Data Visualization

A motivational and inspirational talk

Johan Van Horebeek
October 6th, 2021

Goals:

To make a guided walk with some *best practices*
To present some Tools from Statistics/ML
To provide some useful and inspirational references

CIMAT

Centro de Investigación en Matemáticas, A.C.

Data Visualization

A motivational and inspirational talk

Johan Van Horebeek
October 6th, 2021

Goals:

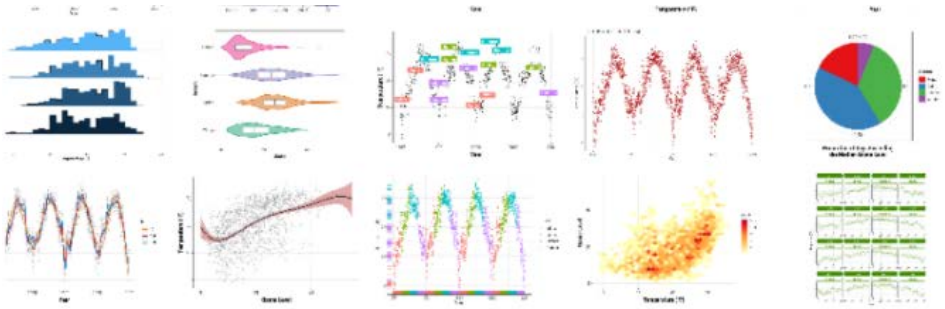
To make a guided walk with some *best practices*
To present some Tools from Statistics/ML
To provide some useful and inspirational references

Not: research talk, technical presentation, to overwhelm

CIMAT
Centro de Investigación en Matemáticas, A.C.

1. Roots and Aims of visualization

Statistical graphics (roots in statistics community)

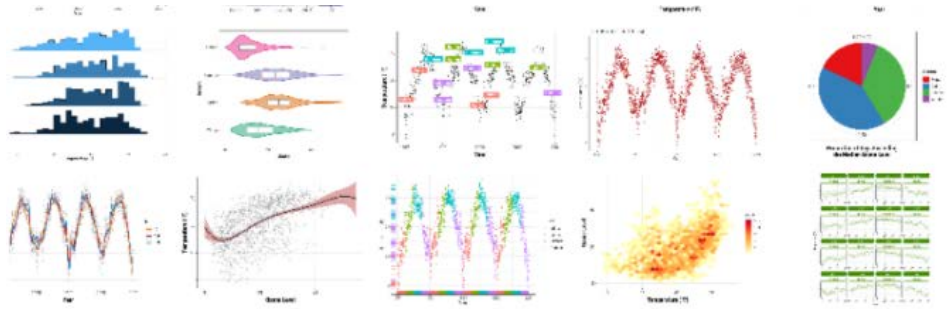


Infographics (roots in graphical design community)

Data Story Telling

1. Roots and Aims of visualization

Statistical graphics (roots in statistics community)



Exploration
Generality
For you (often)

Tool to evaluate models

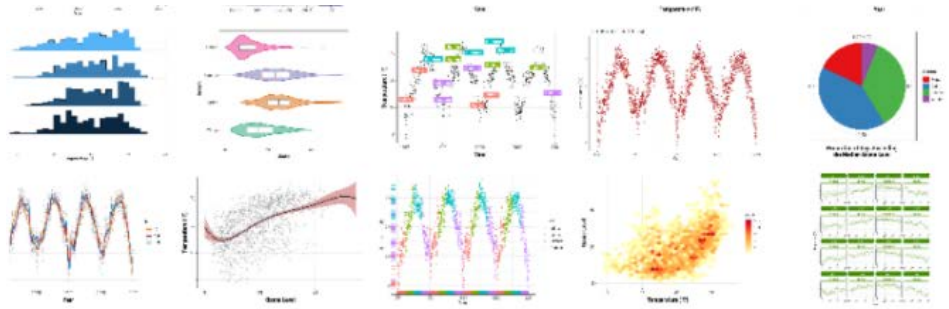
Infographics (roots in graphical design community)

Communication of a message
To engage
Unique
For a particular audience

Data Story Telling

1. Roots and Aims of visualization

Statistical graphics (roots in statistics community)



Exploration
Generality
For you (often)

Tool to evaluate models

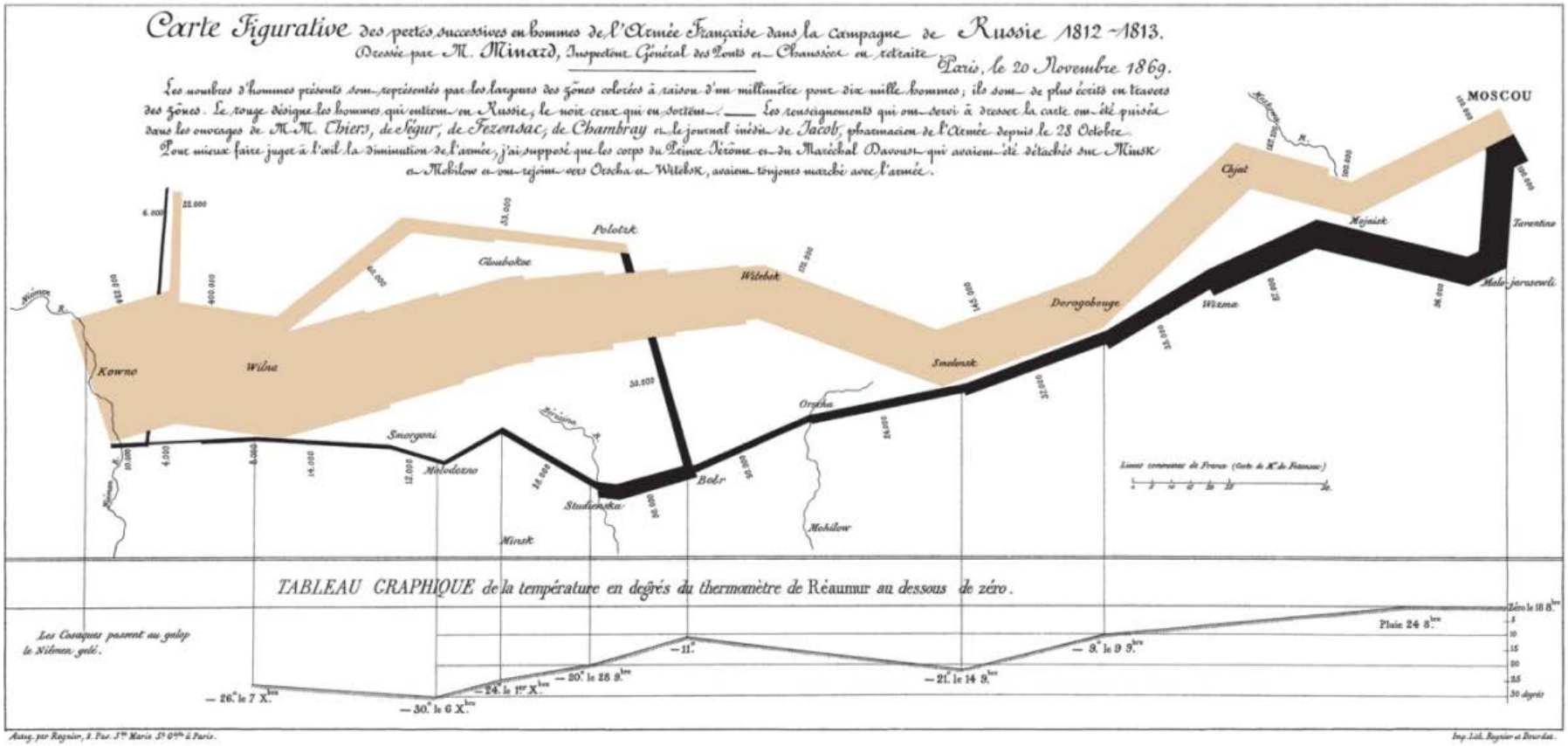
Infographics (roots in graphical design community)

Communication of a message
To engage
Unique
For a particular audience

Data Story Telling

1. Roots and Aims of visualization

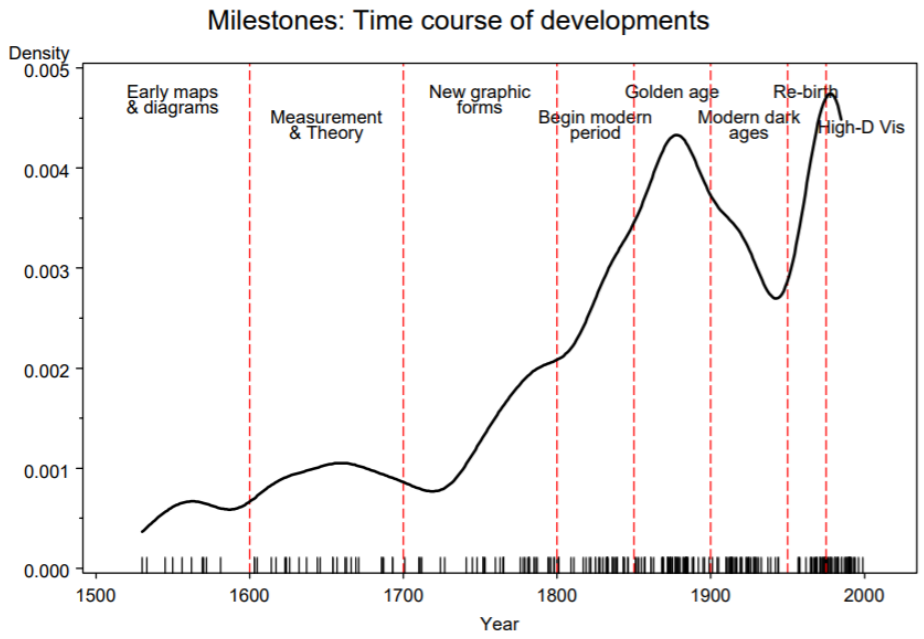
Visualization is from all times



Battles of Napoleon in Russia

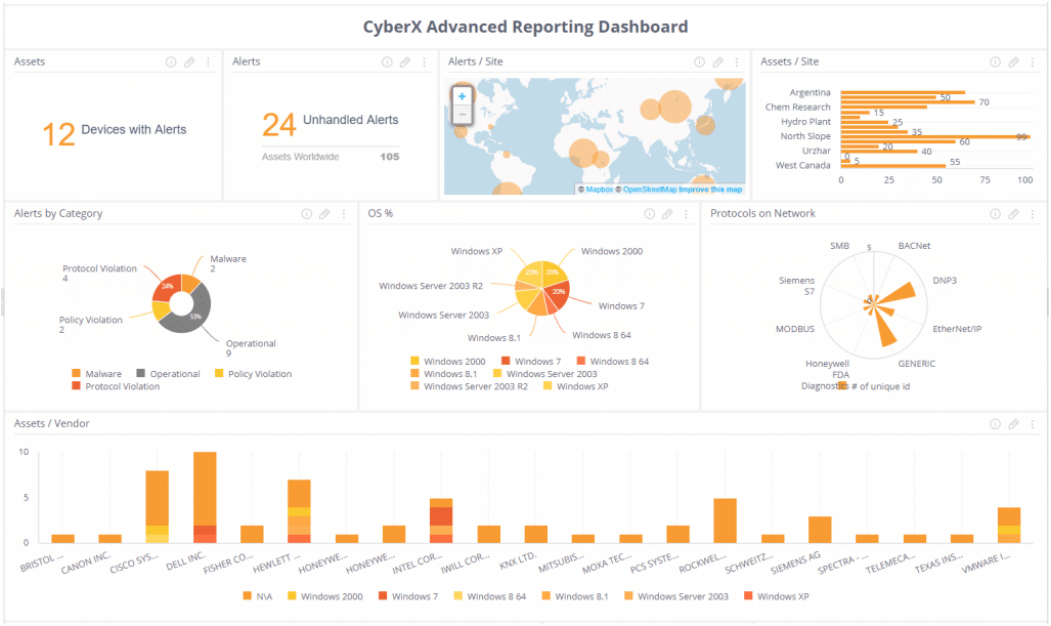
1. Roots and Aims of visualization

Visualization is from all times

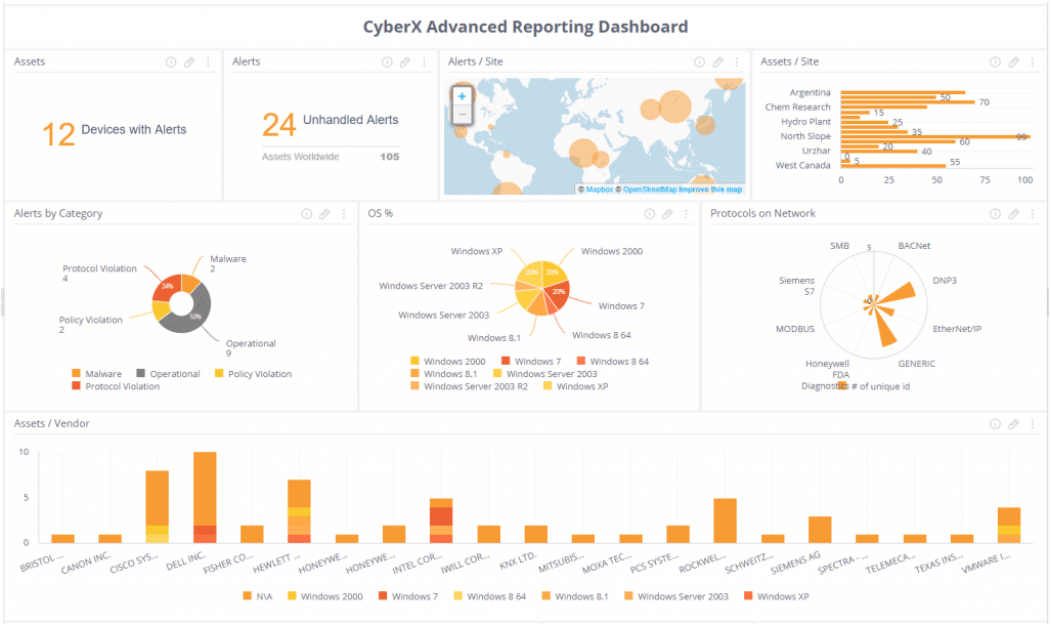


Battles of Napoleon in Russia

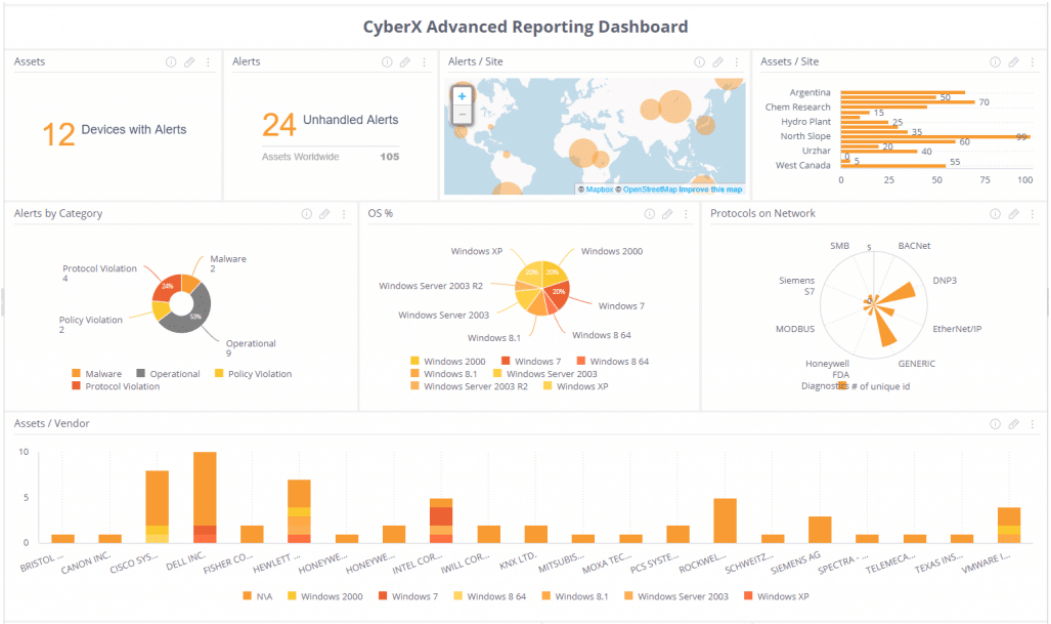
2. Visualization: not just engineering



2. Visualization: not just engineering



2. Visualization: not just engineering



Visual perception

Effective communication

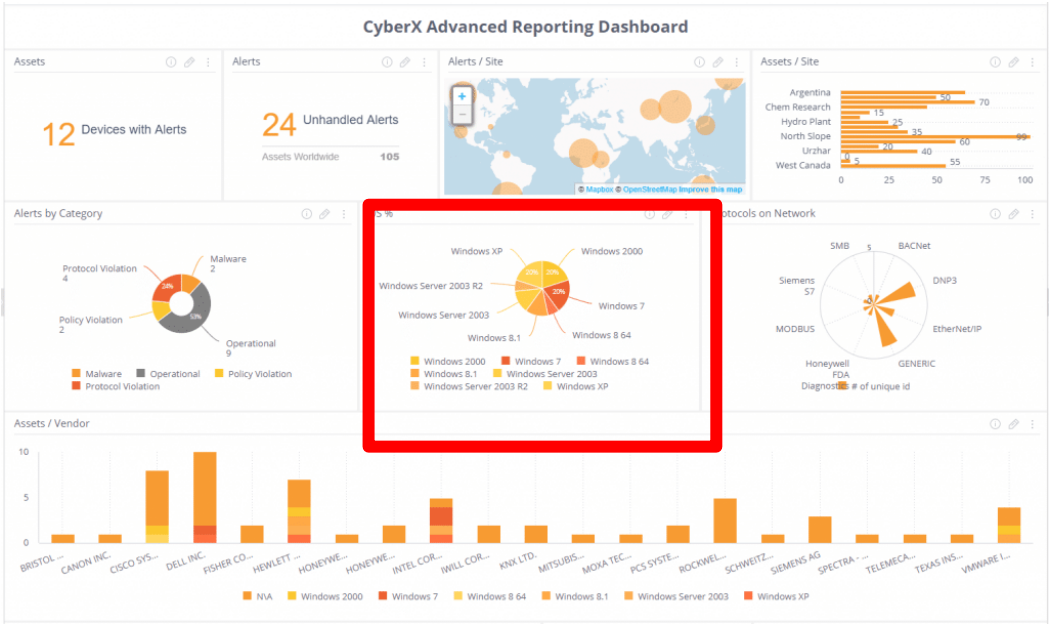
Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

2. Visualization: not just engineering



Visual perception

Effective communication

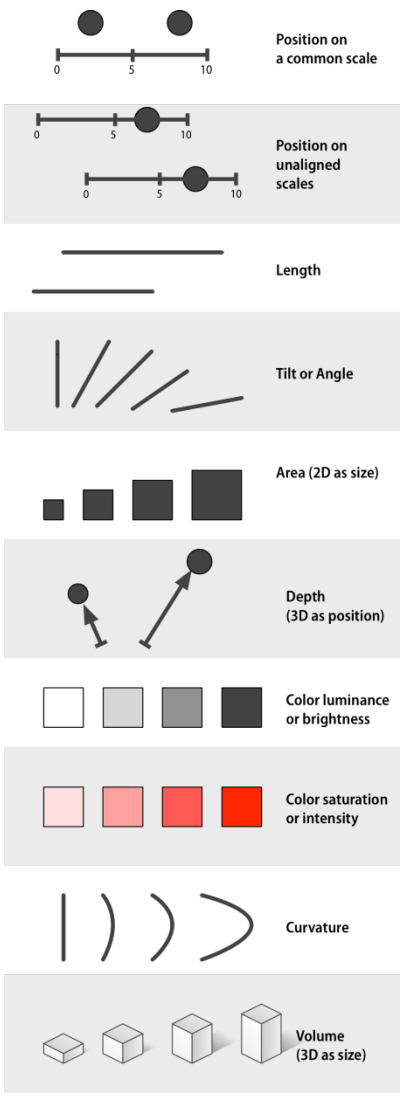
Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

2. Visualization: not just engineering



https://socviz.co

Design no Decoration

Visual perception

Effective communication

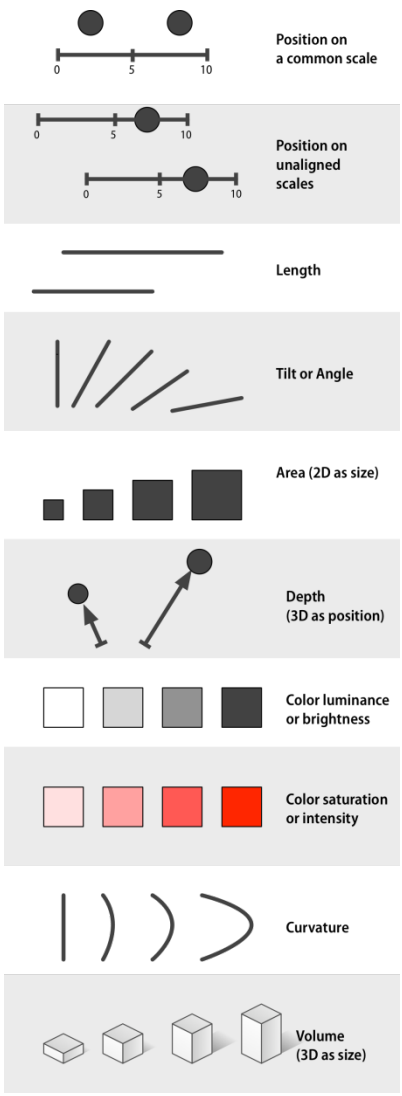
Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

Antoine de Saint-Exupery

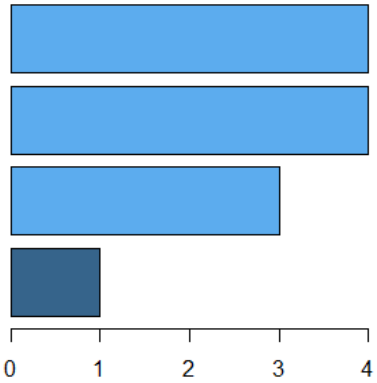
Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

2. Visualization: not just engineering



Design no Decoration



Visual perception

Effective communication

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

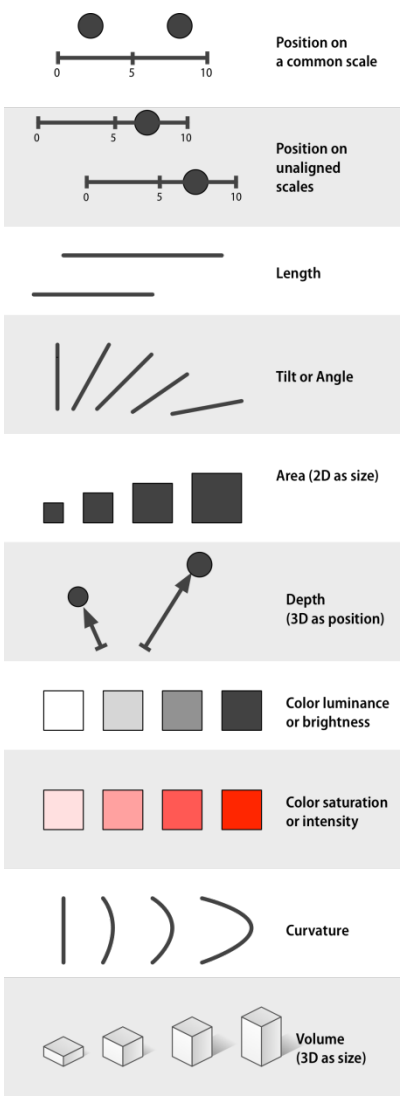
Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

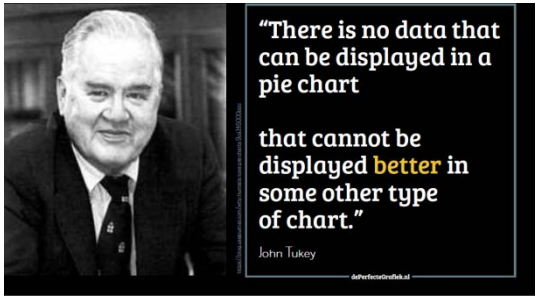
https://socviz.co

2. Visualization: not just engineering



https://socviz.co

Design no Decoration



De perfectegrafiek

Visual perception

Effective communication

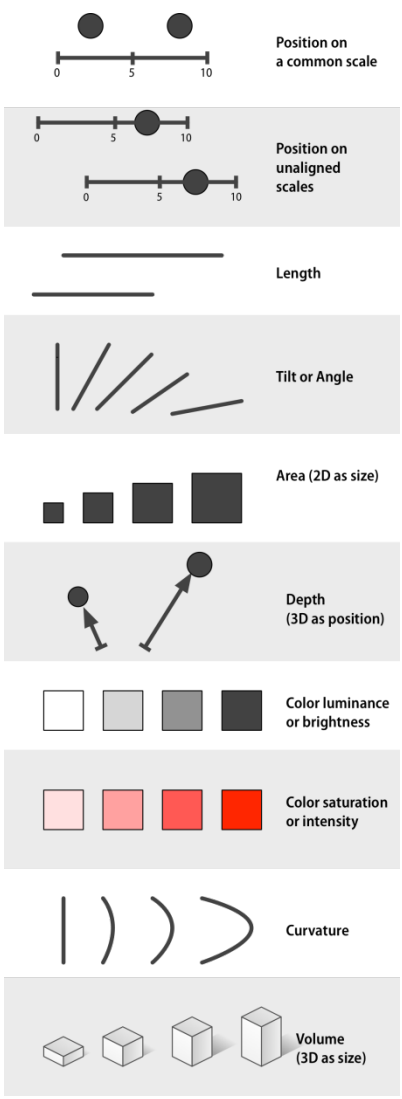
Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

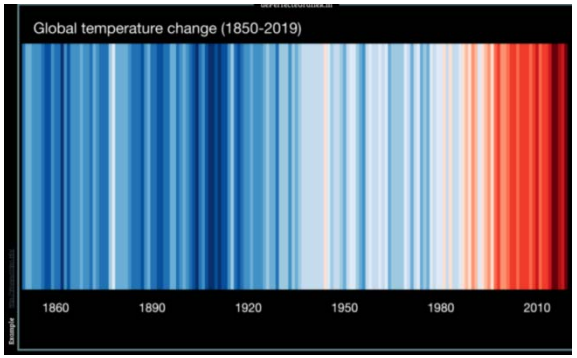
Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

2. Visualization: not just engineering





Design no Decoration

Visual perception

Effective communication

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

https://socviz.co

2. Visualization: not just engineering

Visual perception

Effective communication

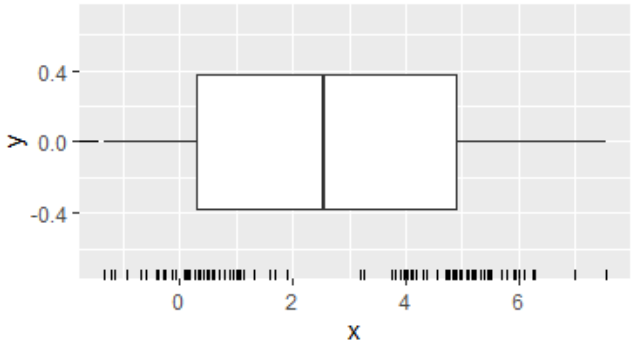
Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

2. Visualization: not just engineering



Visual perception

Effective communication

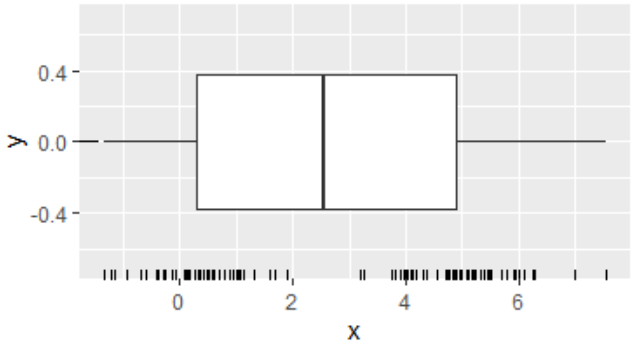
Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

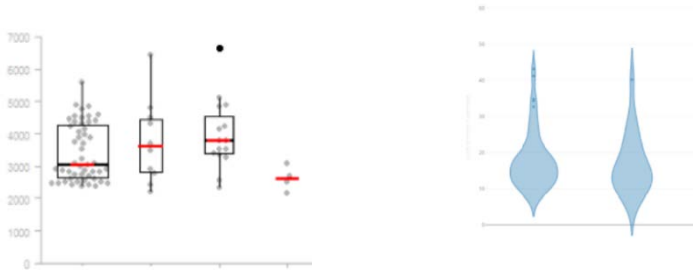
Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

2. Visualization: not just engineering





Do not summarize too fast

Visual perception

Effective communication

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away

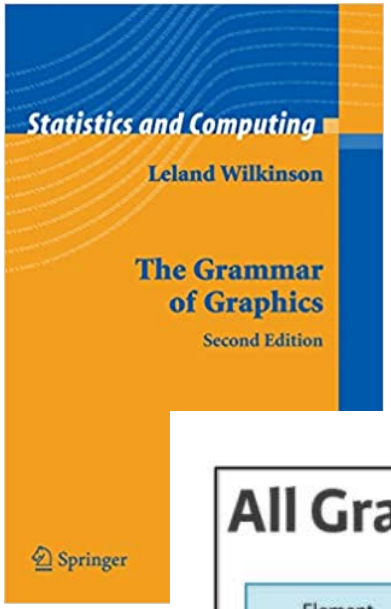
Antoine de Saint-Exupery

Truthful communication

How to avoid artefacts in the visualization that are no intrinsic features?

3. Visualization: a guided walk

3. Visualization: a guided walk



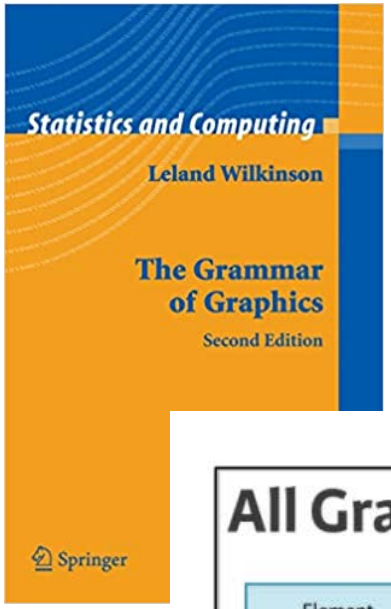
R: ggplot2 (Hadley Wickham)
Python: plotnine

All Grammatical Elements

Element	Description
Data	The dataset being plotted.
Aesthetics	The scales onto which we *map* our data.
Geometries	The visual elements used for our data.
Facets	Plotting small multiples.
Statistics	Representations of our data to aid understanding.
Coordinates	The space on which the data will be plotted.
Themes	All non-data ink.

ggplot2 Layers - Themes

3. Visualization: a guided walk

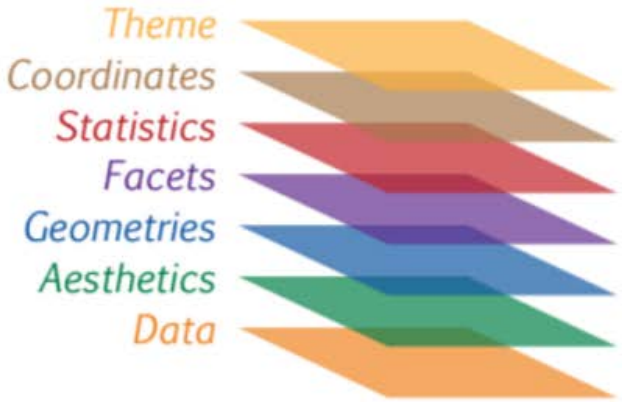


R: ggplot2 (Hadley Wickham)
Python: plotnine

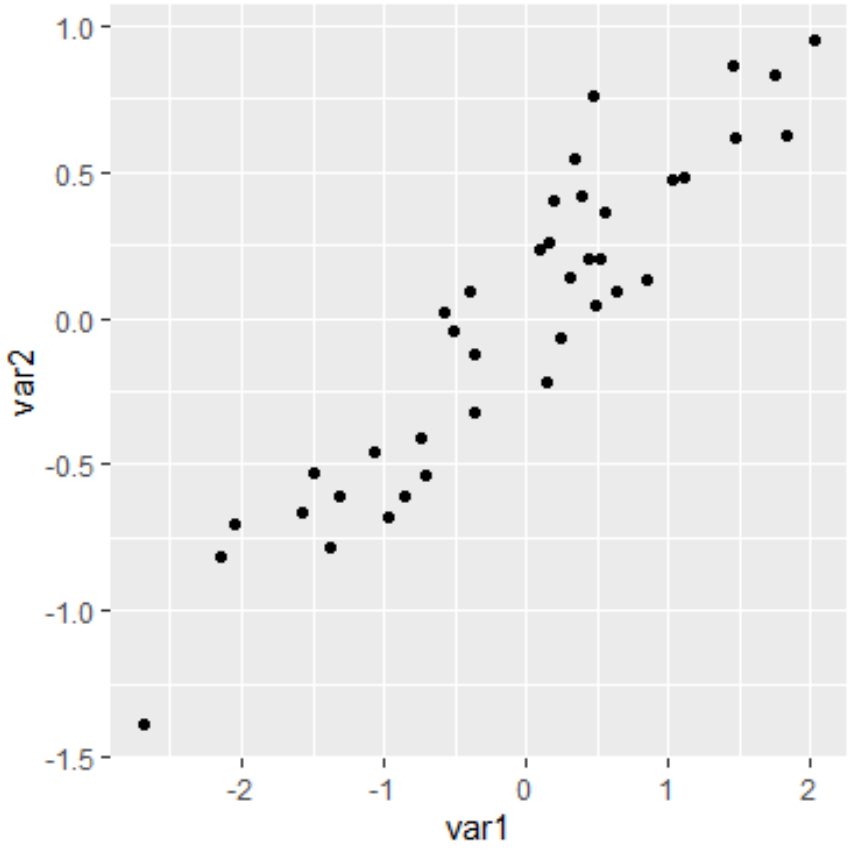
All Grammatical Elements

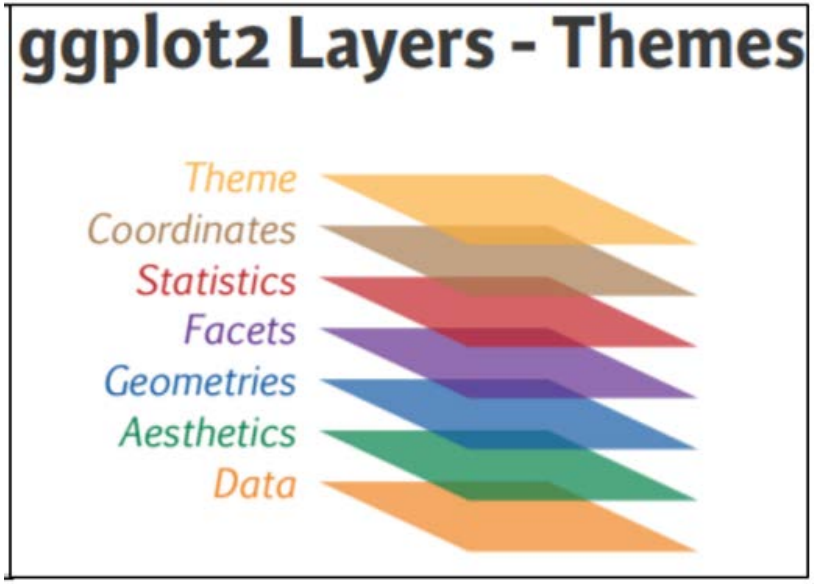
Element	Description
Data	The dataset being plotted.
Aesthetics	The scales onto which we *map* our data.
Geometries	The visual elements used for our data.
Facets	Plotting small multiples.
Statistics	Representations of our data to aid understanding.
Coordinates	The space on which the data will be plotted.
Themes	All non-data ink.

ggplot2 Layers - Themes



3. Visualization: a guided walk

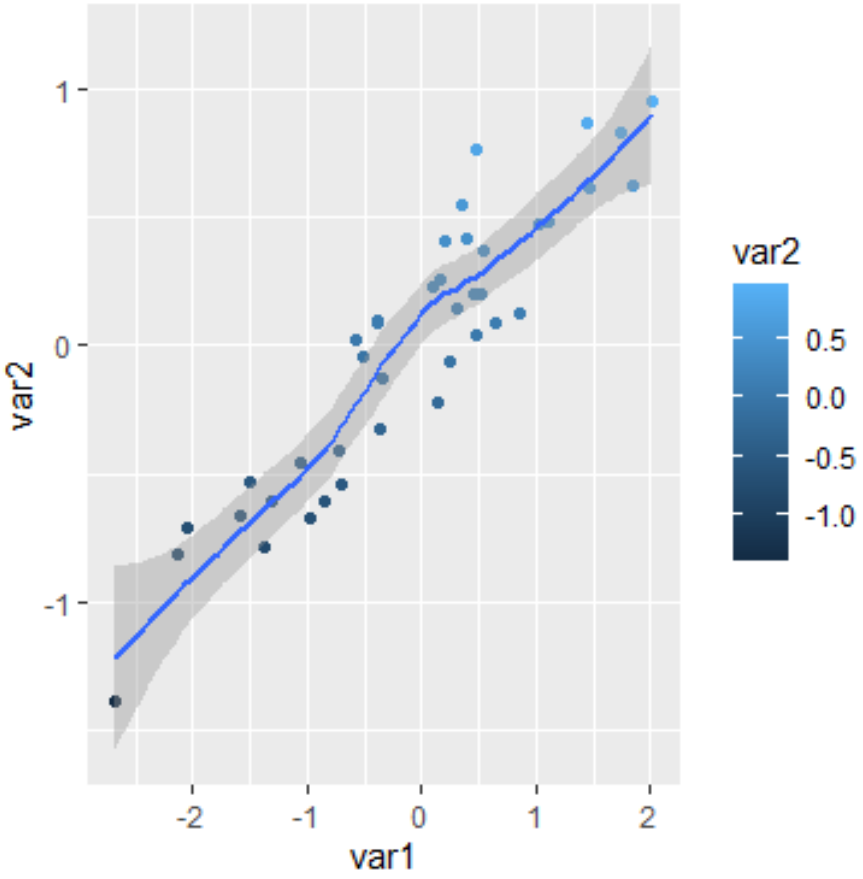


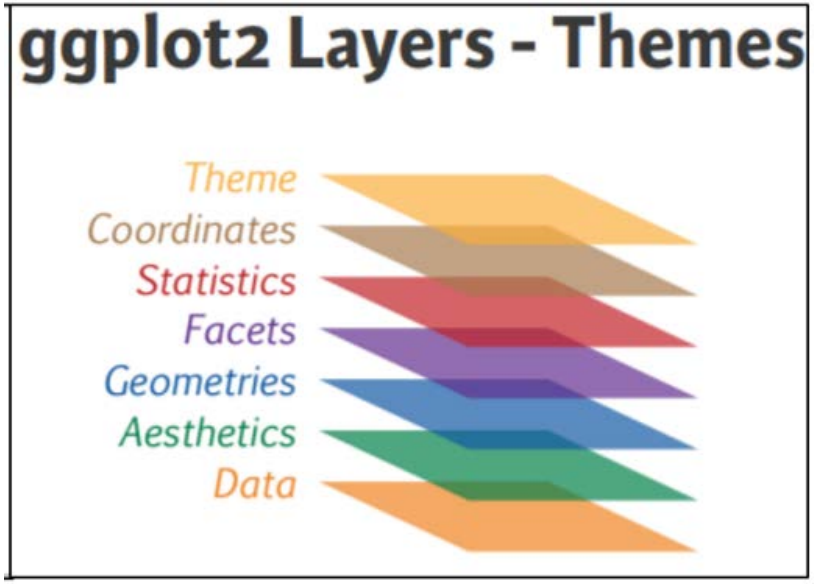



```
d<-data.frame("var1"=x, "var2"=y)
```

```
ggplot(data=d, aes(x=var1, y=var2)) + geom_point()
```

3. Visualization: a guided walk

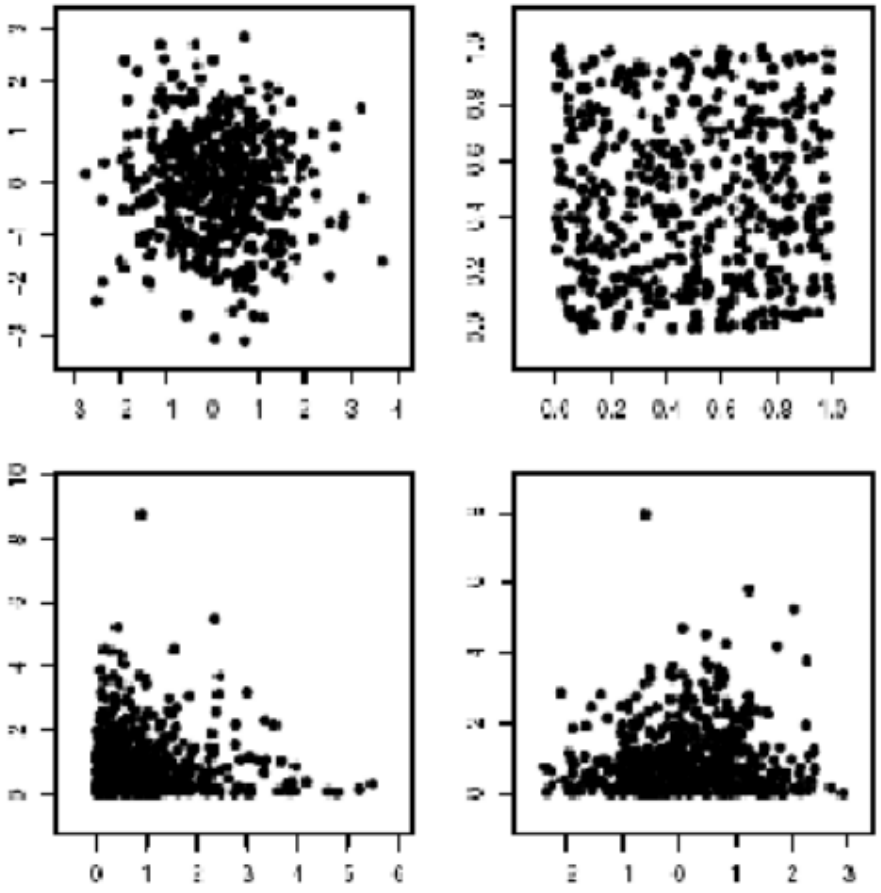





```
d<-data.frame("var1"=x, "var2"=y)
```

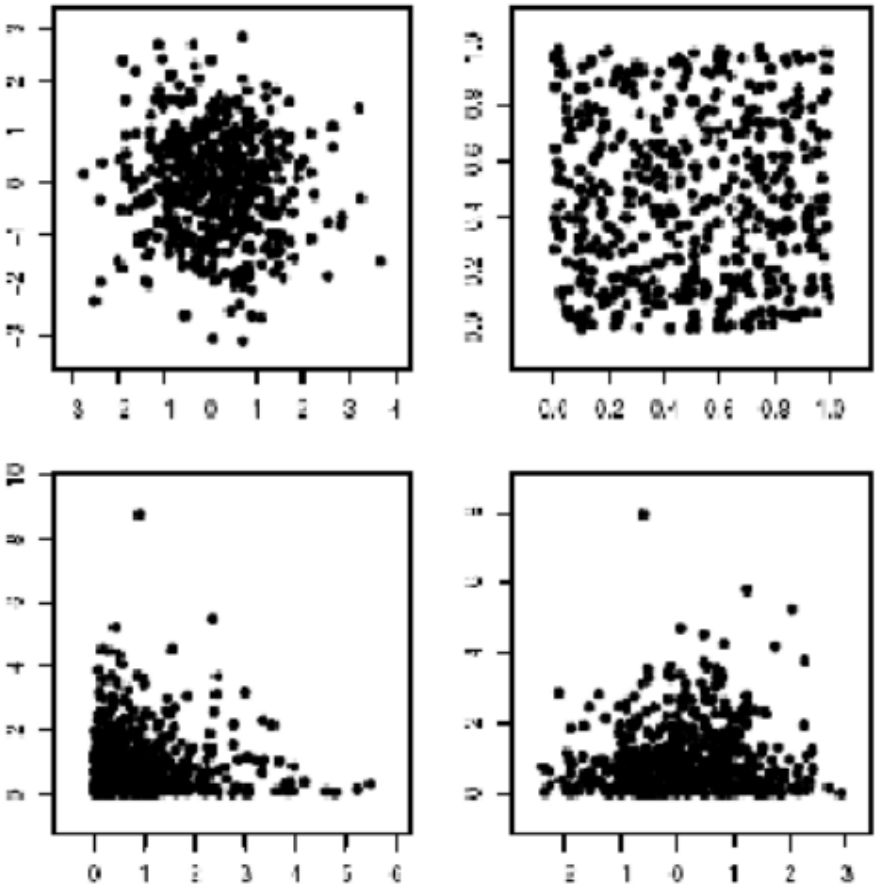
```
ggplot(data=d,aes(x=var1, y=var2, colour=var2)) + geom_point() + geom_smooth()
```

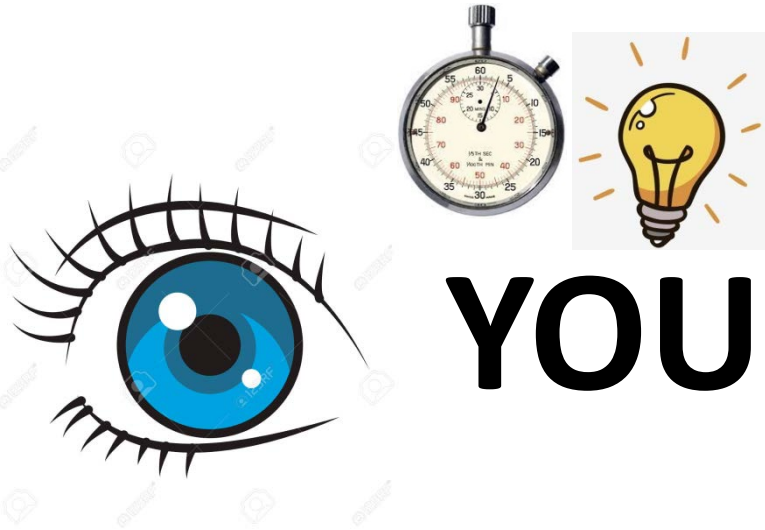
3. Visualization: a guided walk



In which case there is a stronger dependency?

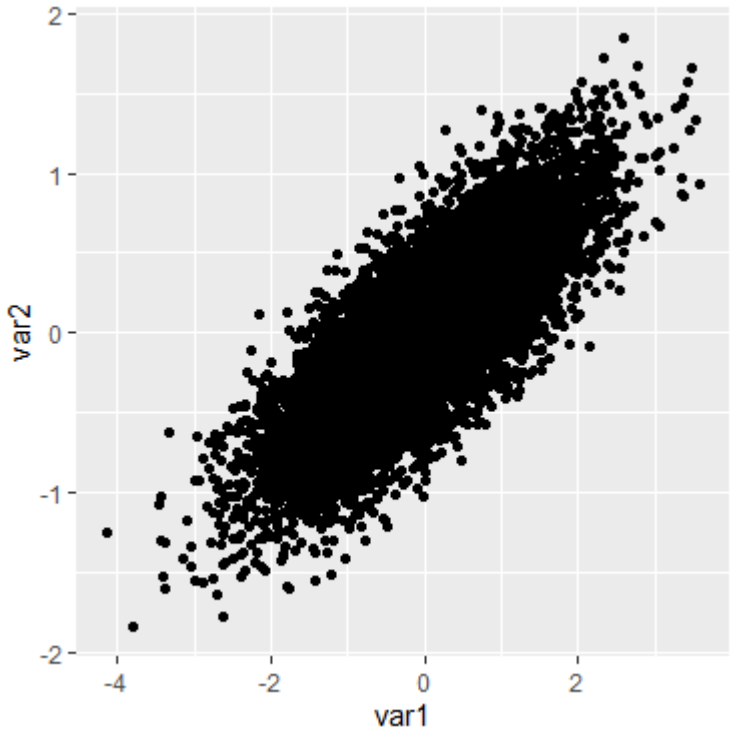
3. Visualization: a guided walk



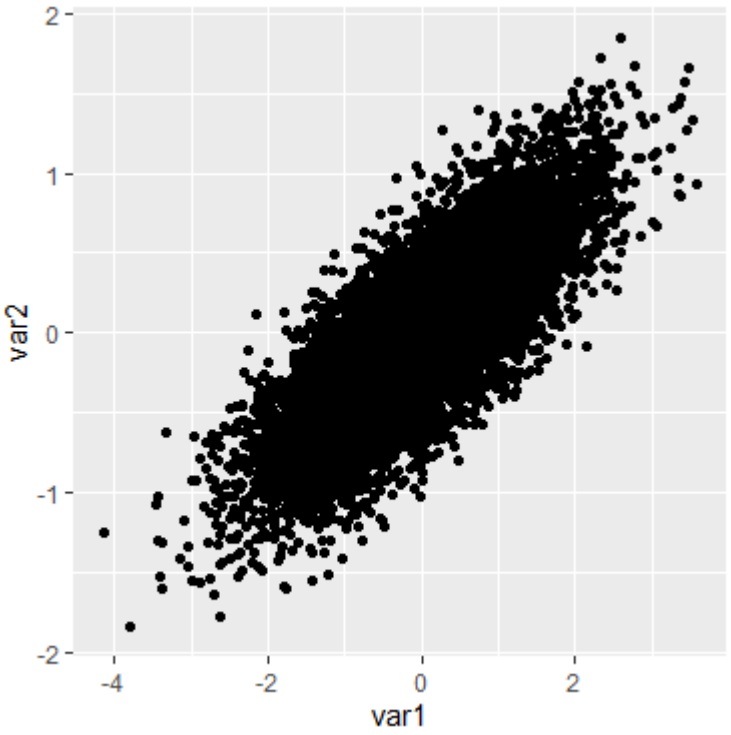


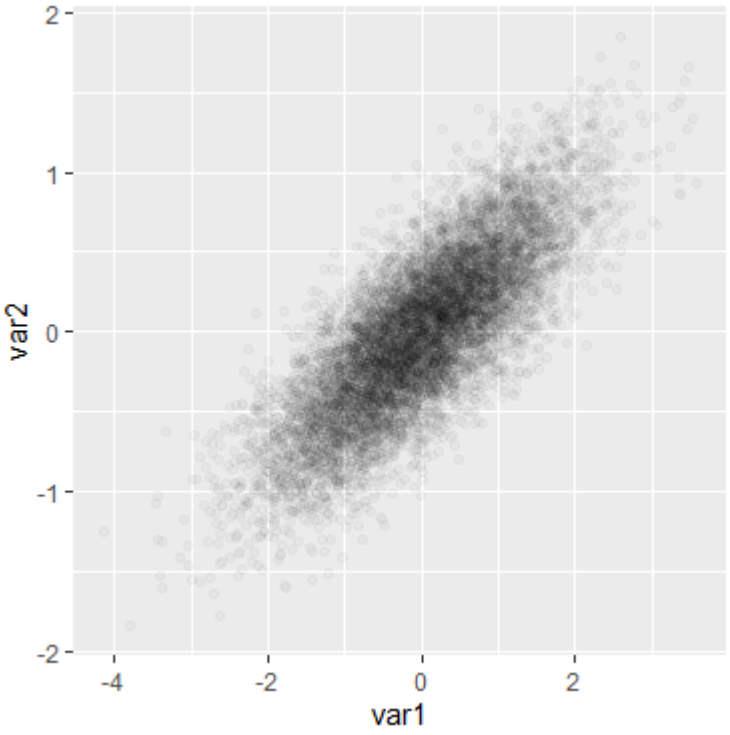
In which case there is a stronger dependency?

3. Visualization: a guided walk

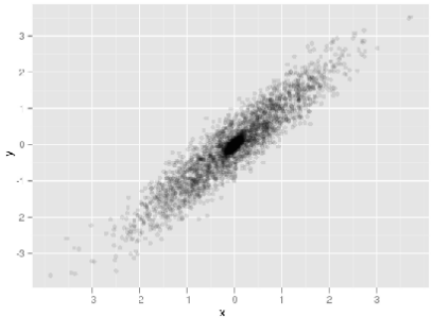


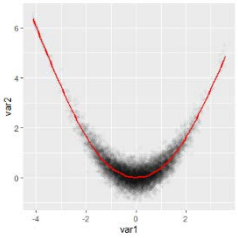
3. Visualization: a guided walk





ggplot(data=d, aes(x=var1,y=var2)) +
geom_point(alpha=0.025)



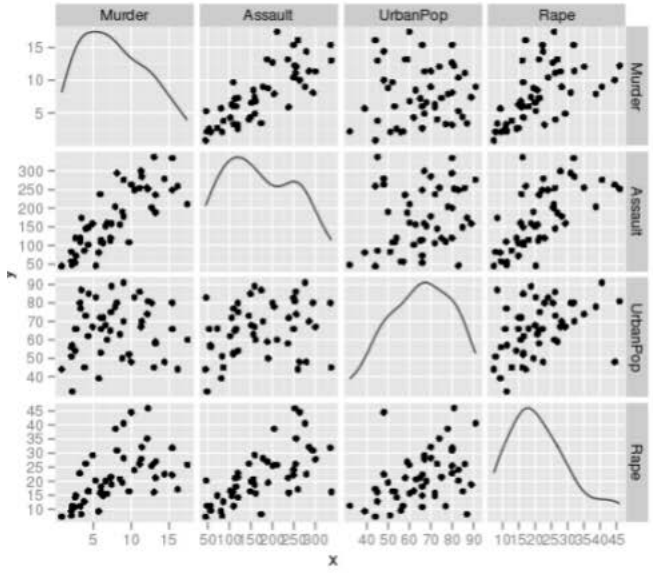


3. Visualization: a guided walk

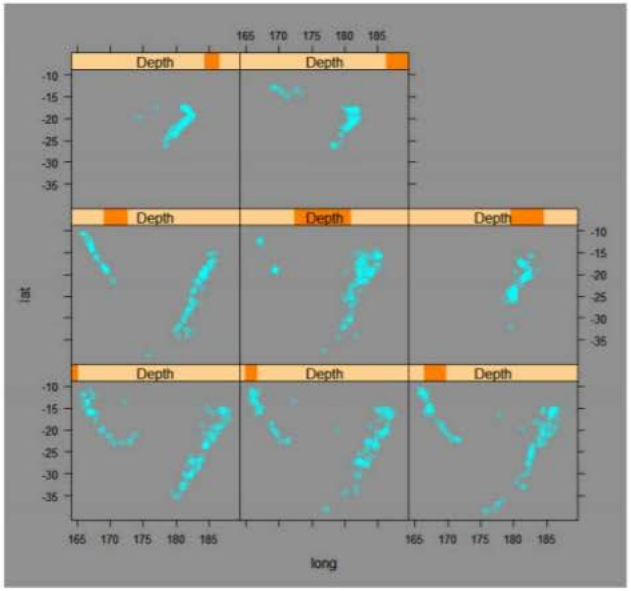
Up to higher dimensions:

abstraction/marginalize

specification/conditionalize



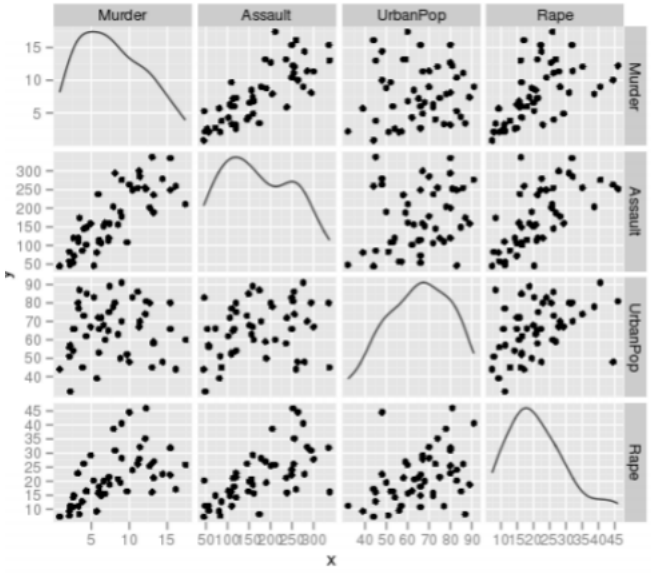
Pairsplots



Trellisplots

3. Visualization: a guided walk

Up to higher dimensions:



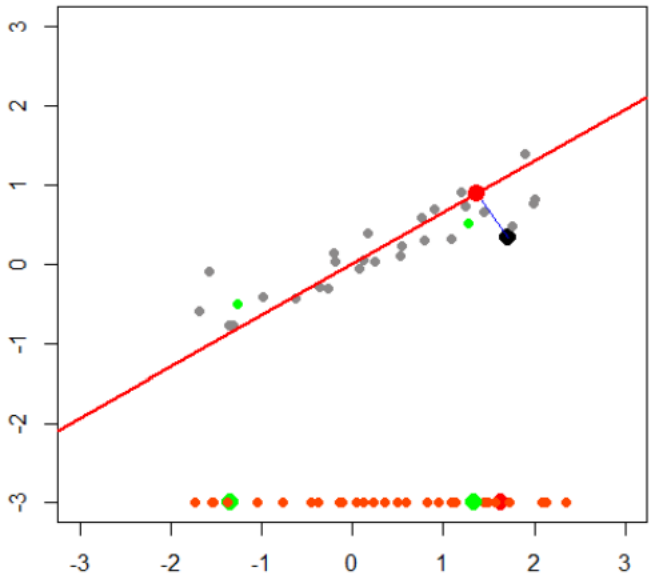
Can we do better?

Dimension reduction techniques:

principal component analysis

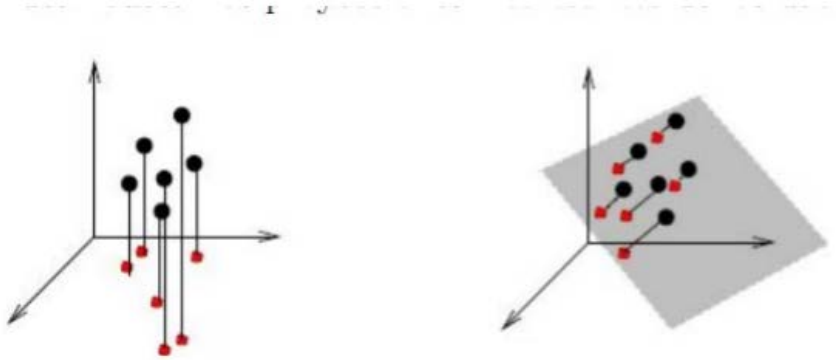
3. Visualization: a guided walk

Up to higher dimensions:



Informative directions (max variance)
Least approximation error
Distance preserving transformation

Can we do better?

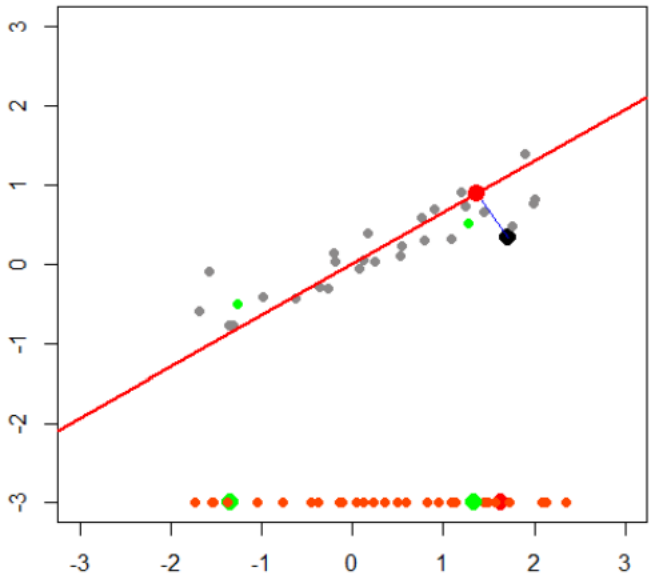


Dimension reduction techniques:

principal component analysis

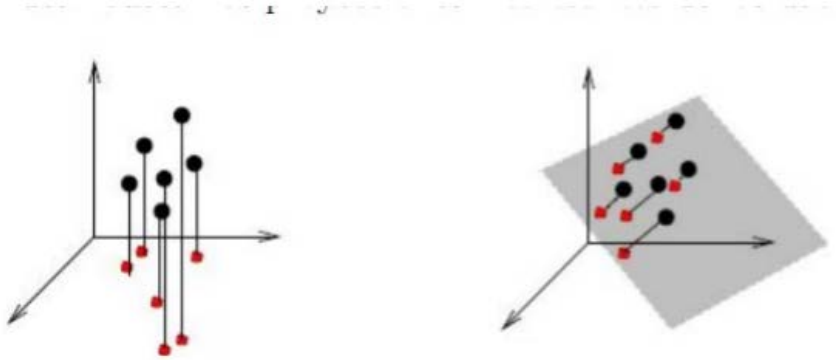
3. Visualization: a guided walk

Up to higher dimensions:



Informative directions (max variance)
Least approximation error
Distance preserving transformation

Can we do better?



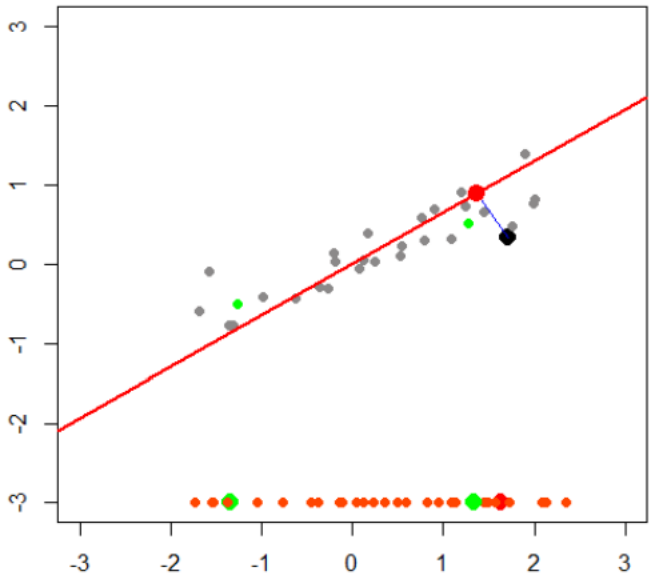
Dimension reduction techniques:

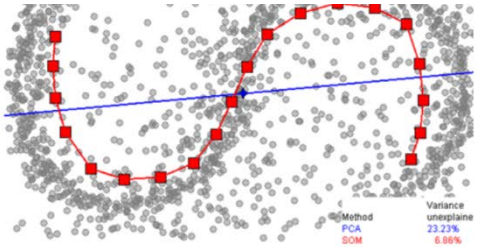
principal component analysis

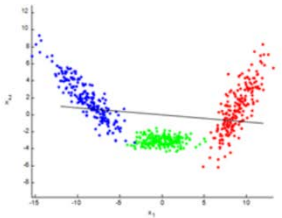
take care : suitable? Optimal?

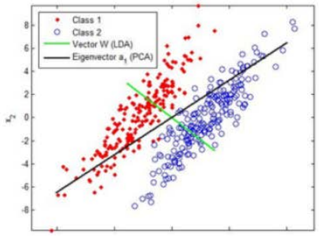
3. Visualization: a guided walk

Up to higher dimensions:









Informative directions (max variance)
Least approximation error
Distance preserving transformation

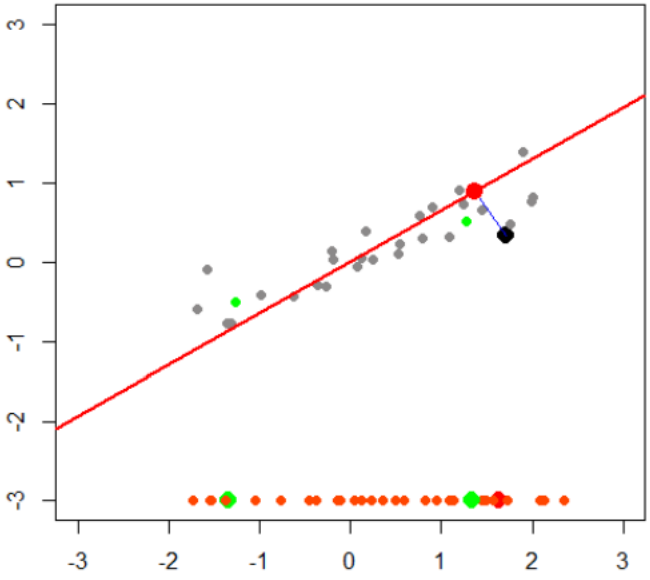
Dimension reduction techniques:

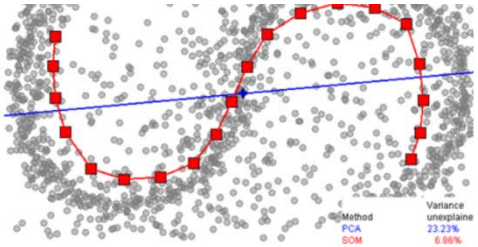
principal component analysis

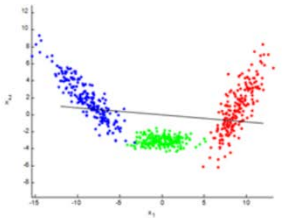
take care : suitable? Optimal?

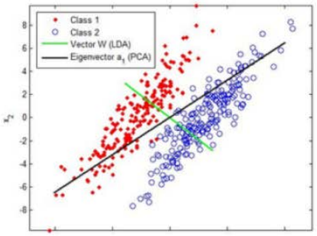
3. Visualization: a guided walk

Up to higher dimensions:









Dimension reduction techniques:

principal component analysis

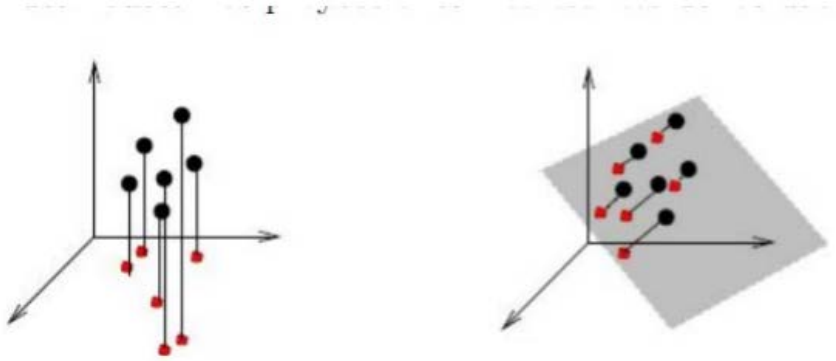
take care : suitable? Optimal?

Informative directions (max variance) (**ICA, PP**)
Least approximation error
Distance preserving transformation
(CDS, KernelPCA, T-SNE, ISOMAP, SOM, …)

3. Visualization: a guided walk

Up to higher dimensions:

Can we do better?



Dimension reduction techniques:

think a lot to decide how to proceed

3. Visualization: a guided walk

Up to higher dimensions:

Can we do better?

Dimension reduction techniques:

Make a **random** walk ⟷ think a lot to decide how to proceed

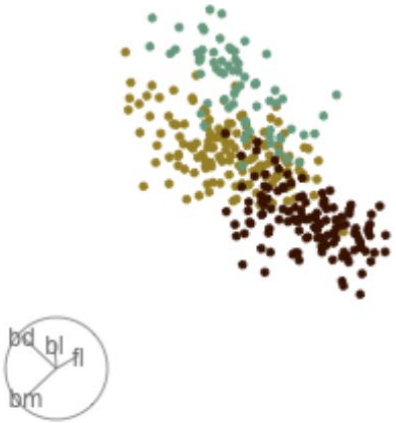
3. Visualization: a guided walk

Up to higher dimensions:

Example (Diana Cook)

Make a **random** walk

challenging rendering problem
nice mathematical results



https://www.dicook.org/files/visec2020/slides_tourr#1

3. Visualization: a guided walk

Up to higher dimensions:

Next step: combine

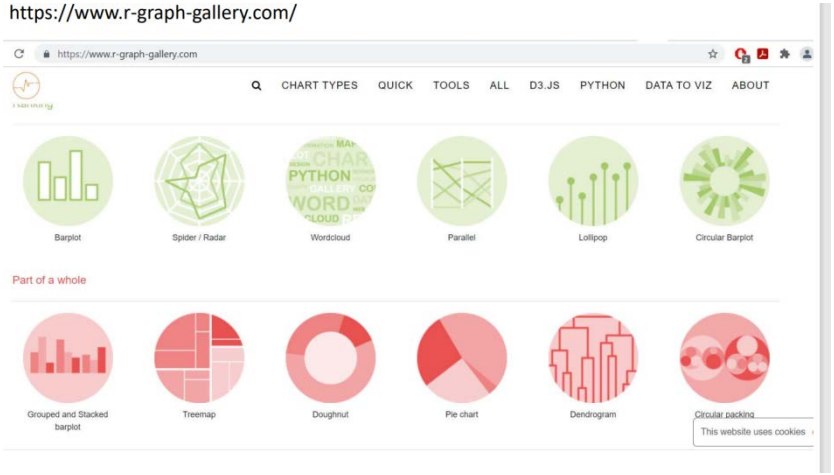
ggobi

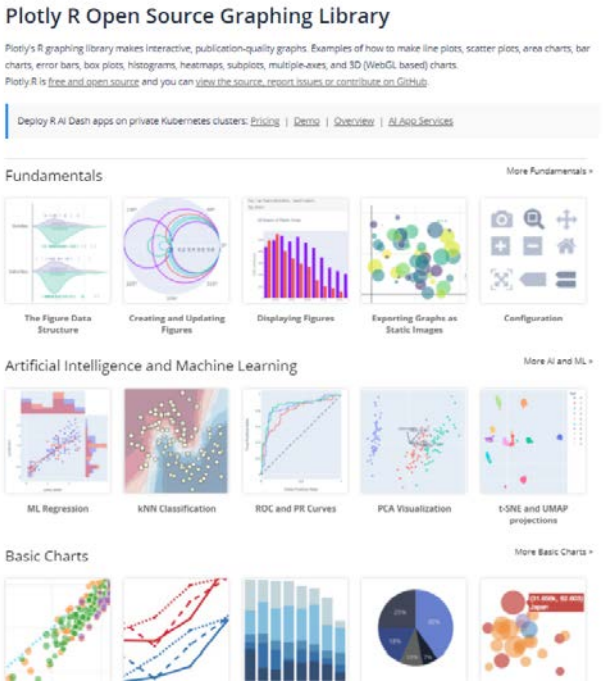
4. Visualization: in practice

Increasing popularity of **Galleries**

Python/R

https://www.r-graph-gallery.com/



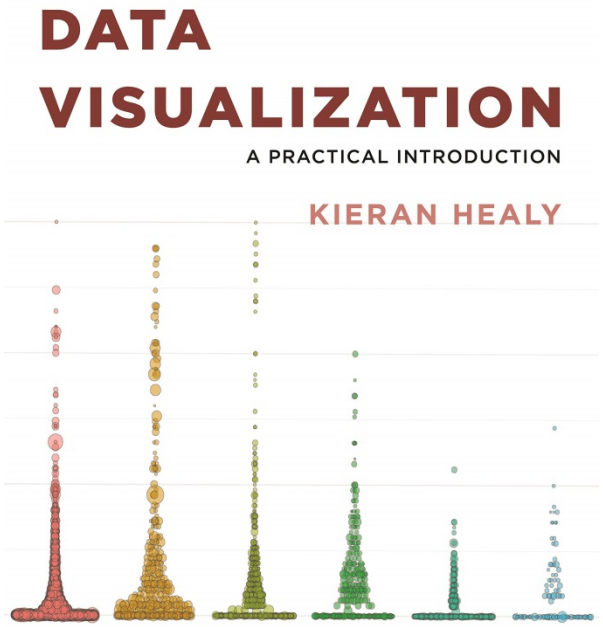


Dashboards: Dash (Python/R),
Shiny (R)

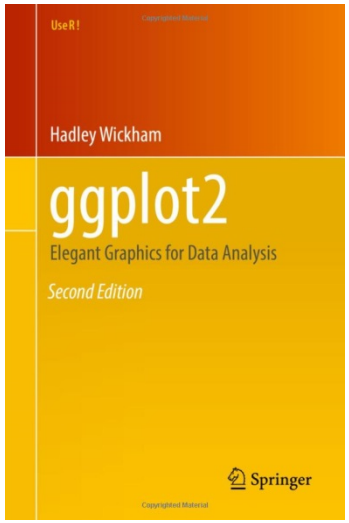
For working with **prototypes**:

https://flourish.studio/

4. Visualization: in practice



https://socviz.co/



https://ggplot2-book.org/

Good practices (Michiel Dullaert)

https://chart.guide/