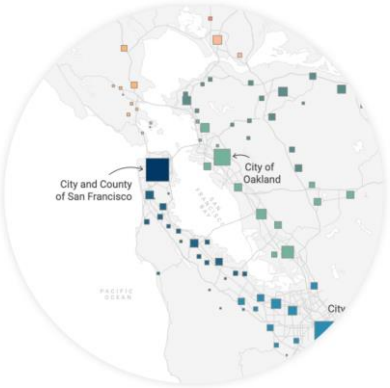
- **DEF:** Web-based platform for operationalizing Python & R models
- **Product:** Plotly Dash
- **Features:**
  - 2D and 3D charts
  - Designer input, i.e. visual customizability
  - Analytics language integrations: Python, R and Matlab
  - Built-in APIs
- **License:** open source, MIT license
- **Used by:** Amazon, Shell, Cisco, Pirelli
- **Link:** https://plot.ly

Datawrapper

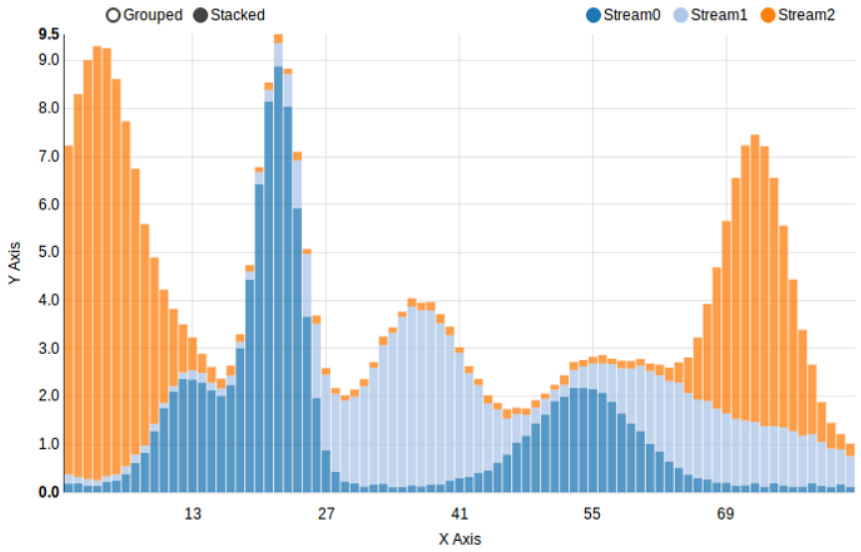
- **DEF:** Datawrapper is an open-source web tool for basic interactive charts.
- **Features:**
  - Free and no sign-up needed
  - 19 chart types, 3 maps + tables (all interactive)
  - Minimum coding skills required
  - Minimum design skills required
  - Interactive charts
- **License:** MIT license
- **Used by:** Fortune, The New York Times, Wired, Süddeutsche Zeitung
- **Link:** https://www.datawrapper.de







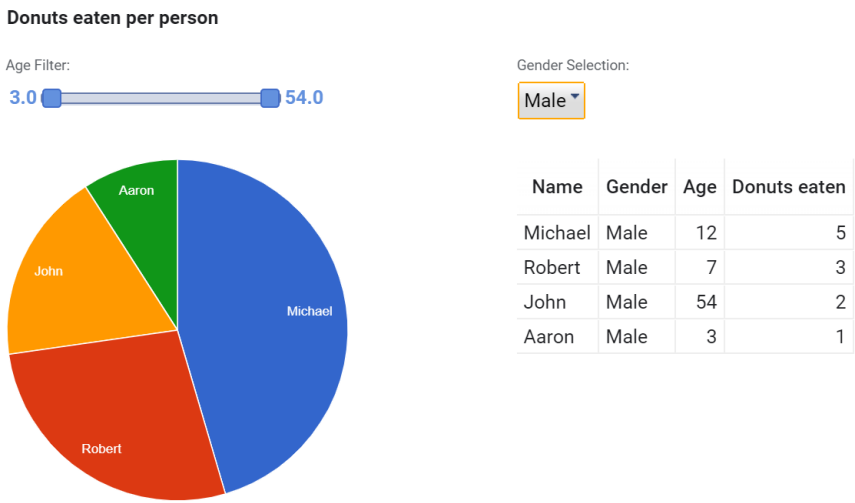
D3.js



- **DEF:** D3.js is a JavaScript library for web-based visualizations
  - Data-Driven Documents
- **Features:**
  - Web-based
  - Interactive viz
  - Downloadable from Github
- **License:** BSD
- **Formats:** Scalable Vector Graphics (SVG), HTML5, and Cascading Style Sheets (CSS)
- **Used by:** Coursera, Akamai
- **Link:** https://d3js.org

Google chart

- **DEF:** Google Charts is Google's big data visualization platform
- **Features:**
  - Completely free
  - Web-based
  - Supported by Google
  - Simple viz types
  - Multi-dimensional viz types
  - Interactive viz
- **License:** Apache 2.0
- **Used by:** BBC, Esquire
- **Link:** https://developers.google.com/chart



Name	Gender	Age	Donuts eaten
Michael	Male	12	5
Robert	Male	7	3
John	Male	54	2
Aaron	Male	3	1

Kibana (part of Elasticsearch)

Kibana features

- Open source data visualization tool
- Visualize ES documents
- Real-time dashboard
- Supports advanced data analytics
- Historical data visualization

Additional features

- Alerting solutions
  - Yelp's ElastAlert
  - Elastic's Watcher
- Shield - authentication and authorization for Kibana

https://logz.io/blog/5-features-weve-added-to-kibana/

More viz tools (A to Z)

- Chartio, https://chartio.com
- Domo, https://www.domo.com
  - Used by: TripAdvisor, Cisco, etc.
- Geckoboard, https://www.geckoboard.com
- Klipfolio, https://www.klipfolio.com
- Sisense, https://www.sisense.com
  - Used by: NASA, NASDAQ, Samsung
  - Merged with: Periscope Data

Summary

- Visualization types
  - Scientific viz
  - Information viz
  - Visual analytics
- Graph types
  - Line, bar, stacked bar, pie
  - Choropleth, scatter, heat
- Interactive visualization
- Viz tools: Tableau, Plotly, Datawrapper, etc.
  - + other viz tools in 2020 listed in the OST intro

Thank you for your attention!