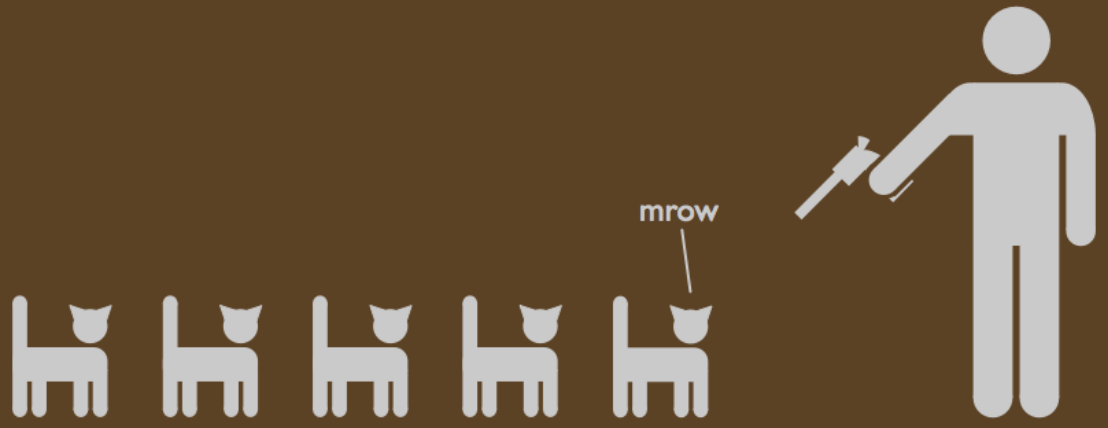
Digital humanities

Vizualizace vědeckých dat

Jindřich Marek

every time you make a powerpoint



edward tufte kills a kitten

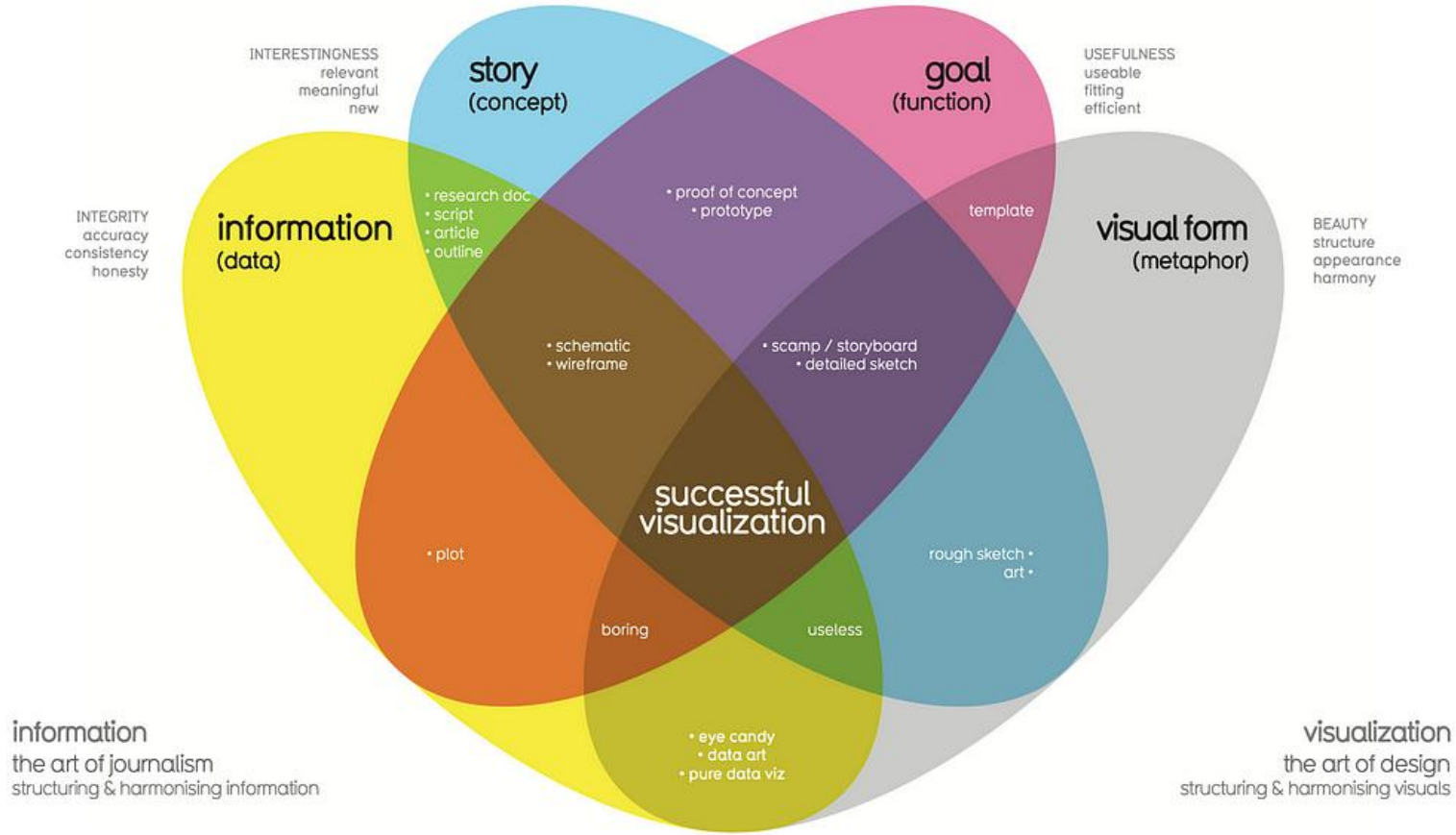
mark goetz

Tufte: grafické zobrazení dat má

- zobrazit data
- přimět diváka, aby se zamyslel nad podstatou, a ne nad metodikou, grafickým designem, technologií grafické tvorby nebo něčím jiným
- vyhnout se zkreslení toho, co data říkají
- prezentovat mnoho čísel na malém prostoru
- scelit velké datové soubory
- podněcovat oči k porovnávání různých částí dat
- odhalit data na několika úrovních podrobnosti, od širokého přehledu až po jemnou strukturu
- sloužit jasnému účelu: popisu, zkoumání, představení struktury dat nebo ozdobě
- být úzce spojeno se statistickým a slovním popisem souboru dat

What Makes a Good Visualization?

explicit (implicit)



information
the art of journalism
structuring & harmonising information

visualization
the art of design
structuring & harmonising visuals

David McCandless
InformationisBeautiful.net

taken from new book
Knowledge is Beautiful

find out more
bit.ly/KIB_Books

Remove

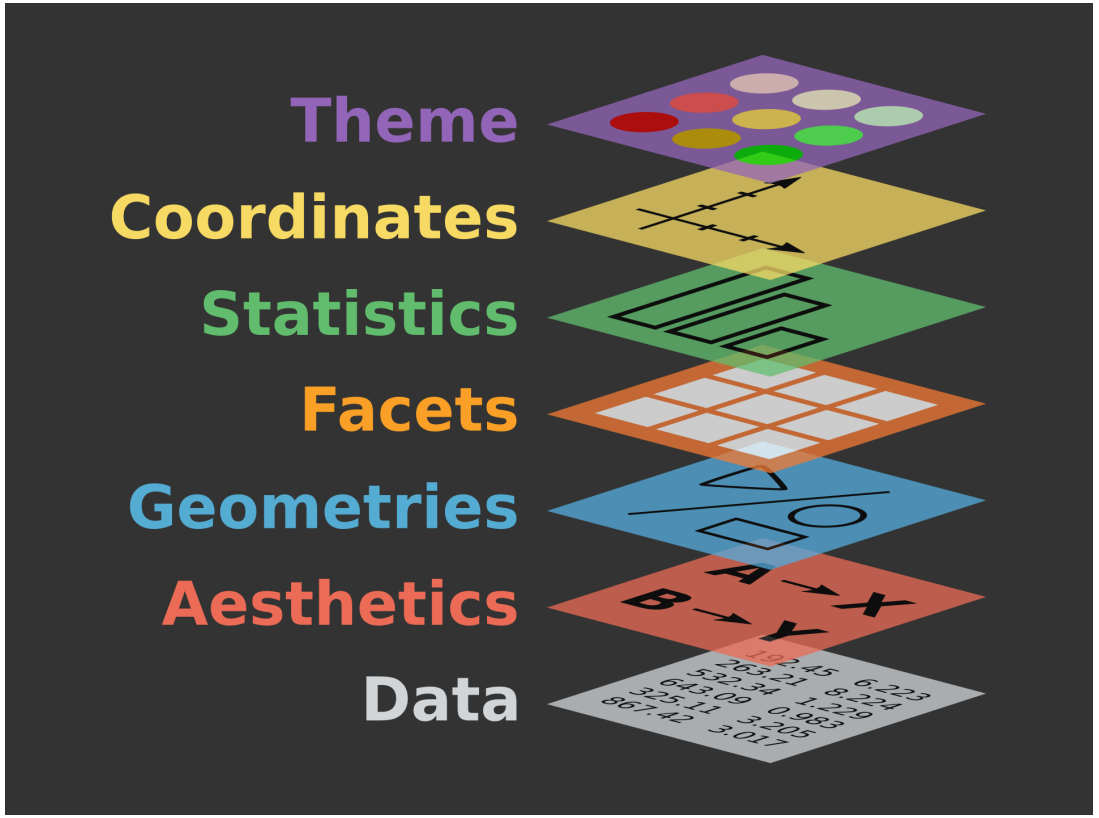
to improve

(the **data-ink** ratio)

Vizualizace: vizuální zobrazení dat

- intuitivní
  - více/méně, časová osa, …
- přístupná
  - relevantní, srozumitelná
- přesná
  - ale případně i shrnující
- jednoduchá
  - abstrakce
- elegantní
  - neodvádějící pozornost

Layered grammar of graphics (Wickham)



- https://vita.had.co.nz/papers/layered-grammar.pdf

- implementace
  - Python: plotnine, případně altair, částečně seaborn
  - R: ggplot2

from Data to Viz

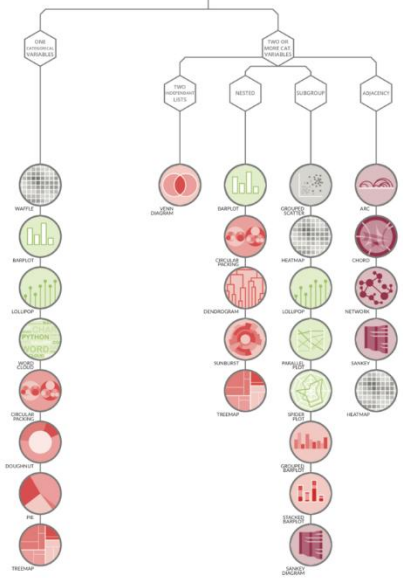
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

1. Identify what type of data you have.
2. Go to the corresponding decision tree and follow it down to a set of possible charts.
3. Choose the chart from the set that will suit your data and your needs best.

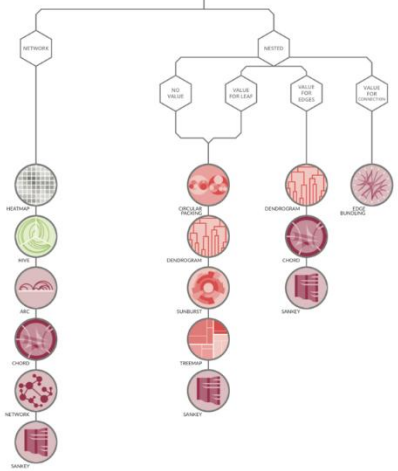
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

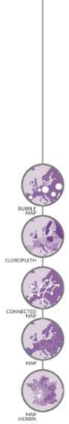
CATEGORIC



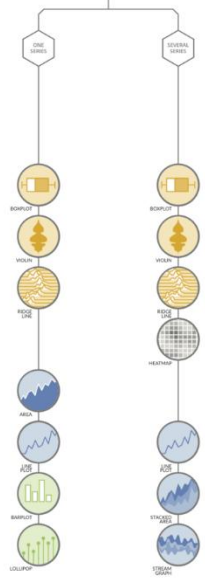
RELATIONAL



MAP



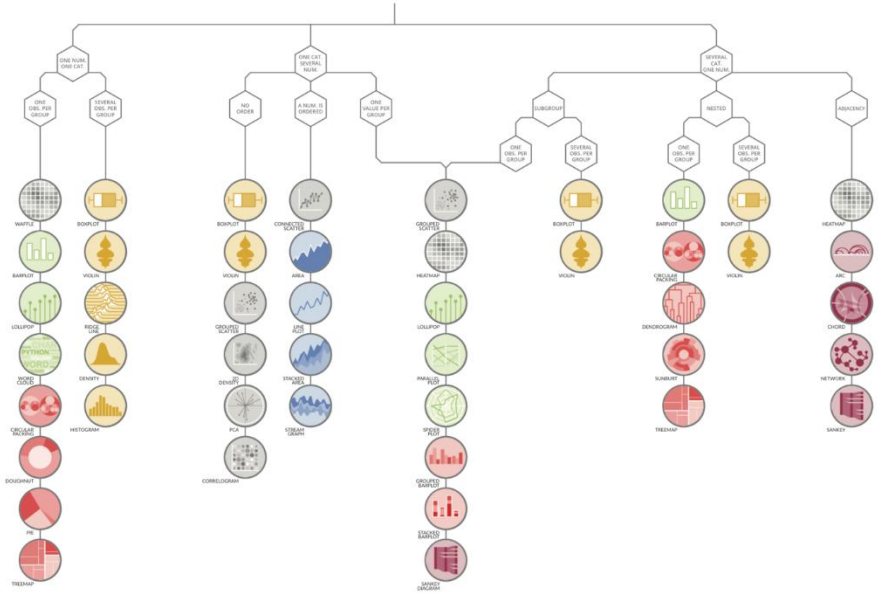
TIME SERIES



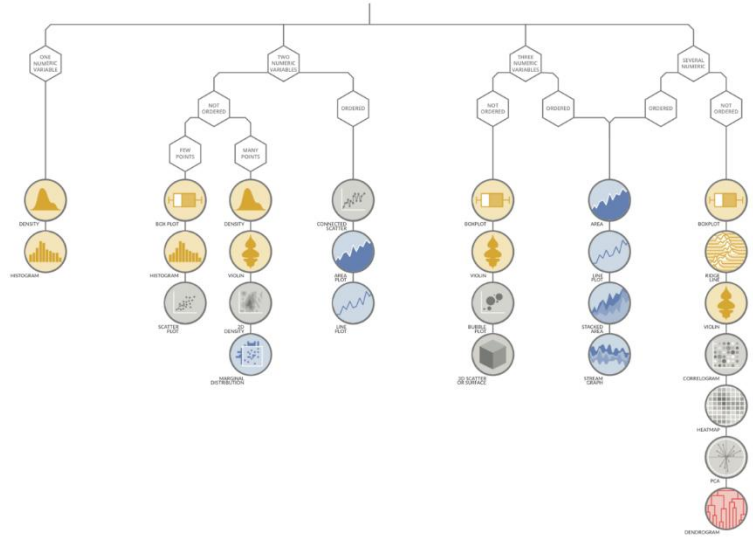
WHAT DO YOU WANT TO SHOW ?

- Distribution
- Correlation
- Ranking
- Part of a whole
- Evolution
- Maps
- Flow

CATEGORIC AND NUMERIC

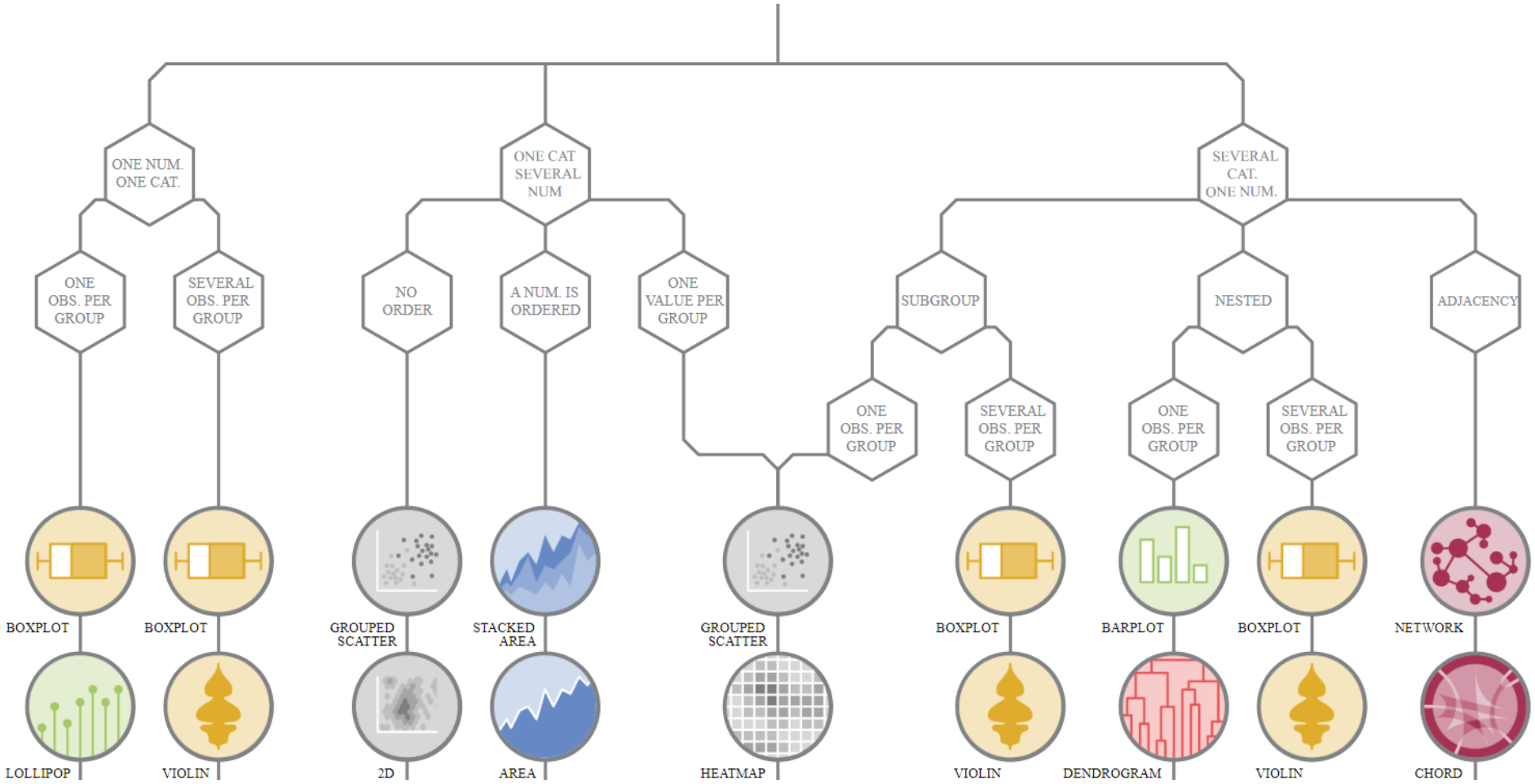


NUMERIC

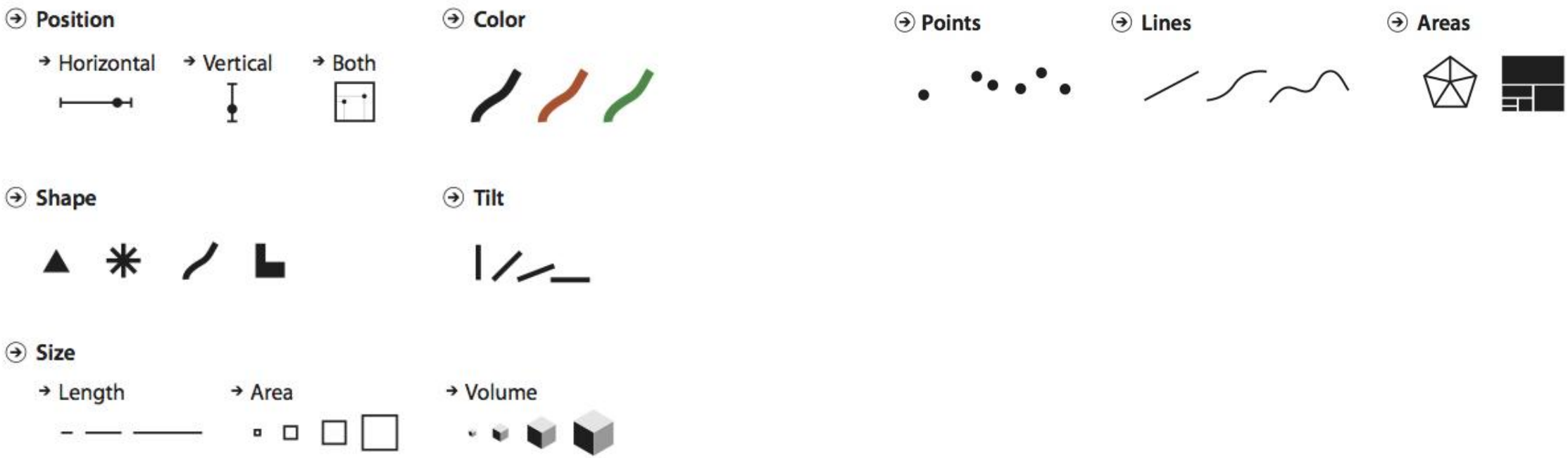


What kind of data do you have? Pick the main type using the buttons below. Then let the decision tree guide you toward your graphic possibilities.

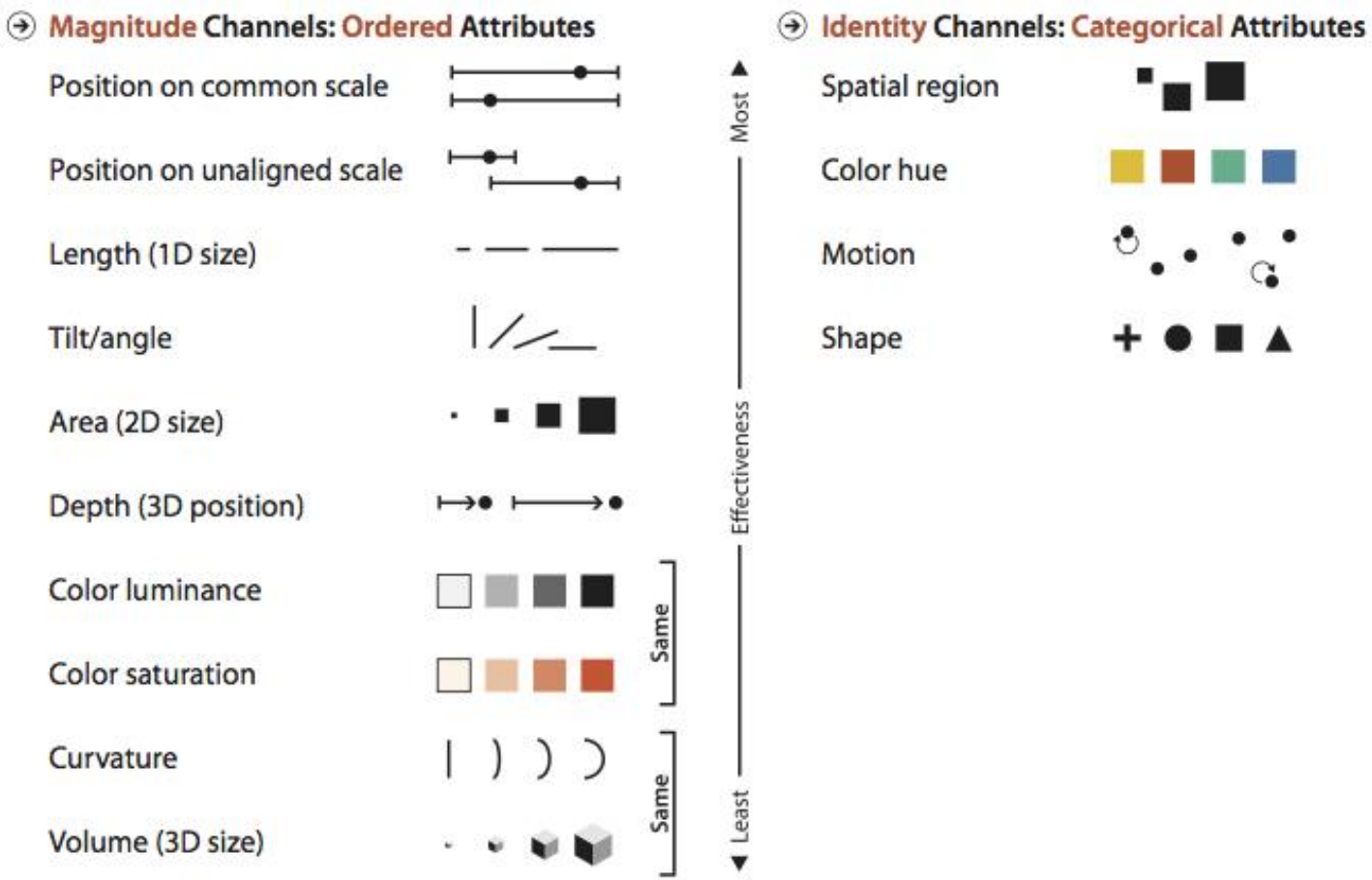
Numeric | Categoric | **Num & Cat** | Maps | Network | Time series



Značky


Position
Horizontal
Vertical
Both
Color
Points
Lines
Areas
Shape
Tilt
Size
Length
Area
Volume

Kanály


Magnitude Channels: Ordered Attributes
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Same
Same
Most
Effectiveness
Least
Identity Channels: Categorical Attributes
Spatial region
Color hue
Motion
Shape

Napoleonovo tažení do Ruska 1812

- Charles Joseph Minard, 1869
  - množství vojáků
  - místa
  - řeky
  - teplota

  - „re-vize“: https://www.datavis.ca/gallery/re-minard.php
  - také mapa k Hannibalovu tažení: https://www.edwardtufte.com/tufte/minard-hannibal

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. ___ Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow en ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

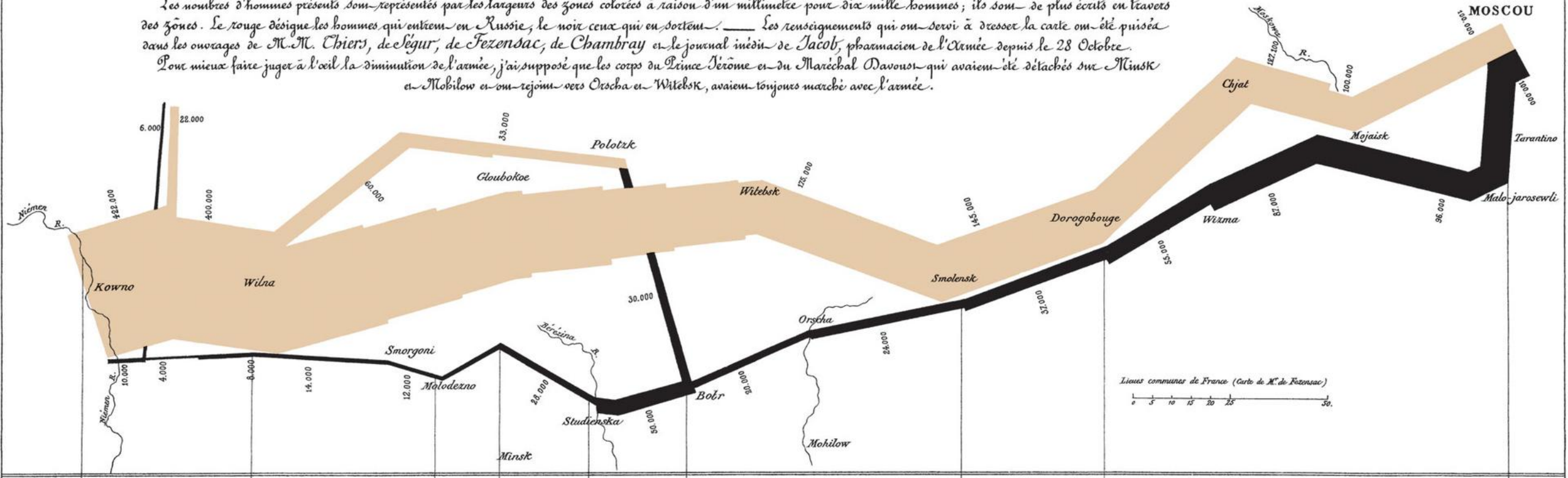
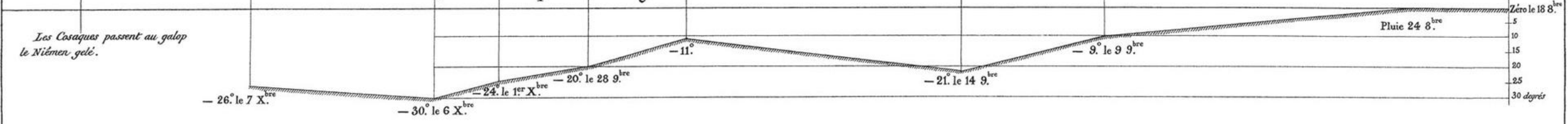


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

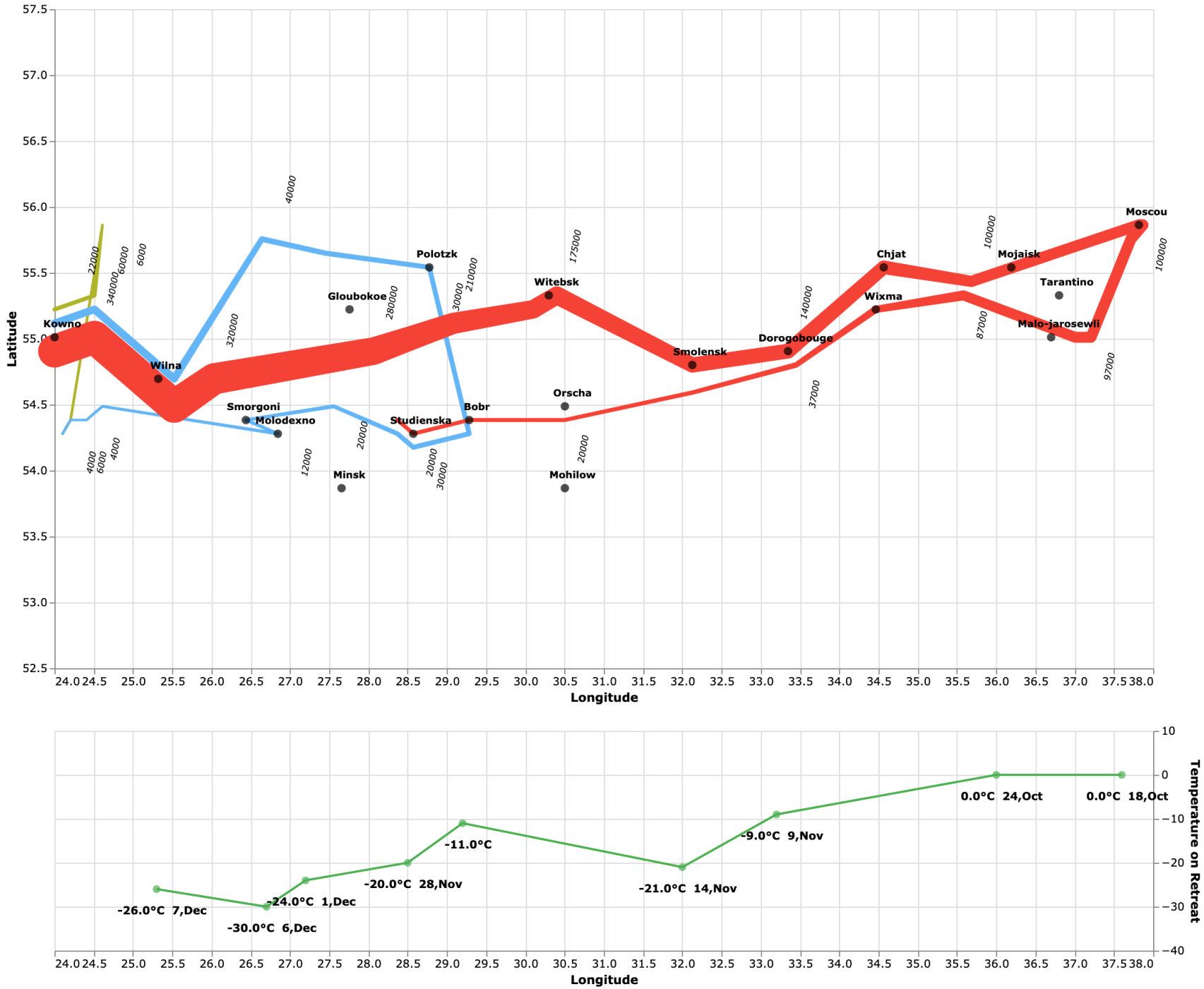


Autog. par Regnier, 8. Pas. Ste Marie St Gain à Paris.

Imp. Lith. Regnier et Dourdet.

Minardův graf v Pythonu

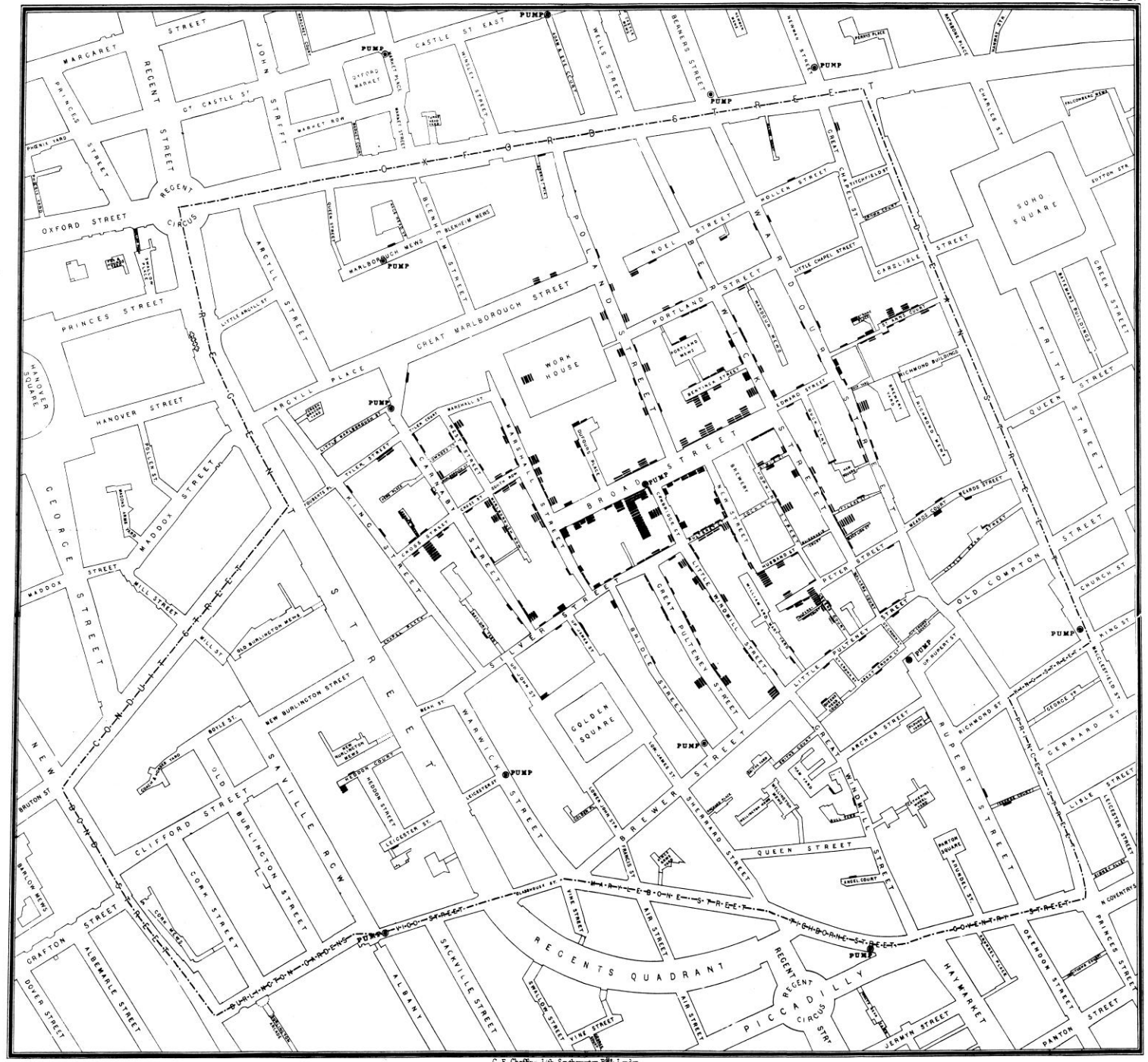
- https://github.com/tulsyanp/data-visualization-python-altair-charles-minards-map


Latitude
Longitude
Kowno
Wilna
Smorgoni
Molodexno
Minsk
Polotzk
Gloubokoe
Studienska
Bobr
Witebsk
Orscha
Mohilow
Smolensk
Dorogobouge
Chjat
Wixma
Mojaisk
Tarantino
Malo-jarosewli
Moscou
22000
60000
6000
340000
40000
320000
4000
6000
4000
12000
20000
280000
30000
210000
20000
30000
175000
20000
140000
37000
87000
100000
97000
100000

Temperature on Retreat
Longitude
-26.0°C 7,Dec
-30.0°C 6,Dec
-24.0°C 1,Dec
-20.0°C 28,Nov
-11.0°C
-21.0°C 14,Nov
-9.0°C 9,Nov
0.0°C 24,Oct
0.0°C 18,Oct

Cholera v Londýně, 1854

- John Snow a jeho mapa
  - https://flowingdata.com/tag/john-snow/
  - Wikipedie: https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak
  - nová reprezentace dat:
    - https://www.leahmeisterlin.com/remappingsnow
    - https://www.r-bloggers.com/2013/03/john-snows-cholera-data-in-more-formats/



C.F. Cheffins, Lith. Southampton Bldgs. London.

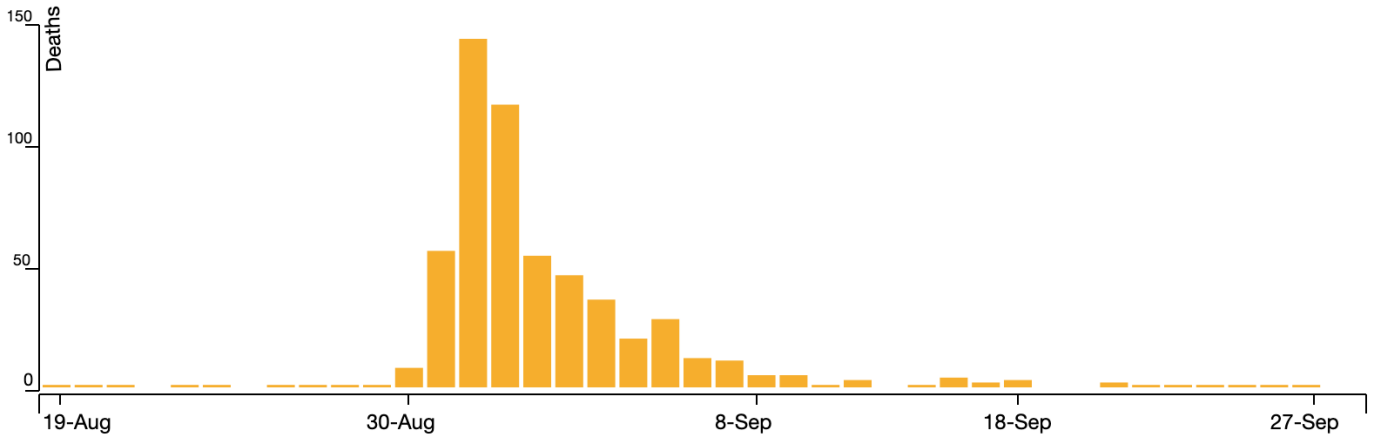
SCALE 30 INCHES TO A MILE.

Snowova mapa v Pythonu

Python:

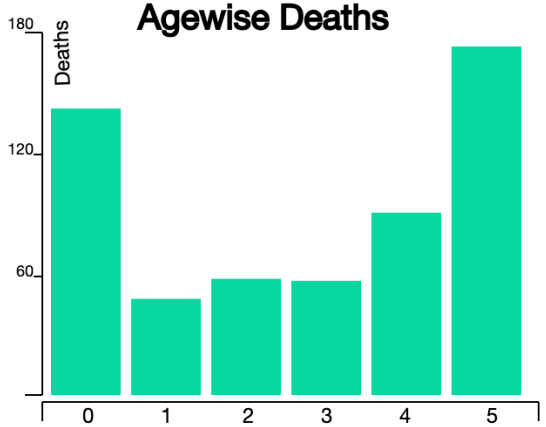
- https://shivaraj9.github.io/portfolio/projects/cholera.html >>
- https://github.com/shukkkur/John-Snows-Ghost-Map
- https://github.com/mdamalie/John-Snow-Cholera-Map-Redesign-1854
  - využití syntetického datasetu

Daywise Deaths

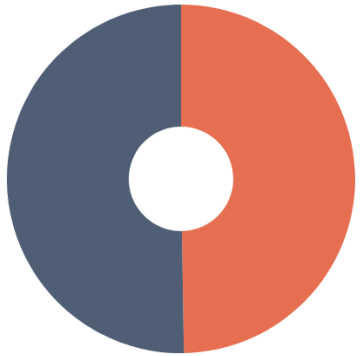


John Snow's Cholera Story Map

Agewise Deaths



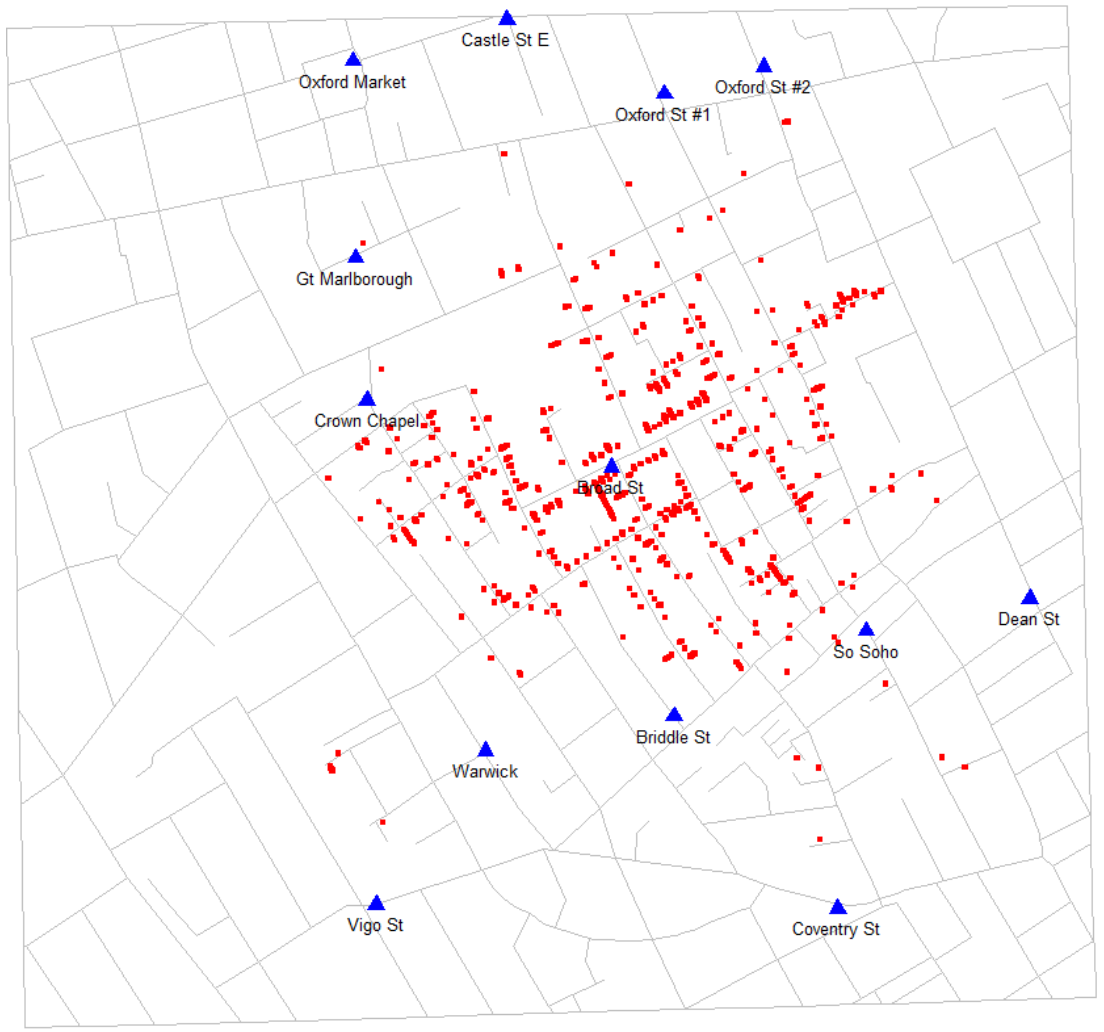
Genderwise Deaths



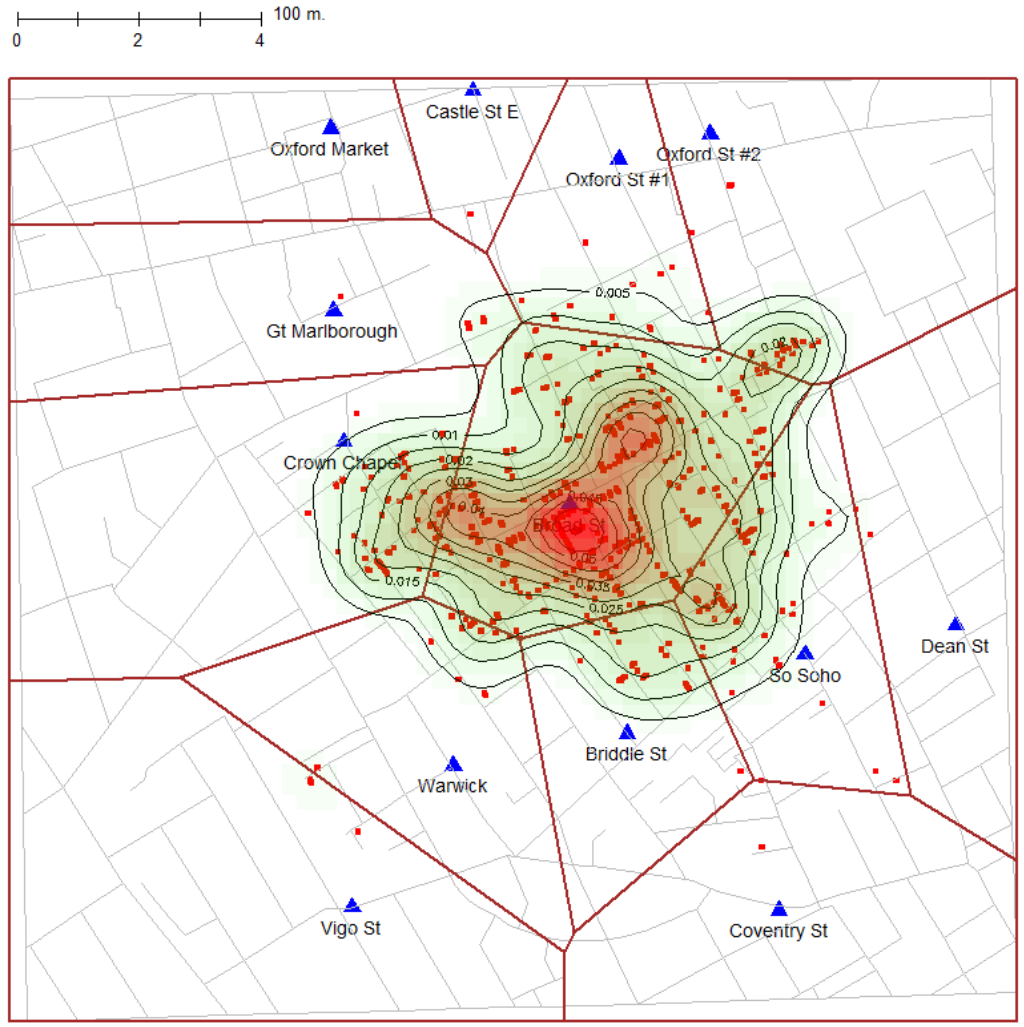
Snowova mapa v R

- https://friendly.github.io/HistData/reference/SnowMap.html >>

Snow's Cholera Map of London (sp)



Snow's Cholera Map of London



Grafy v Pythonu

- různé knihovny
  - soustředíme se na Plotly Express (interaktivní grafy)

  - dále:
  - Plotly (složitější syntaxe než Plotly Express)
  - bokeh (interaktivní, velká data)
  - pandas/matplotlib, seaborn, plotnine (neinteraktivní)

General

Install

```
pip install plotly
```

Import

```
import plotly.express as px
```

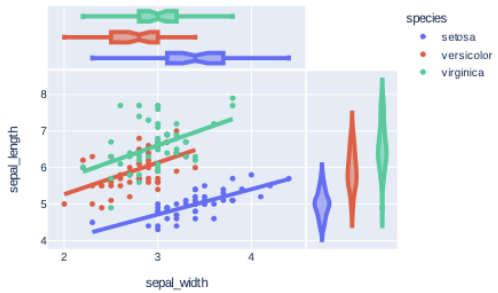
General usage

```
fig = px.chart_type(
    df,
    **chart_specific_parameters,
    title="Chart title",
    labels={"x_column_name": "X column name"},
    width=600,
    height=400,
)
fig.show()
```

Chart types

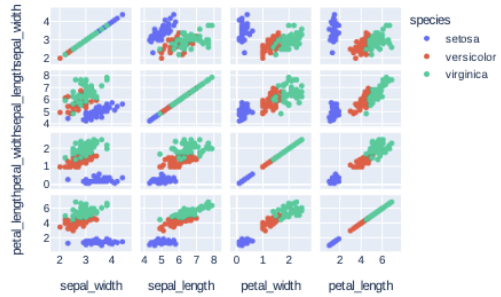
Scatter

```
px.scatter(
    df,
    x="sepal_width",
    y="sepal_length",
    color="species",
    marginal_y="violin",
    marginal_x="box",
    trendline="ols",
    hover_data=["petal_length", "petal_width"],
)
```



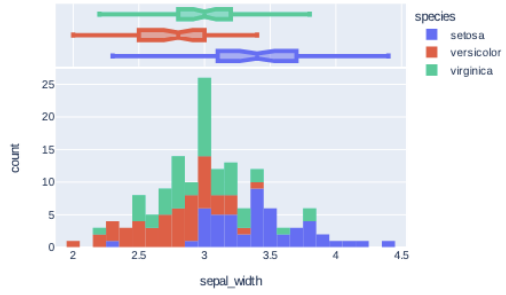
Scatter matrix

```
px.scatter_matrix(
    df,
    dimensions=[
      "sepal_width",
      "sepal_length",
      "petal_width",
      "petal_length",
    ],
    color="species",
)
```



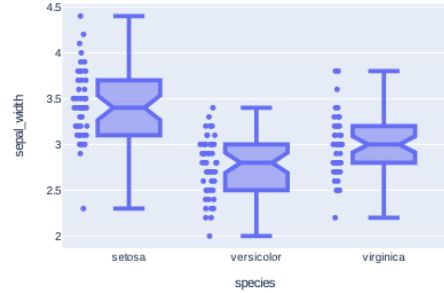
Histogram

```
px.histogram(
    df,
    x="sepal_width",
    color="species",
    marginal="box",  # or "rug", "violin"
    nbins=30,
)
```



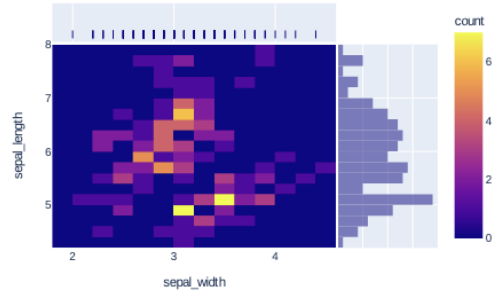
Box

```
px.box(
    df,
    x="species",
    y="sepal_width",
    notched=True,
    points="all"
)
```



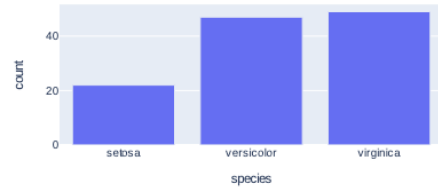
Density heatmap (2D Histogram)

```
px.density_heatmap(
    df,
    x="sepal_width",
    y="sepal_length",
    nbinsx=20,
    nbinsy=20,
    histfunc="count",  # or "sum", "avg"
    marginal_x="rug",
    marginal_y="histogram",
)
```



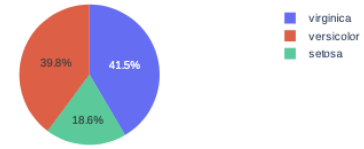
Bar

```
px.bar(count_df, x="species", y="count")
```



Pie

```
px.pie(count_df,
       names="species", values="count")
```



More options

Facet plots

```
px.density_heatmap(
    df,
    x="sepal_width",
    y="sepal_length",
    facet_row="species",
)
```