Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

An Introduction to Data and Information Visualization

Evaluation

Beatriz Sousa Santos
October/ 2021

CLAD
Associação Portuguesa de Classificação e Análise de Dados

Evaluation in Visualization

Main Issues for evaluation planning

- Motivation/ goal (why? / what for?)
- Test data (which data sets? How many?)
- Evaluation methods (which methods?)
- Collected data (which measures? Which observations?)
- Data analysis (which methods?)

Much related with the methods

- Visualization (Visual data mining) solutions currently are interactive

- Should be **usable** as any other interactive system

- Usability is, according to ISO 9241-11:

"the extent to which a product can be used by **specified users** to achieve **specified goals** with **effectiveness**, **efficiency** and **satisfaction** in a **specified context** of use"

- How to measure it??

Methods used to evaluate the usability of interactive systems

Some methods from HCI have been adapted to evaluate **Visualization solutions**

- **Analytical** (without users)
  - **Heuristic Evaluation** ←
  - Cognitive Walkthrough
  - Model based methods
  - Review methods
  - ...
- **Empirical** (involving users)
  - Observation
  - Query
  - Controlled Experiments
  - ...

> **usability tests** ←

Heuristic Evaluation (Nielsen and Molich 1990)

- A "**discount usability engineering method**" for quick, inexpensive, and easy evaluation of a UI design
- Most popular usability inspection method; yet is **subjective**
- It is a **systematic inspection** of a design for usability
- Meant to find the usability problems in the design so that they can be attended to as part of an iterative design process
- Involves a small set of analysts judging the UI against a list of usability principles ("**heuristics**")

How to conduct Heuristic Evaluation

- Should be performed by **several evaluators** (one person will never be able to find all the problems)

- Evaluators should work independently:
  - First get a general idea of the UI
  - Then perform a detailed inspection using a set of heuristics
  - Listing **potential usability problems**

    (heuristics not followed and severity degree)

- Findings of all evaluators should be integrated in the same report

1- Each evaluator inspects the interface alone (not communicating among them) (this ensures **independent** and **unbiased** evaluations)

giving for **each potential problem** found:

- a **description** (including the usability heuristic not followed)
- a **severity rate**

2- The results can be recorded as written reports or by having the evaluators verbalize their comments to an observer as they go through the interface

3- Findings are aggregated only after all evaluations have been completed

4- A **list of problems** (description/severity) is given to the development team

https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/

The following 0 to 4 rating scale can be used to **rate the severity** of usability problems:

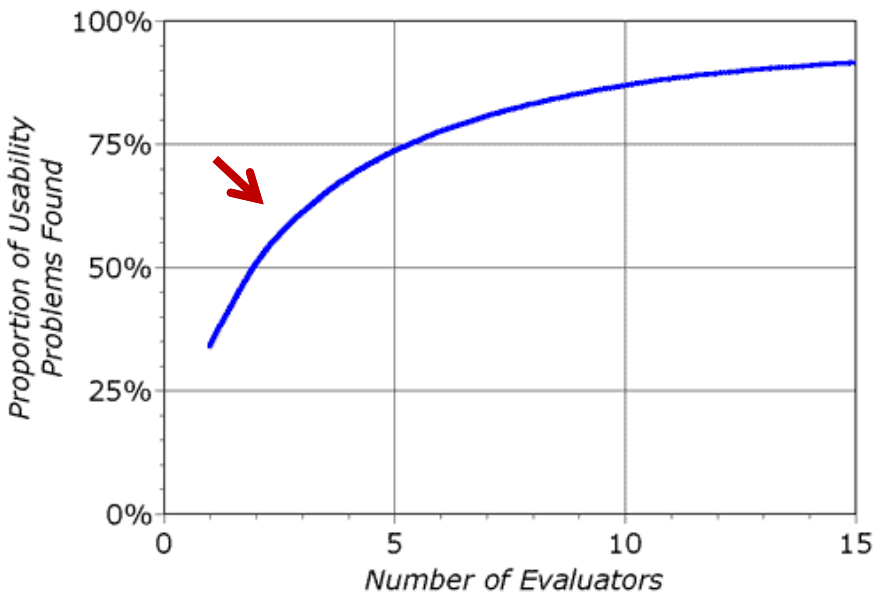
0 = I don't agree that this is a usability problem at all
1 = Cosmetic problem only: need not be fixed unless extra time is available on project
2 = Minor usability problem: fixing this should be given low priority
3 = Major usability problem: important to fix, so should be given high priority
4 = Usability catastrophe: imperative to fix this before product can be released

https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

How to determine the number of evaluators

- Fairly poor results are achieved when relying on single evaluators

- Averaged over six of Nielsen's projects:



http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation

Substantially better performance is possible by aggregating the evaluations from several evaluators

In general :
3 to 5 evaluators seems reasonable in many cases

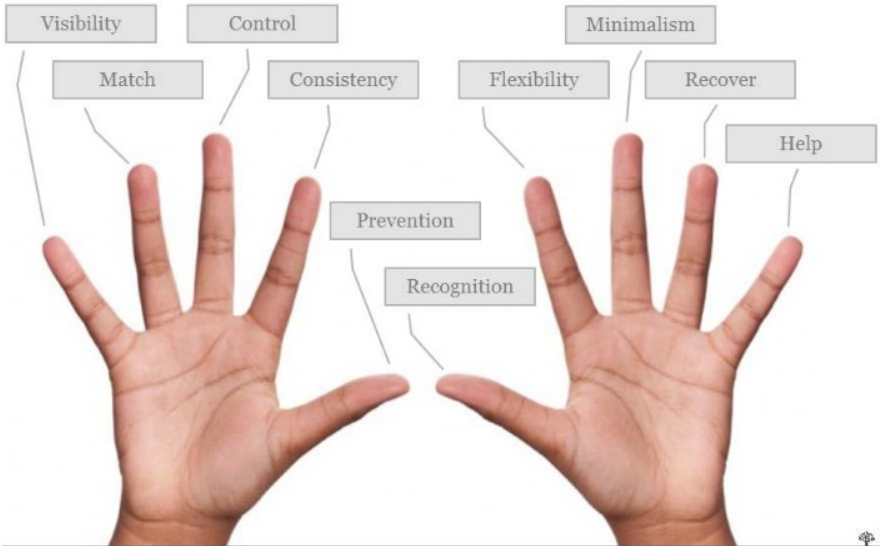
More evaluators (or more experienced) should be used when usability is critical or large payoffs can be expected

- Nielsen proposed **10 general usability heuristics**, yet **there are other sets** (e.g., for web, mobile, visualization applications, for seniors or children…)

- More details on how to conduct a heuristic evaluation at: http://www.nngroup.com/articles/how-to-conduct-a-heuristic evaluation

- And how to rate severity of the usability problems found: http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

- **The list of problems and severity rates should help the development team to priorityse problem fixing**



List of recognized usability principles (“the heuristics”)

1-Visibility of system status

2-Match between system and the real world

3-User control and freedom

4-Consistency and standards

5-Error prevention

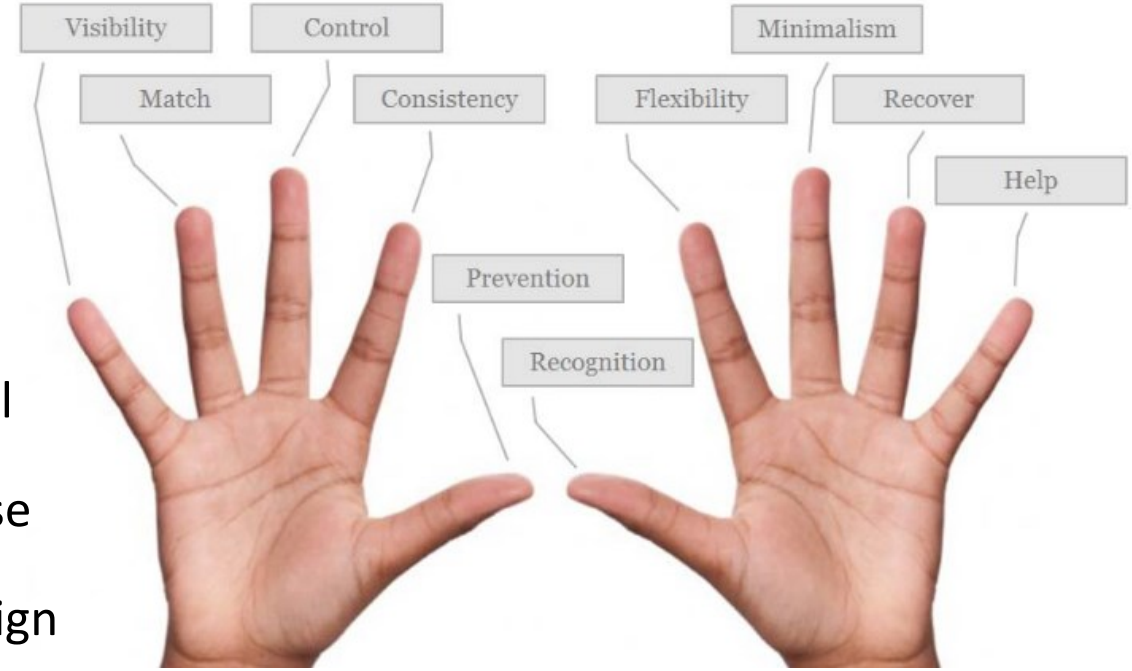
6-Recognition rather than recall

7-Flexibility and efficiency of use

8-Aesthetic and minimalist design

9-Help users recognize, diagnose, and recover from errors

10-Help and documentation



https://www.nngroup.com/articles/ten-usability-heuristics/

Visualization Specific Heuristics

Several lists have been proposed to evaluate visualizations:

- Zuk and Carpendale's Perceptual and Cognitive heuristics (Zuk et al., 2006)
- Forsell's et al. heuristic set for evaluation in InfoVis (Forsell and Johanson, 2010)
- Shneiderman's "Visual Information-Seeking Mantra"
- Freitas's et al. Ergonomic Criteria for Hierarchical Information Visualization Techniques
- Amar and Stasko's Knowledge and task-based framework ….

- ... This is still an active research topic

Zuk and Carpendale's (2006) heuristics

1-Ensure visual variable has sufficient length

2-Don't expect reading order from color

3-Color perception varies with size of colored item

4-Local contrast affects color & gray perception

5-Consider people with color blindness

6-Preattentive benefits increase with field of view

7-Quantitative assessment requires position or size variation

8-Preserve data to graphics dimensionality

9-Put the most data in the least space

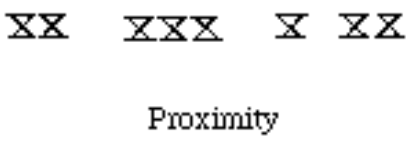
10-Remove the extraneous (ink)

11-Consider Gestalt Laws

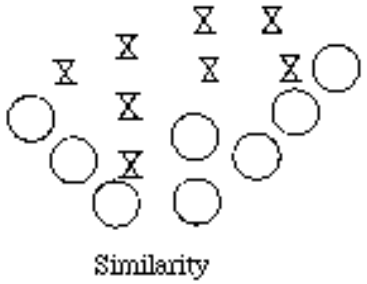
12-Provide multiple levels of detail

13-Integrate text whenever relevant

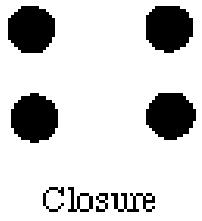
Gestalt Laws



Near stimuli are perceived as a group



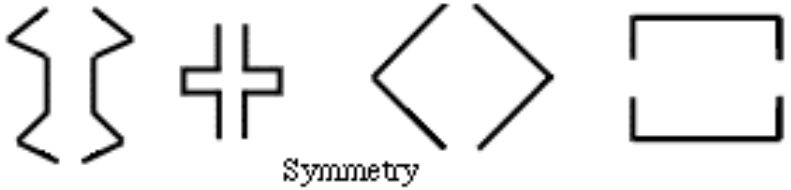
Similar stimuli tend to be grouped
(may override proximity)



Stimuli tend to be grouped
in complete figures

Ambiguous stimuli tend to be resolved
Using the simplest explanation

Stimuli tend to be grouped as to minimize
variations or discontinuities



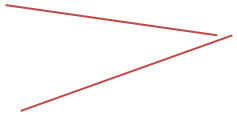
Regions delimited by symmetric tend
to be perceived as coherent figures

Limitations of Analytical Methods

- Are **subjective**
- Involve several usability experts
- Cannot find all usability problems

Thus, empirical methods (involving users) are needed

observation
query
controlled experiments (scientific approach)



Usability test (engineering approach)

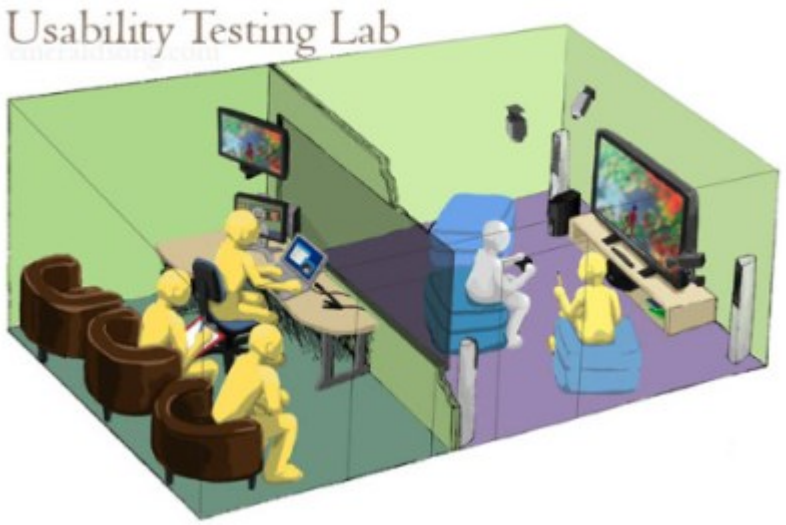
Recommendation: use analytical methods in early phases to debug major usability problems before using empirical methods

Ethics in applying empirical methods

- Involving users implies specific cautions:
  - Asking for explicit consent
  - Confidentiality
  - Safety (avoid any risk)
  - Freedom (users may give up at any time)
  - Limit stress

It's the system that is under evaluation not the user!

Observation



Has many variants from very simple to very complex and expensive:

- Direct: observer takes notes
- Undirect: through audio/ vídeo – more complex and time consuming
- Think Aloud: users are asked to explain what they are doing
- Logging: users activity is logged by the system
- Combinations of the previous, etc

Query

- Two main variants:
  - Questionnaire (reach more people; less flexible)
  - Interview
- Should be carefully prepared and **tested**
- Collected data should be carefully analyzed

System Usability Scale (SUS) (a widely used questionnaire)

- Provides a "**quick and dirty**", tool for assessing the usability
- It includes 10 questions with five response options (Likert-type scale)
- It allows to evaluate a wide variety of products and services (H/W, S/W, mobile devices, websites and applications)
- Has become an **industry standard**, with references in many publications

Benefits of using a SUS

- Is a very easy scale to administer to participants
- Can be used on small sample sizes with good results
- It can differentiate between usable and unusable systems

https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

Usability tests

- Involve **observation** and **query**

- Main aspects:
  - Participants
  - Tasks
  - Test facilities and systems
  - Protocol
  - Usability measures
  - Questions
  - Data analysis

- Have a complex logistics

- May be performed in different ways: very simple to very complex

Example of a usability test of a visual exploration app

based on a web questionnaire to be answered by a user
while observed by an experimenter

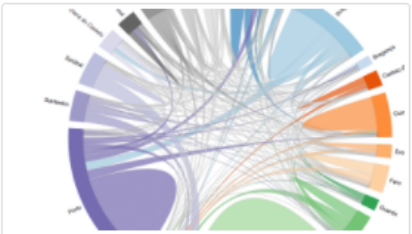
Data

The candidates and institutions data were provided by Direcção Geral do Ensino Superior. The data represents three years (2012, 2013 and 2014) of Portuguese students applications to universities and polytechnic institutions. The dataset has 115636 students applications from 20 districts to 305 institutions.
The geography shapes data is from Direção Geral do Território.

Visualizations

Adjacency Matrix

The adjacency matrix of the network is shown as a two-dimensional grid; each grid cell encodes the number of applicants moving from one district (on the left) to another district (at the top). Adjacency matrices are great for finding clusters (with appropriate sorting),



Chord Diagram

A chord diagram arranges graph nodes (districts) radially, drawing thick curves between nodes. The thickness of a chord encodes the number of applicants moving between districts. Like matrix diagrams, chord diagrams reveal asymmetries: if a chord is tapered,

Map

The map diagram allows you to explore migrations with a geo-spatial reference. Each district is a node, you can click in a district node to visualize the applicants migration; color will help you to understand the net balance of each district and destination.

https://forms.ua.pt/index.php?r=survey/index&sid=489227

In a nut shell:
Do you have a lot of data?

- Visualization may be the solution (or part of it)
- Visualization is **human-in the loop and thus UCD should be used**
- There are **a lot** of visualization techniques
- Should be selected according to the **phenomenon**, **data, users, tasks, and context of use**
- And should **always be evaluated (with users)** before giving them to users (or even use it yourself!)

Practical activity on evaluation

(groups of 3-5 students)

- Select a (preferably not very usable) Visualization/Visual Data Mining application and evaluate it using heuristic evaluation with one of these heuristics sets:
  - Nielsen
  - Zuk et al.
  - Other …

- You may find interesting examples at:
  - Tableau public – gallery
  - Spotfire – gallery

    …
- Each evaluator should read carefully and try to understand each heuristic and analyze independently the application registering the potential problems and their classification

- Discuss the problems with the other group members and consolidate a list of problems

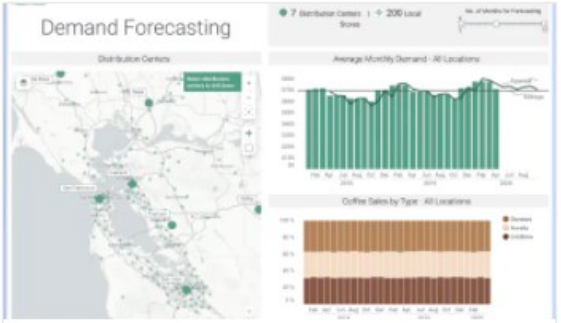
- Prepare a 5 min presentation with the main potential problems you found and send it to bss@ua.pt

TIBCO CUSTOMERS SOLUTIONS SERVICES PRODUCTS PARTNERS CONTACT US TRIALS

Home › Products › TIBCO Spotfire® › TIBCO Spotfire® - Learn

Overview Solutions Interactive Demos Learn Editions & Pricing Free Trial Learn More

COVID-19 Geospatial Hotspot Identification



Spot Coffee Demand Forecasting and Route Optimization

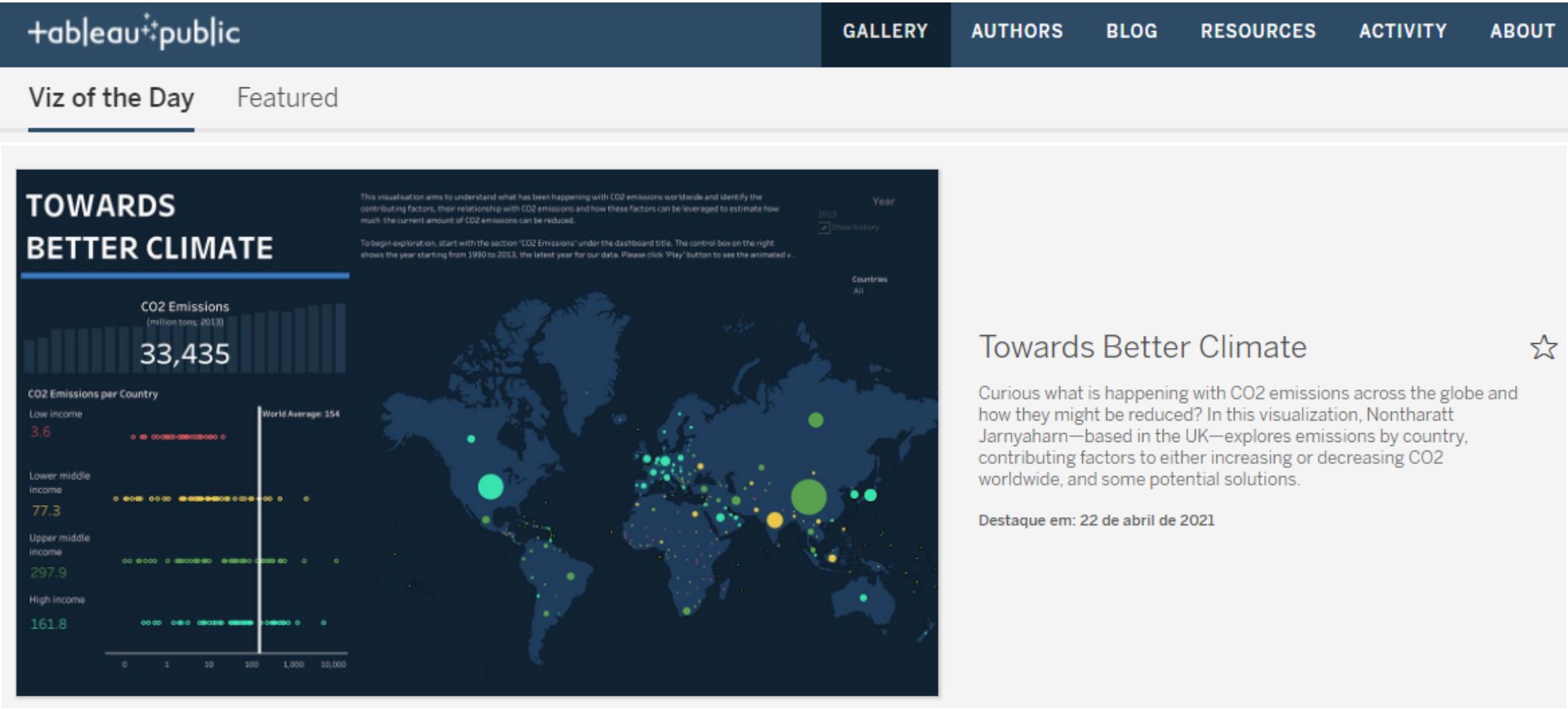
Expense Analyzer Dashboard

Grape Price Elasticity

Delivery Routing

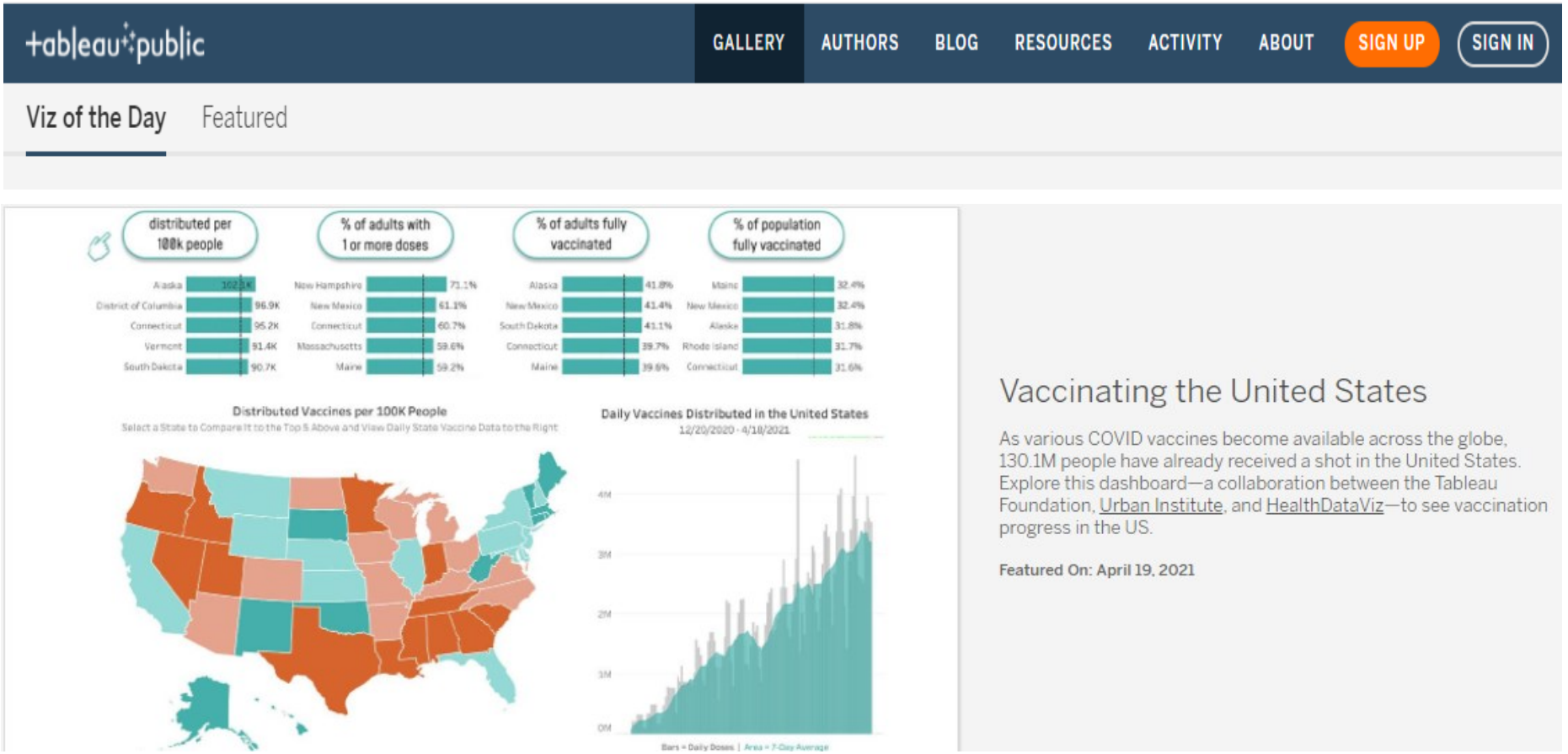
Sales and Marketing

https://www.tibco.com/products/tibco-spotfire/learn/demos



https://public.tableau.com/en-us/s/gallery

Example: vaccinating the united states



https://public.tableau.com/en-us/gallery/vaccinating-united-states?tab=viz-of-the-day&type=viz-of-the-day

Evaluation in Visualization Bibliography

- Carpendale, S. Evaluating Information Visualizations Challenges in Evaluating Information Visualizations. In A. et al. Karren (Ed.), *Information Visualization , Human-centered issues and perspectives*, 2008, pp. 19 – 45.
- Munzner, T. *Visualization Analysis and Design*. A K Peters/CRC Press, 2014.
- Ellis, G. and Dix, A. An Explorative Analysis of User Evaluation Studies in Information Visualisation. *BELIV'06 Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006.
- Forsell, C. , Johanson, J. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces AVI2010,* 2010, pp. 199–206.
- Munzner, T.. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*. 15, 6, 2009, pp. 921–8.
- Zuk, T., Schlesier, L., Neumann, P., Hancock, M. S., & Carpendale, S. Heuristics for Information Visualization Evaluation. In *BELIV'06, 2006,* pp. 1–6.