

DSSP

Table Wrangling & Data Visualization

A. Bichat - E. Le Pennec



2019 - 2020

Outline

Introduction



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Table Wrangling and Visualization

Introduction



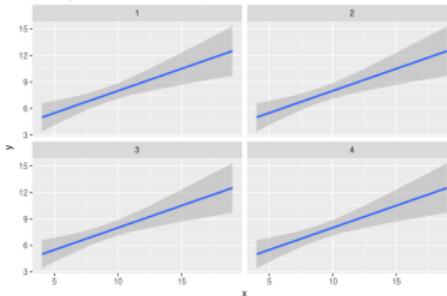
Anscombe Quartet

| | x1 | y1 | x2 | y2 | x3 | y3 | x4 | y4 |
|----|----|-------|----|------|----|-------|----|------|
| 1 | 10 | 8.04 | 10 | 9.14 | 10 | 7.46 | 8 | 6.58 |
| 2 | 8 | 6.95 | 8 | 8.14 | 8 | 6.77 | 8 | 5.76 |
| 3 | 13 | 7.58 | 13 | 8.74 | 13 | 12.74 | 8 | 7.71 |
| 4 | 9 | 8.81 | 9 | 8.77 | 9 | 7.11 | 8 | 8.84 |
| 5 | 11 | 8.33 | 11 | 9.28 | 11 | 7.81 | 8 | 8.47 |
| 6 | 14 | 9.96 | 14 | 8.1 | 14 | 8.64 | 8 | 7.04 |
| 7 | 6 | 7.24 | 6 | 6.13 | 6 | 6.08 | 8 | 5.25 |
| 8 | 4 | 4.26 | 4 | 3.1 | 4 | 5.39 | 19 | 12.5 |
| 9 | 12 | 10.84 | 12 | 9.13 | 12 | 8.15 | 8 | 5.56 |
| 10 | 7 | 4.82 | 7 | 7.26 | 7 | 6.42 | 8 | 7.91 |
| 11 | 5 | 5.68 | 5 | 4.74 | 5 | 5.73 | 8 | 6.89 |

From

to

Anscombe Quartet
Linear regression with confidence bar and points



From Table to Graph

- Need to manipulate tables.
- Need to visualize tables.

Outline

Table Wrangling



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Tidy Table

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

Tidy Table

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

- Each observation forms a row.

Tidy Table

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

- Each observation forms a row.
- Each variable forms a column.

Tidy Table

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

- Each observation forms a row.
- Each variable forms a column.

- Columns are named, rows are not.
- Columns made of values of the same type.

Tidy Table

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

- Each observation forms a row.
- Each variable forms a column.
- Each type of observational unit forms a table.

- Columns are named, rows are not.
- Columns made of values of the same type.
- Codd's 3rd normal form...

Tidy Table

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

- Each observation forms a row.
- Each variable forms a column.
- Each type of observational unit forms a table.

- Columns are named, rows are not.
- Columns made of values of the same type.
- Codd's 3rd normal form...

In practice

- Definition of observation may depend on the task.
- Tidying data is a real work!

Table Actions

Table Wrangling



| | a | b | c | d |
|---|---|---|---|---|
| 5 | | | | A |
| 1 | | | | B |
| 4 | | | | A |
| 5 | | | | B |
| 2 | | | | B |

Verbs

Table Actions

Table Wrangling



| | a | b | c | d |
|---|---|---|---|---|
| 5 | | | | A |
| 1 | | | | B |
| 4 | | | | A |
| 5 | | | | B |
| 2 | | | | B |

Filter

Verbs

- Rows: Filter,

Table Actions

Table Wrangling



| | a | b | c | d |
|---|---|---|---|---|
| 5 | | | | A |
| 1 | | | | B |
| 4 | | | | A |
| 5 | | | | B |
| 2 | | | | B |

| | a | d |
|---|---|---|
| 5 | | A |
| 1 | | B |
| 4 | | A |
| 5 | | B |
| 2 | | B |

Select

Verbs

- Rows: Filter,
- Columns: Remove,

Table Actions

Table Wrangling



| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |
| 3 | | | A |

Insert

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

Verbs

- Rows: Filter, Insert
- Columns: Remove,

Table Actions

Table Wrangling



| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d | e |
|---|---|---|---|----|
| 5 | | | A | 10 |
| 1 | | | B | 2 |
| 4 | | | A | 8 |
| 5 | | | B | 10 |
| 2 | | | B | 4 |

Add

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add

Table Actions

Table Wrangling



| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d | e |
|---|---|---|---|----|
| 5 | | | A | 10 |
| 1 | | | B | 2 |
| 4 | | | A | 8 |
| 5 | | | B | 10 |
| 2 | | | B | 4 |

Transform

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, ,

a'
17
Summarize

| | a | b | c | d |
|---|---|---|---|---|
| 5 | | | | A |
| 1 | | | | B |
| 4 | | | | A |
| 5 | | | | B |
| 2 | | | | B |

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize,

Table Actions

Table Wrangling



| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d |
|---|---|---|---|
| 1 | | | B |
| 2 | | | B |
| 4 | | | A |
| 5 | | | A |
| 5 | | | B |

Sort

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort

Table Actions

Table Wrangling



| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |
| 1 | | | B |
| 5 | | | B |
| 2 | | | B |

Split

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort
- Split

Table Actions

Table Wrangling



| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |
| 1 | | | B |
| 5 | | | B |
| 2 | | | B |

Split

| d | a' |
|---|----|
| A | 9 |
| B | 8 |

| d | a' |
|---|----|
| A | 9 |
| B | 8 |

Split/Apply/Combine

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort
- Split/Apply/Combine (Summarize)

Table Actions

Table Wrangling



a'
17
Summarize

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |
| 3 | | | A |

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |
| 3 | | | A |

Insert

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

Filter

| a | b | c | d | e |
|---|---|---|---|----|
| 5 | | | A | 10 |
| 1 | | | B | 2 |
| 4 | | | A | 8 |
| 5 | | | B | 10 |
| 2 | | | B | 4 |

Add/Transform

| a | d |
|---|---|
| 5 | A |
| 1 | B |
| 4 | A |
| 5 | B |
| 2 | B |

Select

| a | b | c | d |
|---|---|---|---|
| 1 | | | B |
| 2 | | | B |
| 4 | | | A |
| 5 | | | A |
| 5 | | | B |
| 2 | | | B |

Sort

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |
| 1 | | | B |
| 5 | | | B |
| 2 | | | B |

Split

| d | a' |
|---|----|
| A | 9 |
| B | 8 |

Split/Apply/Combine

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort
- Split/Apply/Combine (Summarize)

Table Actions

Table Wrangling



a'
17
Summarize
SELECT

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 3 | | | A |

Insert
INSERT/UNION

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |

Filter
WHERE

| a | b | c | d | e |
|---|---|---|---|----|
| 5 | | | A | 10 |
| 1 | | | B | 2 |
| 4 | | | A | 8 |
| 5 | | | B | 10 |
| 2 | | | B | 4 |

Add/Transform
SELECT

| a | d |
|---|---|
| 5 | A |
| 1 | B |
| 4 | A |
| 5 | B |
| 2 | B |

Select
SELECT

| a | b | c | d |
|---|---|---|---|
| 1 | | | B |
| 2 | | | B |
| 4 | | | A |
| 5 | | | A |
| 5 | | | B |
| 2 | | | B |

Sort
ORDER BY

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |
| 1 | | | B |
| 5 | | | B |
| 2 | | | B |

Split
GROUP BY

| d | a' |
|---|----|
| A | 9 |
| B | 8 |

Split/Apply/Combine
SELECT + GROUP BY

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort
- Split/Apply/Combine (Summarize)

Table Actions

Table Wrangling



a'
17
Summarize
summarize

| a | b | c | d | e |
|---|---|---|----|---|
| 5 | | A | 10 | |
| 1 | | B | 2 | |
| 4 | | A | 8 | |
| 5 | | B | 10 | |
| 2 | | B | 4 | |

Add/Transform
bind_cols/mutate

| a | b | c | d |
|---|---|---|---|
| 5 | | A | |
| 1 | | B | |
| 4 | | A | |
| 5 | | B | |
| 2 | | B | |

Select
select

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |
| 3 | | | A |

Insert
bind_rows

| a | b | c | d |
|---|---|---|---|
| 5 | | A | |
| 4 | | | A |

Filter
filter

| a | b | c | d |
|---|---|---|---|
| 1 | | | B |
| 2 | | | B |
| 4 | | | A |
| 5 | | A | |
| 2 | | B | |

Sort
arrange

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |
| 1 | | | B |
| 5 | | | B |
| 2 | | | B |

Split
group_by

| d | a' |
|---|----|
| A | 9 |
| B | 8 |

Split/Apply/Combine
group_by + summarize

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort
- Split/Apply/Combine (Summarize)

Table Actions

Table Wrangling



a'
17
Summarize
agg

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 1 | | | B |
| 4 | | | A |
| 5 | | | B |
| 2 | | | B |
| 3 | | | A |

Insert
concat

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |

Filter
query

| a | b | c | d | e |
|---|---|---|---|----|
| 5 | | | A | 10 |
| 1 | | | B | 2 |
| 4 | | | A | 8 |
| 5 | | | B | 10 |
| 2 | | | B | 4 |

| a | d |
|---|---|
| 5 | A |
| 1 | B |
| 4 | A |
| 5 | B |
| 2 | B |

| a | b | c | d |
|---|---|---|---|
| 1 | | | B |
| 2 | | | B |
| 4 | | | A |
| 5 | | | A |
| 5 | | | B |
| 2 | | | B |

| a | b | c | d |
|---|---|---|---|
| 5 | | | A |
| 4 | | | A |
| 1 | | | B |
| 5 | | | B |
| 2 | | | B |

| d | a' |
|---|----|
| A | 9 |
| B | 8 |

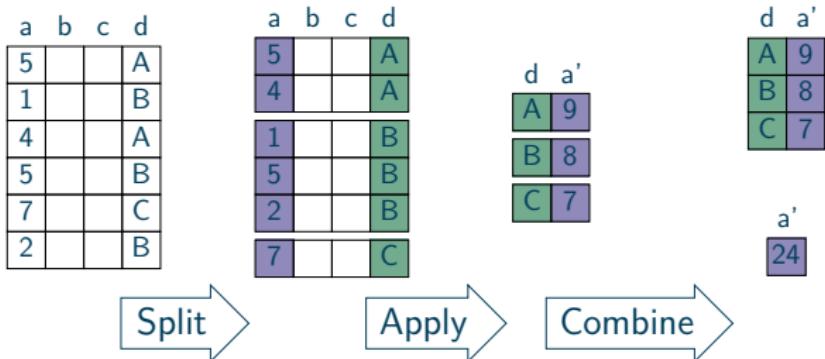
Split/Apply/Combine
groupby + agg

Verbs

- Rows: Filter, Insert
- Columns: Remove, Add
- Rows/Columns: Transform, Summarize, Sort
- Split/Apply/Combine (Summarize)

Split/Apply/Combine

Table Wrangling

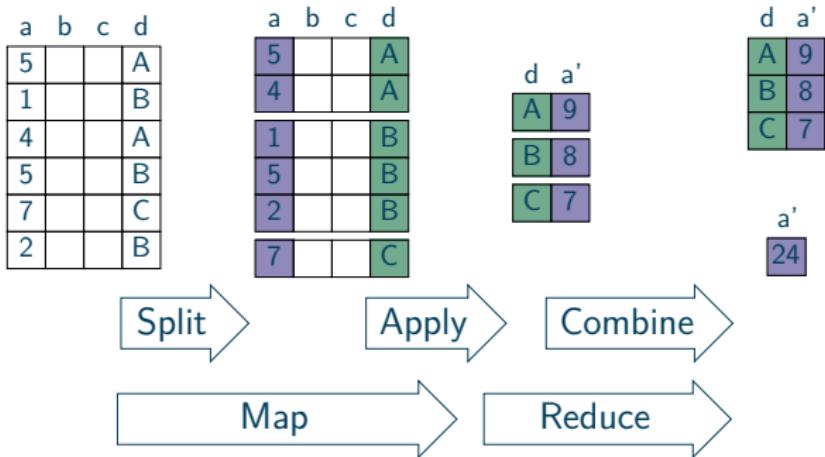


Split/Apply/Combine

- **Split** the data by some **grouping** variable
- **Apply** some function to each group independently
- **Combine** the data into some **output**.

Split/Apply/Combine

Table Wrangling



Split/Apply/Combine

- **Split** the data by some **grouping** variable
- **Apply** some function to each group independently
- **Combine** the data into some **output**.
- **Map/Reduce** ~ **Split/Apply/ Combine**.

Joining Two Tables

Table Wrangling



| | d | k | e |
|---|---|---|---|
| 1 | A | | |
| 2 | B | | |
| 3 | A | | |
| 4 | D | | |

| | a | b | c | k |
|---|---|---|---|---|
| 1 | | | | A |
| 2 | | | | B |
| 3 | | | | A |
| 4 | | | | C |
| 5 | | | | B |

Join between two tables along a key

Joining Two Tables

Table Wrangling



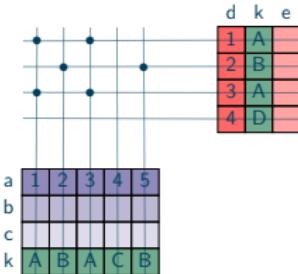
| | d | k | e |
|---|---|---|---|
| 1 | A | | |
| 2 | B | | |
| 3 | A | | |
| 4 | D | | |

| | | | | | |
|---|---|---|---|---|---|
| a | 1 | 2 | 3 | 4 | 5 |
| b | | | | | |
| c | | | | | |
| k | A | B | A | C | B |

Join between two tables along a key

Joining Two Tables

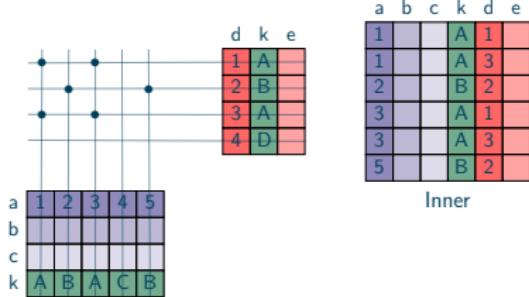
Table Wrangling



Join between two tables along a key

Joining Two Tables

Table Wrangling

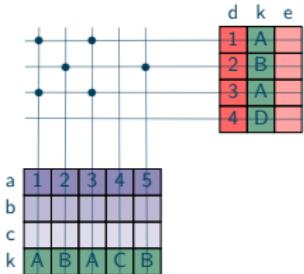


Join between two tables along a key

- Inner

Joining Two Tables

Table Wrangling



| | a | b | c | k | d | e |
|---|---|---|---|---|---|---|
| 1 | | | | A | 1 | |
| 1 | | | | A | 3 | |
| 2 | | | | B | 2 | |
| 3 | | | | A | 1 | |
| 3 | | | | A | 3 | |
| 4 | | | | C | ? | ? |
| 5 | | | | B | 2 | |

Left

Join between two tables along a key

- Inner
- Outer: left

Joining Two Tables

Table Wrangling



| | d | k | e |
|---|---|---|---|
| 1 | A | | |
| 2 | B | | |
| 3 | A | | |
| 4 | D | | |

| a | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| b | | | | | |
| c | | | | | |
| k | A | B | A | C | B |

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 3 | | | A | 1 | |
| 2 | | | B | 2 | |
| 5 | | | B | 2 | |
| 1 | | | A | 3 | |
| 3 | | | A | 3 | |
| ? | ? | ? | D | 4 | |

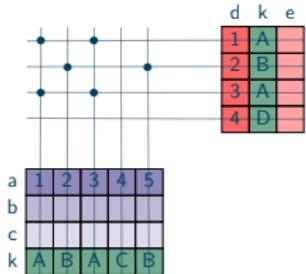
Right

Join between two tables along a key

- Inner
- Outer: left, right

Joining Two Tables

Table Wrangling



| a | b | c | k | d | e |
|---|---|---|---|---|-----|
| 1 | | | A | 1 | red |
| 1 | | | A | 3 | red |
| 2 | | | B | 2 | red |
| 3 | | | A | 1 | red |
| 3 | | | A | 3 | red |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | red |
| ? | ? | ? | D | 4 | red |
| ? | ? | ? | | | |

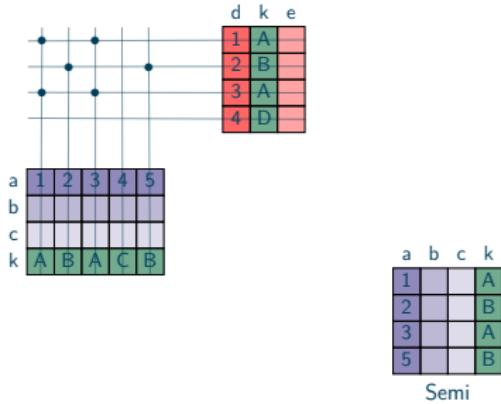
Full

Join between two tables along a key

- Inner
- Outer: left, right, full

Joining Two Tables

Table Wrangling

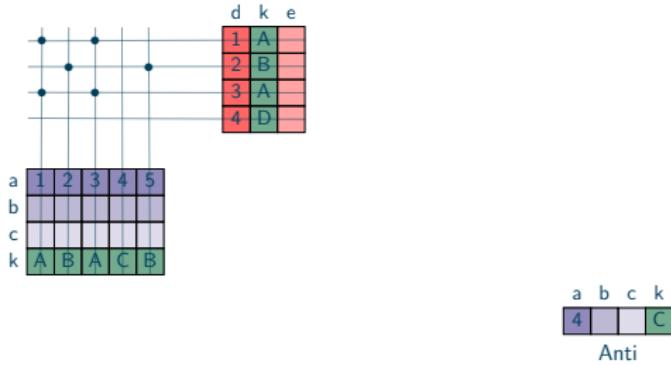


Join between two tables along a key

- Inner
- Outer: left, right, full
- Other: semi

Joining Two Tables

Table Wrangling



Join between two tables along a key

- Inner
- Outer: left, right, full
- Other: semi, anti

Joining Two Tables

Table Wrangling



| | d | k | e |
|---|---|---|---|
| 1 | A | | |
| 2 | B | | |
| 3 | A | | |
| 4 | D | | |

| a | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| b | | | | | |
| c | | | | | |
| k | A | B | A | C | B |

Inner

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 3 | | | B | 2 | |
| 5 | | | | | |

Left

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 3 | | | C | ? | ? |
| 4 | | | | | |
| 5 | | | B | 2 | |
| 5 | | | | | |
| ? | | | | | |
| ? | | | | | |
| ? | | | | | |

Right

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 3 | | | C | ? | ? |
| 4 | | | A | 3 | |
| 3 | | | A | 3 | |
| ? | | | D | 4 | |
| ? | | | | | |
| ? | | | | | |
| ? | | | | | |

Full

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |
| 5 | | | B | 2 | |
| ? | | | D | 4 | |
| ? | | | | | |
| ? | | | | | |
| ? | | | | | |

Semi

| a | b | c | k |
|---|---|---|---|
| 1 | | | A |
| 2 | | | B |
| 3 | | | A |
| 5 | | | B |

Anti

| a | b | c | k |
|---|---|---|---|
| 4 | | | C |

Join between two tables along a key

- Inner
- Outer: left, right, full
- Other: semi, anti

Joining Two Tables

Table Wrangling



| | d | k | e | |
|---|---|---|---|--|
| 1 | A | | | |
| 2 | B | | | |
| 3 | A | | | |
| 4 | D | | | |

| a | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| b | | | | | |
| c | | | | | |
| k | A | B | A | C | B |

Inner JOIN

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 5 | | | B | 2 | |

Left LEFT JOIN

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |

Right RIGHT JOIN

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 3 | | | A | 1 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 5 | | | B | 2 | |
| 1 | | | A | 3 | |
| 3 | | | A | 3 | |
| ? | ? | ? | D | 4 | |

Full FULL JOIN

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |
| ? | ? | ? | D | 4 | |

Semi SEMI JOIN

| a | b | c | k |
|---|---|---|---|
| 1 | | | A |
| 2 | | | B |
| 3 | | | A |
| 5 | | | B |

Anti ANTI JOIN

| a | b | c | k |
|---|---|---|---|
| 4 | | | C |

Join between two tables along a key

- Inner
- Outer: left, right, full
- Other: semi, anti

Joining Two Tables

Table Wrangling



| | d | k | e | |
|---|---|---|---|--|
| 1 | A | | | |
| 2 | B | | | |
| 3 | A | | | |
| 4 | D | | | |

| a | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| b | | | | | |
| c | | | | | |
| k | A | B | A | C | B |

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 3 | | | B | 2 | |

Inner
inner_join

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 3 | | | B | 2 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |

Left
left_join

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 1 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 5 | | | B | 2 | |
| 1 | | | A | 3 | |
| 3 | | | A | 3 | |
| ? | | | D | 4 | |

Right
right_join

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |
| ? | | | D | 4 | |

Full
full_join

| a | b | c | k |
|---|---|---|---|
| 1 | | | A |
| 2 | | | B |
| 3 | | | A |
| 5 | | | B |

Semi
semi_join

| a | b | c | k |
|---|---|---|---|
| 4 | | | C |

Anti
anti_join

Join between two tables along a key

- Inner
- Outer: left, right, full
- Other: semi, anti

Joining Two Tables

Table Wrangling

| | d | k | e |
|---|---|---|---|
| 1 | A | | |
| 2 | B | | |
| 3 | A | | |
| 4 | D | | |

| a | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| b | | | | | |
| c | | | | | |
| k | A | B | A | C | B |

Inner merge (inner)

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 3 | | | B | 2 | |

Left merge (left)

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |

Right merge (right)

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |
| ? | | | D | 4 | |

Full merge (full)

| a | b | c | k | d | e |
|---|---|---|---|---|---|
| 1 | | | A | 1 | |
| 1 | | | A | 3 | |
| 2 | | | B | 2 | |
| 2 | | | B | 2 | |
| 3 | | | A | 1 | |
| 3 | | | A | 3 | |
| 4 | | | C | ? | ? |
| 5 | | | B | 2 | |
| ? | | | D | 4 | |

Semi merge (inner) + loc

Anti merge (inner) + loc

| a | b | c | k |
|---|---|---|---|
| 1 | | | A |
| 2 | | | B |
| 3 | | | A |
| 5 | | | B |

| a | b | c | k |
|---|---|---|---|
| 4 | | | C |

Join between two tables along a key

- Inner
- Outer: left, right, full
- Other: semi, anti

| | a | b | c | d | e |
|---|---|---|---|---|---|
| E | A | 5 | | | |
| E | B | 1 | | | |
| F | A | 4 | | | |
| F | B | 5 | | | |
| G | C | 7 | | | |
| G | B | 2 | | | |

- Several ways to organize information.

Table Reshaping

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | E | A | 5 | | |
| b | E | B | 1 | | |
| c | F | A | 4 | | |
| d | F | B | 5 | | |
| e | G | C | 7 | | |
| f | G | B | 2 | | |

Pivot

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | E | A | 5 | | |
| b | E | B | 1 | | |
| c | F | A | 4 | | |
| d | F | B | 5 | | |
| e | G | C | 7 | | |
| f | G | B | 2 | | |

 Pivot

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

 Pivot/Unpivot

| a | b | c |
|---|---|---|
| E | A | 5 |
| E | B | 1 |
| F | A | 4 |
| F | B | 5 |
| G | C | 7 |
| G | B | 2 |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns
- Unpivot: Columns to categories

| | a | b | c | d | e |
|---|---|---|---|---|---|
| E | A | 5 | | | |
| E | B | 1 | | | |
| F | A | 4 | | | |
| F | B | 5 | | | |
| G | C | 7 | | | |
| G | B | 2 | | | |

Pivot
PIVOT

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

Pivot/Unpivot
PIVOT/UNPIVOT

| a | b | c |
|---|---|---|
| E | A | 5 |
| E | B | 1 |
| F | A | 4 |
| F | B | 5 |
| G | C | 7 |
| G | B | 2 |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns
- Unpivot: Columns to categories

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | E | A | 5 | | |
| b | E | B | 1 | | |
| c | F | A | 4 | | |
| d | F | B | 5 | | |
| e | G | C | 7 | | |
| f | G | B | 2 | | |

Pivot
pivot_wider

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

Pivot/Unpivot
pivot_wider/pivot_longer

| a | b | c |
|---|---|---|
| E | A | 5 |
| E | B | 1 |
| F | A | 4 |
| F | B | 5 |
| G | C | 7 |
| G | B | 2 |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns
- Unpivot: Columns to categories

| | a | b | c | d | e |
|---|---|---|---|---|---|
| E | A | 5 | | | |
| E | B | 1 | | | |
| F | A | 4 | | | |
| F | B | 5 | | | |
| G | C | 7 | | | |
| G | B | 2 | | | |

 Pivot

pivot

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

 Pivot/Unpivot

pivot/melt

| | a | b | c |
|---|---|---|---|
| E | A | 5 | |
| E | B | 1 | |
| F | A | 4 | |
| F | B | 5 | |
| G | C | 7 | |
| G | B | 2 | |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns
- Unpivot: Columns to categories

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | E | A | 5 | | |
| b | E | B | 1 | | |
| c | F | A | 4 | | |
| d | F | B | 5 | | |
| e | G | C | 7 | | |
| f | G | B | 2 | | |

 Pivot

Wide

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

 Pivot/Unpivot

Long

| a | b | c |
|---|---|---|
| E | A | 5 |
| E | B | 1 |
| F | A | 4 |
| F | B | 5 |
| G | C | 7 |
| G | B | 2 |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns
- Unpivot: Columns to categories
- Wide / Long format

| | a | b | c | d | e |
|---|---|---|---|---|---|
| E | A | 5 | | | |
| E | B | 1 | | | |
| F | A | 4 | | | |
| F | B | 5 | | | |
| G | C | 7 | | | |
| G | B | 2 | | | |

 Pivot

| a | A | B | C |
|---|---|---|---|
| E | 5 | 1 | ? |
| F | 5 | 5 | ? |
| G | ? | 2 | 7 |

 Pivot/Unpivot

| a | b | c |
|---|---|---|
| E | A | 5 |
| E | B | 1 |
| F | A | 4 |
| F | B | 5 |
| G | C | 7 |
| G | B | 2 |

- Several ways to organize information.

Table Reshaping

- Pivot: Categories to columns
- Unpivot: Columns to categories

- Best format depends on the task

Local Files: CSV, JSON, xls,...

- Dedicated import(/export) libraries
- Often weakly typed.
- Cleaning often required.
- Typed variants exist: arrow, feather...

Database: SQL, NoSQL,...

- Dedicated DB connectors.
- SQL as a common language.
- Often already cleaned.
- Local files may be accessed through a DB interface...

Web: HTML

- Web scraping library
- Often a lot of cleaning
- Web API ~ Local files...

Outline

Visualization



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Outline

Visualization



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Data visualization can:

- provide a clear understanding of patterns in data
- detect hidden structures in data
- condense information

Why Data Visualization?

Visualization



Data visualization can:

- provide a clear understanding of patterns in data
- detect hidden structures in data
- condense information
- Anscombe's quartet example:

Anscombe Quartet

| | x1 | y1 | x2 | y2 | x3 | y3 | x4 | y4 |
|----|----|-------|----|------|----|-------|----|------|
| 1 | 10 | 8.04 | 10 | 9.14 | 10 | 7.46 | 8 | 6.58 |
| 2 | 8 | 6.95 | 8 | 8.14 | 8 | 6.77 | 8 | 5.76 |
| 3 | 13 | 7.58 | 13 | 8.74 | 13 | 12.74 | 8 | 7.71 |
| 4 | 9 | 8.81 | 9 | 8.77 | 9 | 7.11 | 8 | 8.84 |
| 5 | 11 | 8.33 | 11 | 9.26 | 11 | 7.81 | 8 | 8.47 |
| 6 | 14 | 9.96 | 14 | 8.1 | 14 | 8.84 | 8 | 7.04 |
| 7 | 6 | 7.24 | 6 | 6.13 | 6 | 6.08 | 8 | 5.25 |
| 8 | 4 | 4.26 | 4 | 3.1 | 4 | 5.39 | 19 | 12.5 |
| 9 | 12 | 10.84 | 12 | 9.13 | 12 | 8.15 | 8 | 5.56 |
| 10 | 7 | 4.82 | 7 | 7.26 | 7 | 6.42 | 8 | 7.91 |
| 11 | 5 | 5.68 | 5 | 4.74 | 5 | 5.73 | 8 | 6.89 |

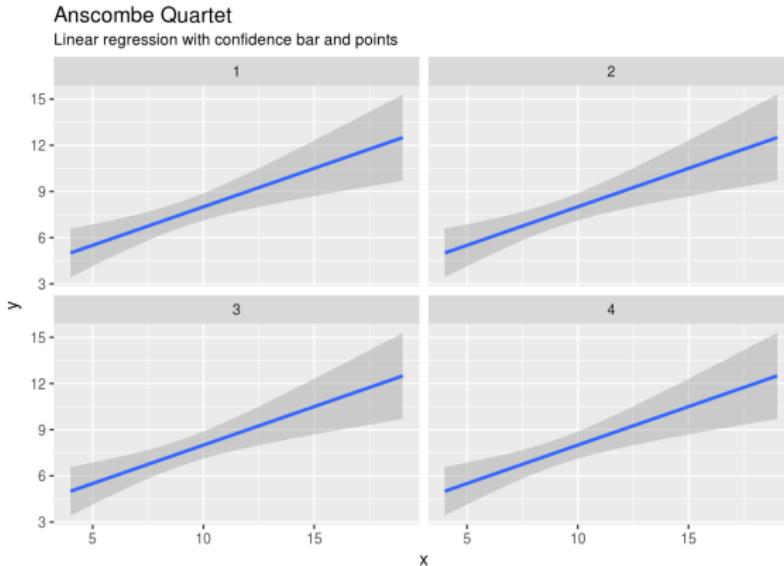
Why Data Visualization?

Visualization



Data visualization can:

- provide a clear understanding of patterns in data
- detect hidden structures in data
- condense information
- Anscombe's quartet example:



Why Data Visualization?

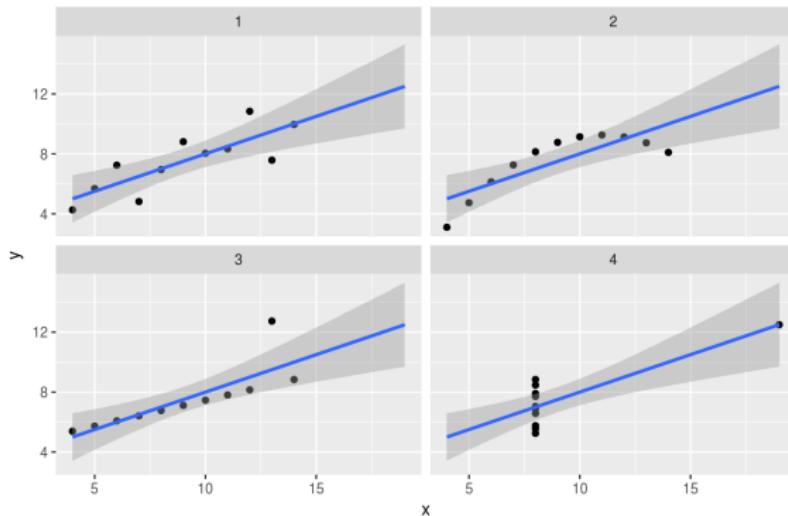
Visualization



Data visualization can:

- provide a clear understanding of patterns in data
- detect hidden structures in data
- condense information
- Anscombe's quartet example:

Anscombe Quartet
Linear regression with confidence bar



Focus of today

- Standard data visualization techniques,
- Review of various graphical techniques,
- Principle of good data presentation,
- Example of implementation with R.

Not the focus of this lecture

- *Infographics*
- Cognitive aspect of data perception...

Goal

- Exposure to various plotting techniques.
- Proof of concept with R.
- Visualize the power of appropriate data graphics techniques

Outline

Visualization



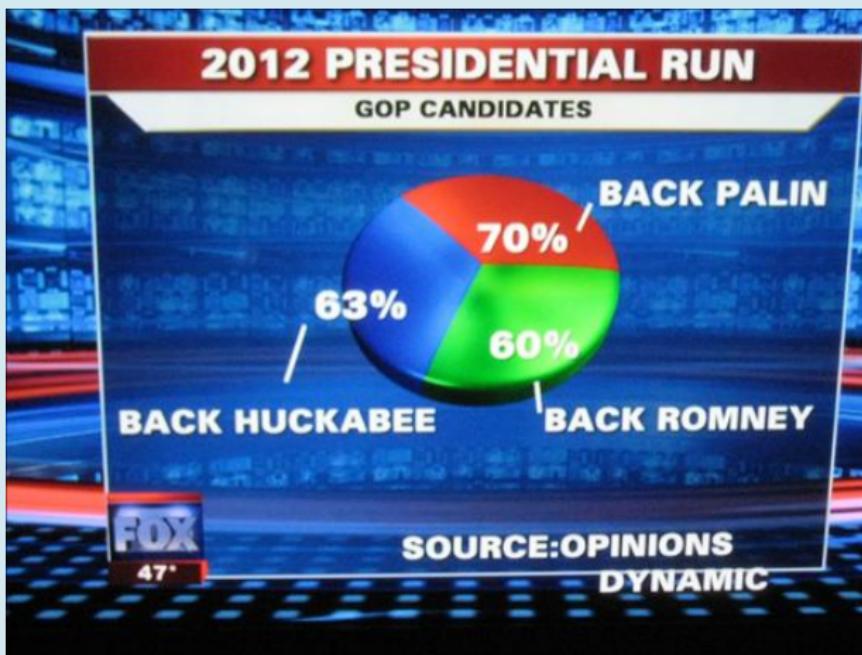
- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Bad Data Visualization

Visualization



No Comment!



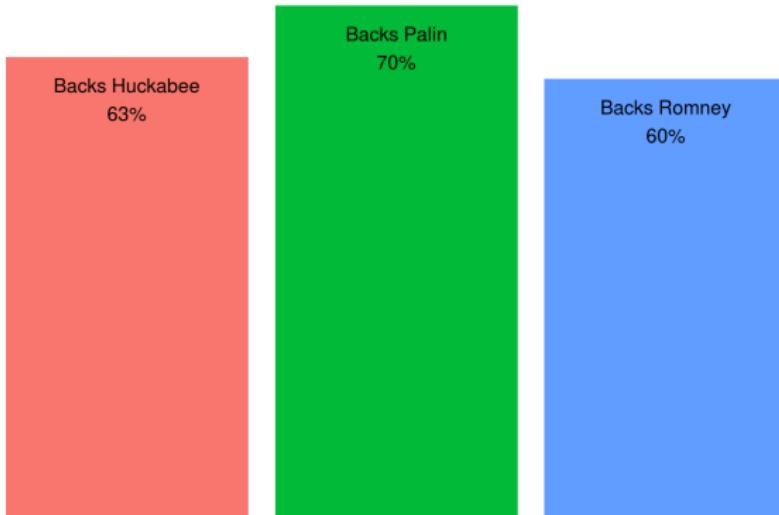
Bad Data Visualization

Visualization



A possible fix

Bar Plot
2012 Presidential Run



Bad Data Visualization

Visualization



Scale Issue + Missing Data

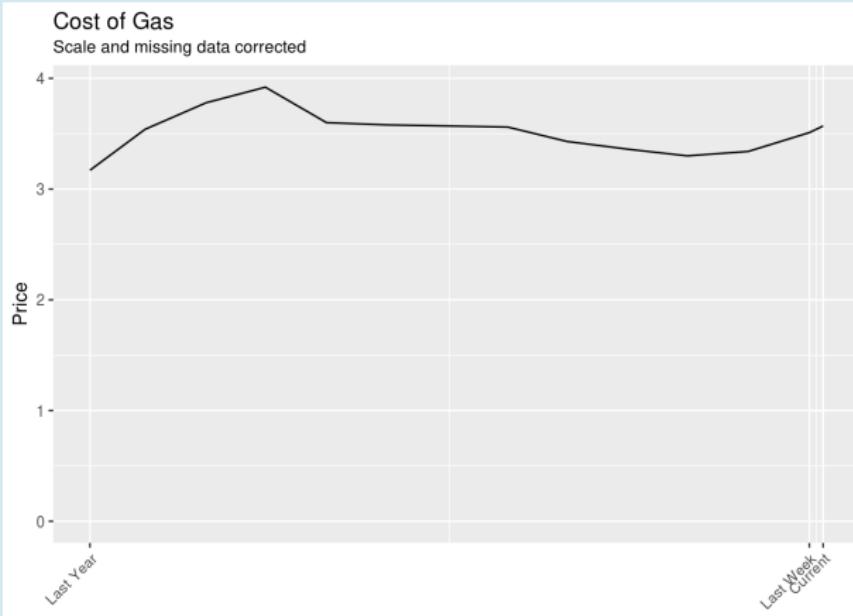


Bad Data Visualization

Visualization



Scale Issue Corrected + Missing Data Corrected



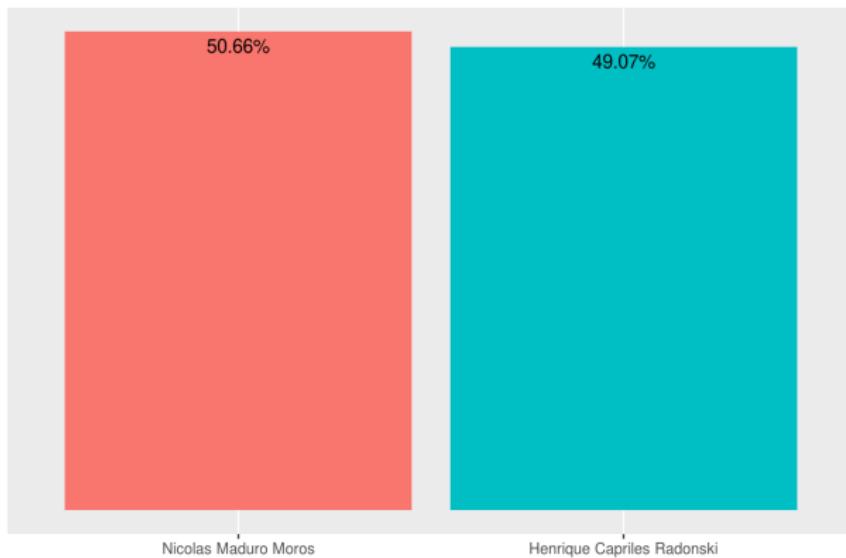
Truncated Axis Issue



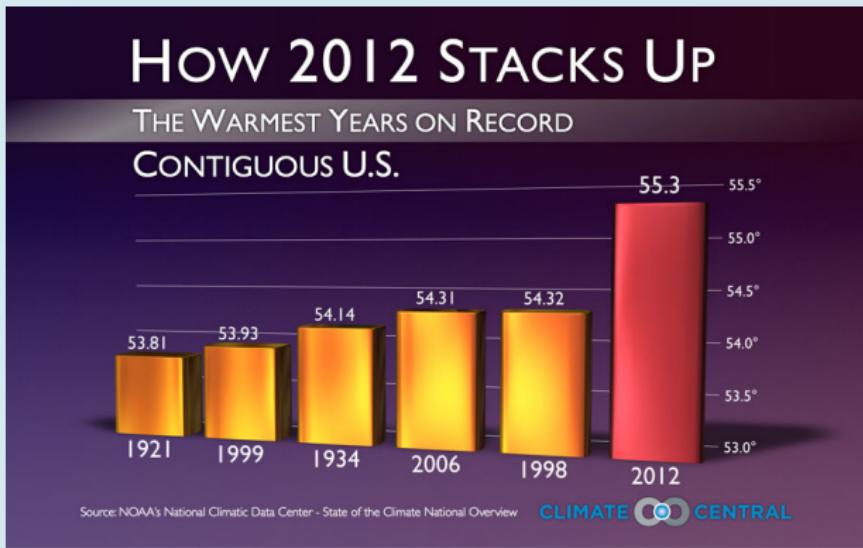
Truncated Axis Issue Corrected

2013 Venezuelan presidential election

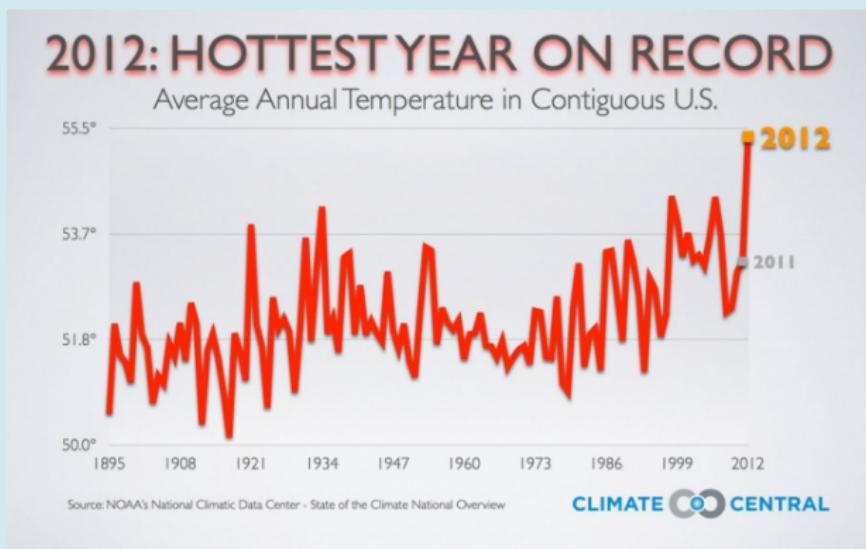
Truncated axis corrected



Scale Issue + Selection Issue



Scale Issue Corrected + Selection Issue Corrected

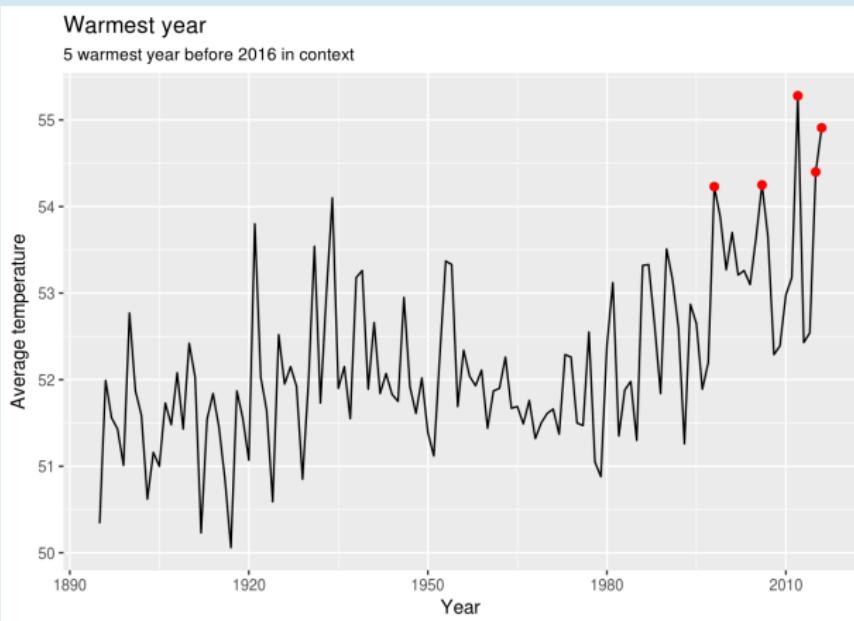


Bad Data Visualization

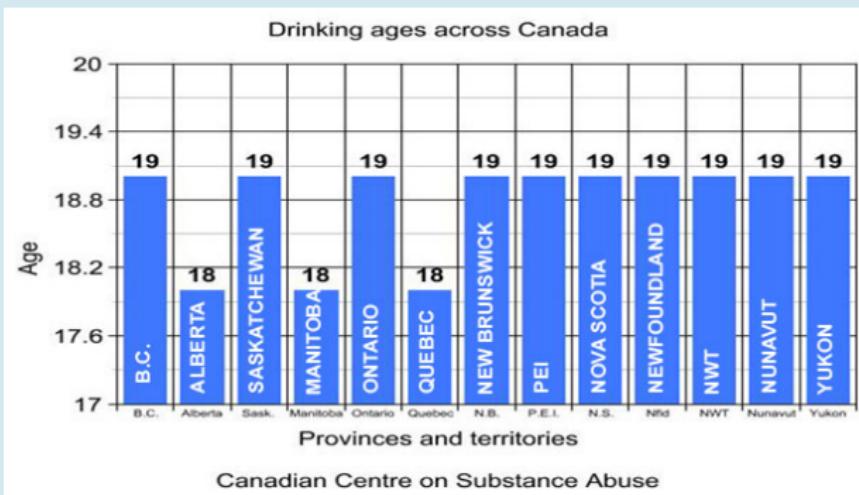
Visualization



Scale Issue Corrected + Selection Issue Corrected (2016)



Truncated Axis + Clutter Issue



Truncated Axis + Clutter Issue Corrected?

Drinking ages across Canada

Less clutter issue?

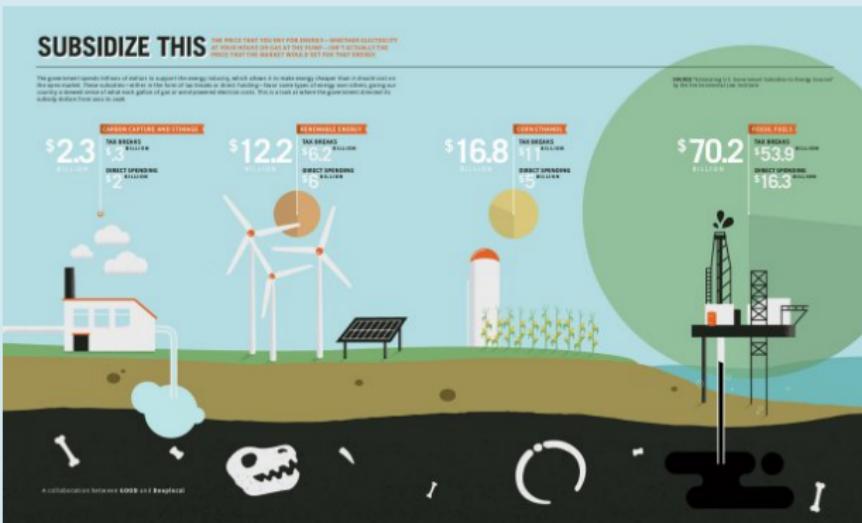
| | |
|---------------|----|
| Alberta | 18 |
| Manitoba | 18 |
| Quebec | 18 |
| B.C. | 19 |
| Saskatchewan | 19 |
| Ontario | 19 |
| New Brunswick | 19 |
| PEI | 19 |
| Nova Scotia | 19 |
| Newfoundland | 19 |
| NWT | 19 |
| Nunavut | 19 |
| Yukon | 19 |

Bad Data Visualization

Visualization



Area Issue

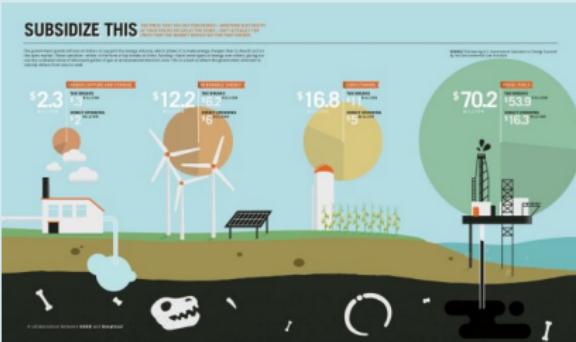
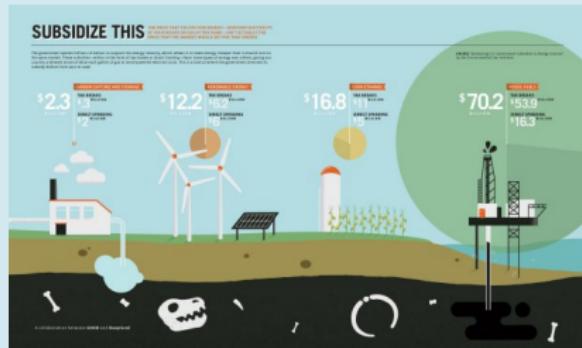


Bad Data Visualization

Visualization



Area Issue Corrected



Bad Data Visualization

Visualization



Area Issue



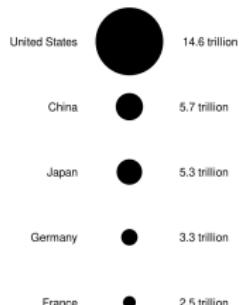
Bad Data Visualization

Visualization

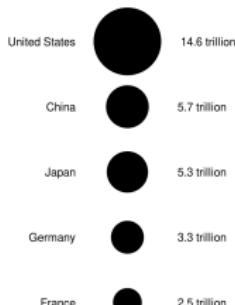


Area Issue Corrected

GDP 2012
Size issue



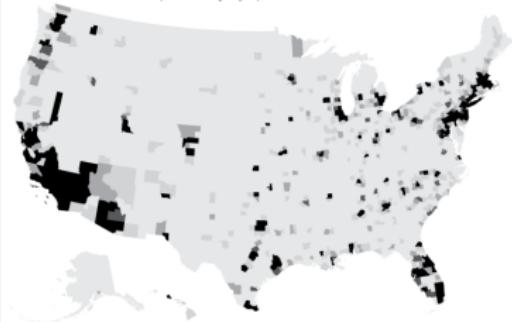
GDP 2012
Size issue corrected



Map Issue

SEEING ONLY IN ABSOLUTES

This is just population. When comparing across places, categories, or groups, you must compare fairly and consider relative values.

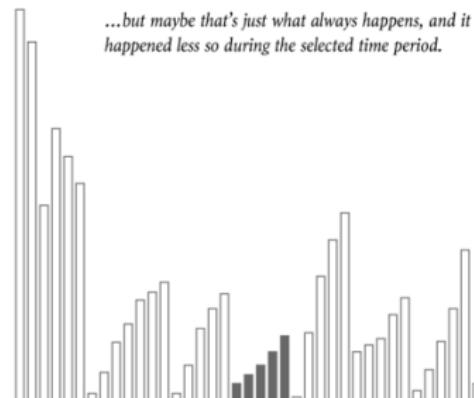


Limited Scope Issue (Corrected)

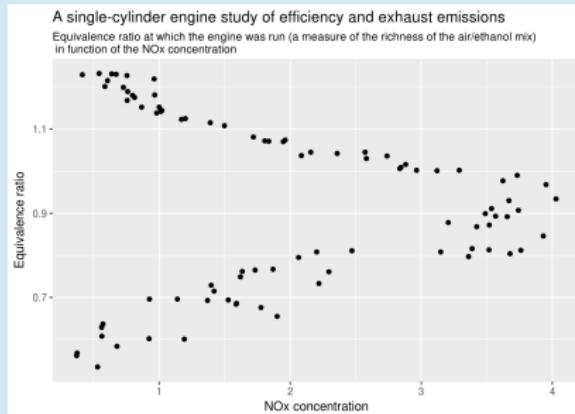
LIMITED SCOPE

It looks like something increased a lot...

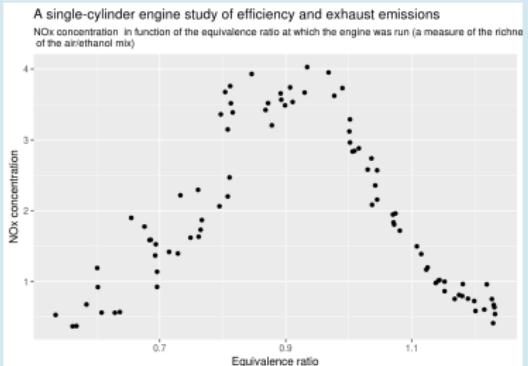
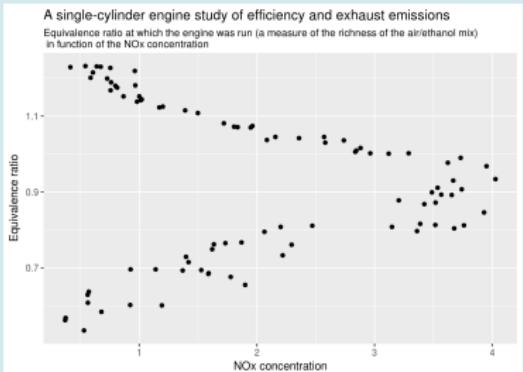
...but maybe that's just what always happens, and it happened less so during the selected time period.



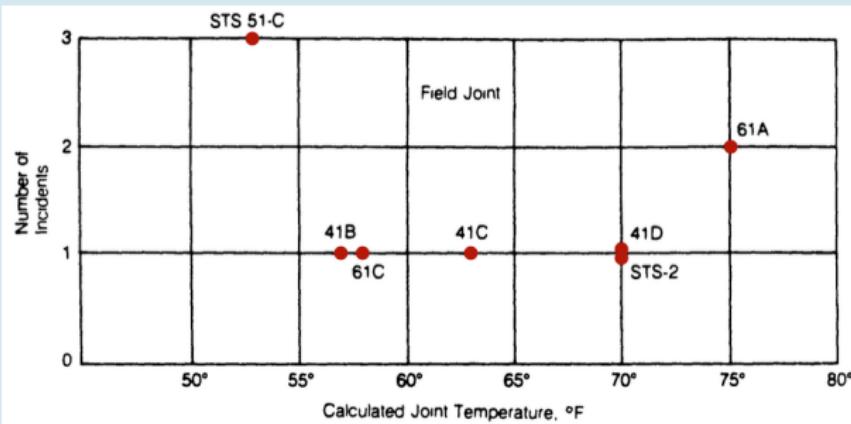
Unusual Axis Issue



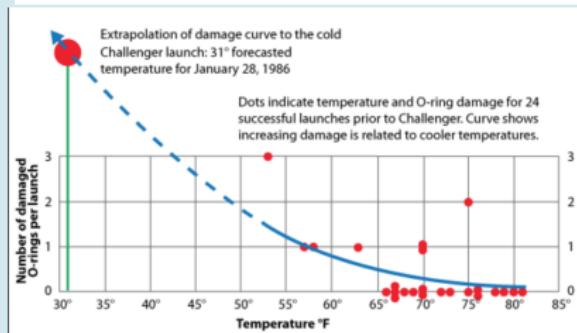
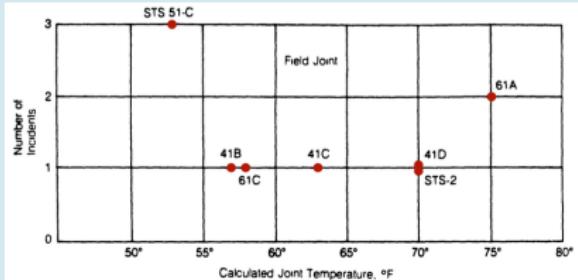
Unusual Axis Issue Corrected



Catastrophic Issue



Catastrophic Issue Corrected



Outline

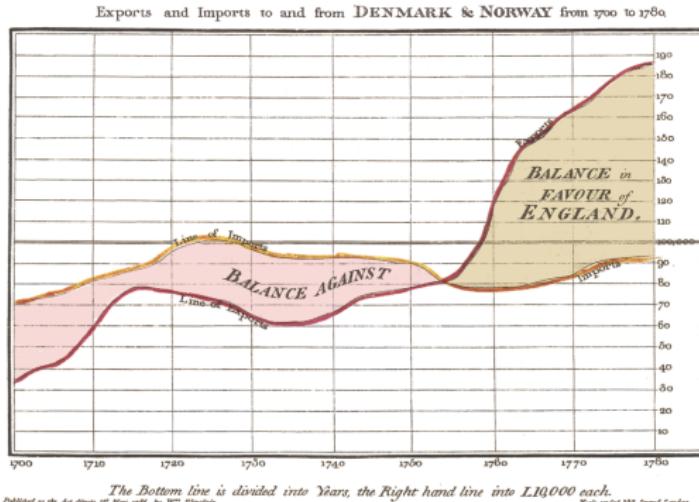
Historical Milestones

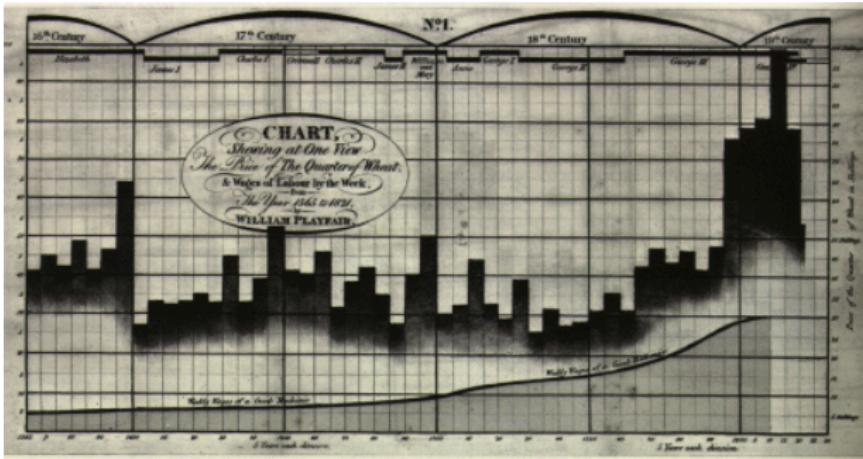


- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

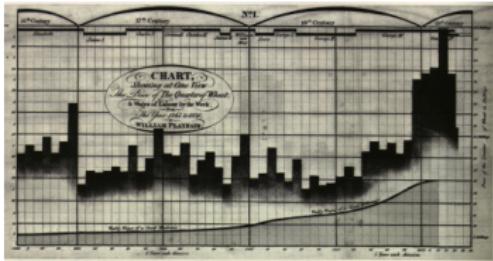
William Playfair (1759-1823)

- Generally viewed as the inventor of most of the common graphical forms used to display data: line plots, bar chart and pie chart





- Unfortunately often flawed...

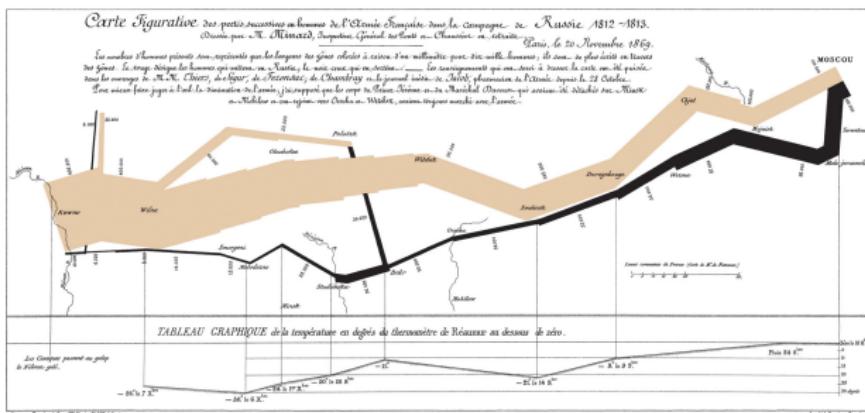


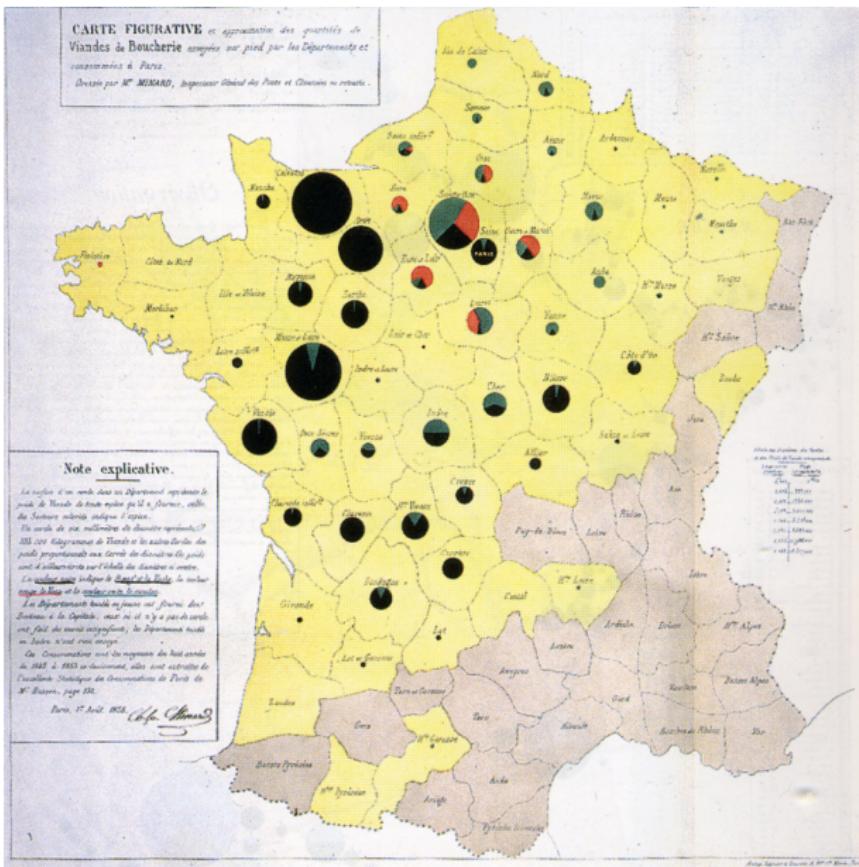
Issue Resolved?



Charles Minard (1781-1870)

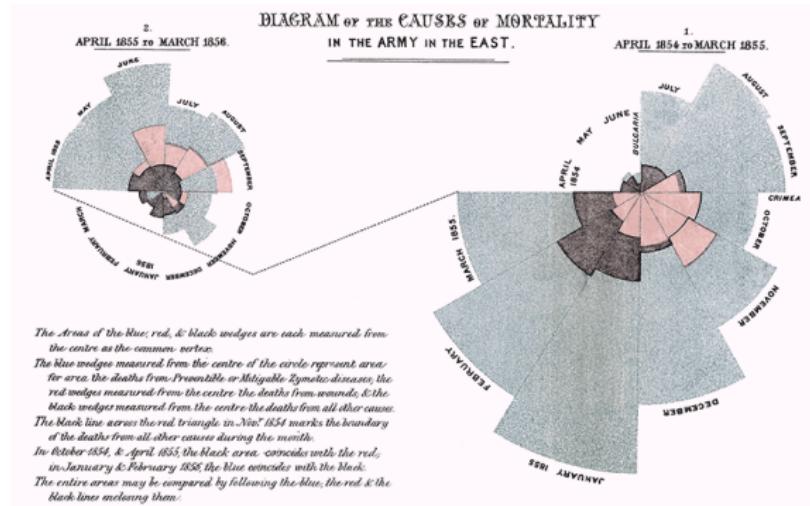
- contributed significantly in the field of information graphics in civil engineering and statistics and in particular in geographic maps.





Florence Nightingale (1820-1910)

- Mostly famous as the mother of modern nursing. She also contributed to the use of graphical representation.

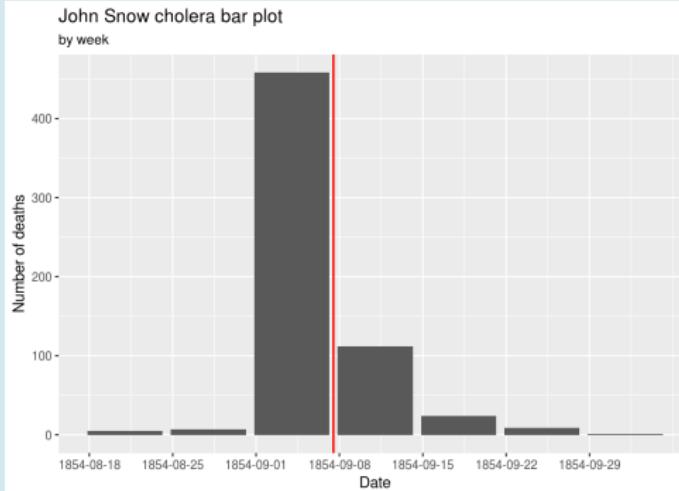


John Snow (1813-1858)

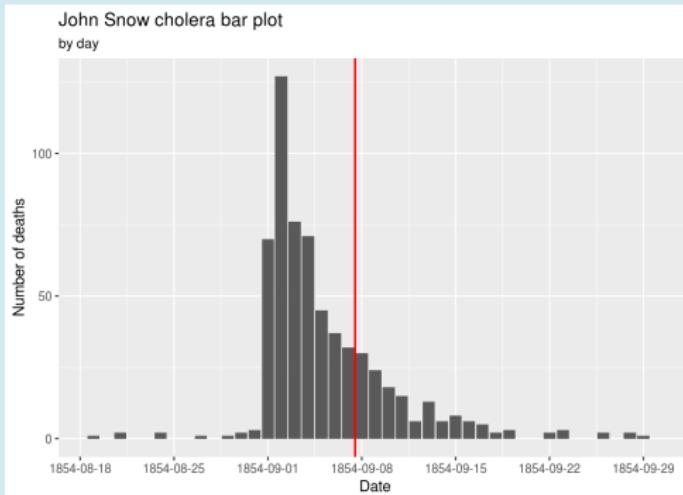
- An English physician famous for tracing the source of a cholera outbreak in London.



Story

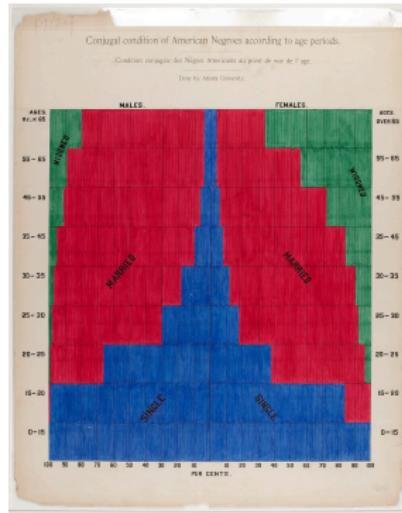


Reality



Du Bois

Historical Milestones

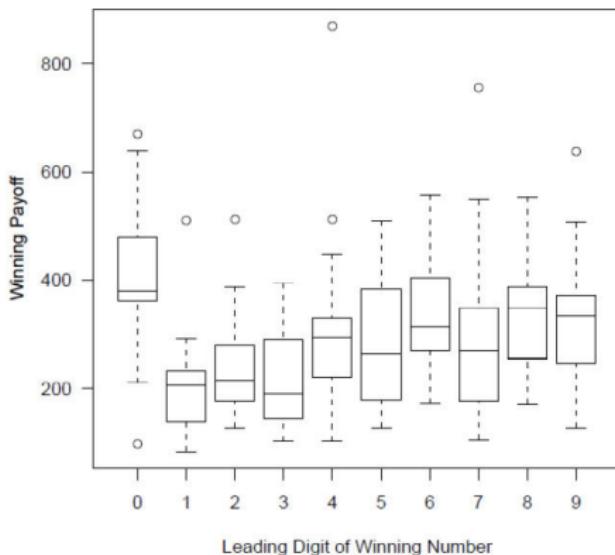


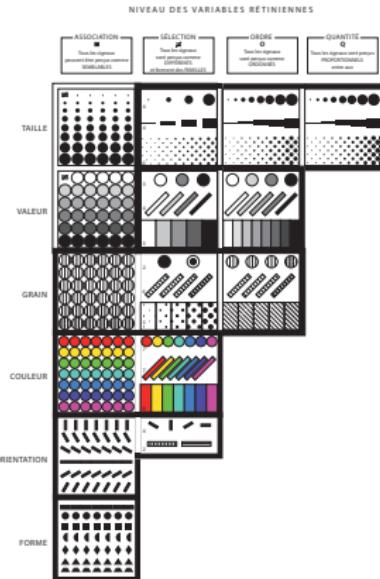
W. E. B. Du Bois (1868-1963)

- African American activist and sociologist
- “Exhibit of American Negroes” organized to represent black contributions to the United States at the 1900 Exposition Universelle in Paris

Ronald Fisher (1890-1962) and John Tukey (1915-2000)

- Advance graphical methods for the analysis of data.
- Fisher: plot the data to understand relationships.
- Tukey promoted Exploratory Data Analysis!
- Tukey created the box plot and the stem and leaf plot.





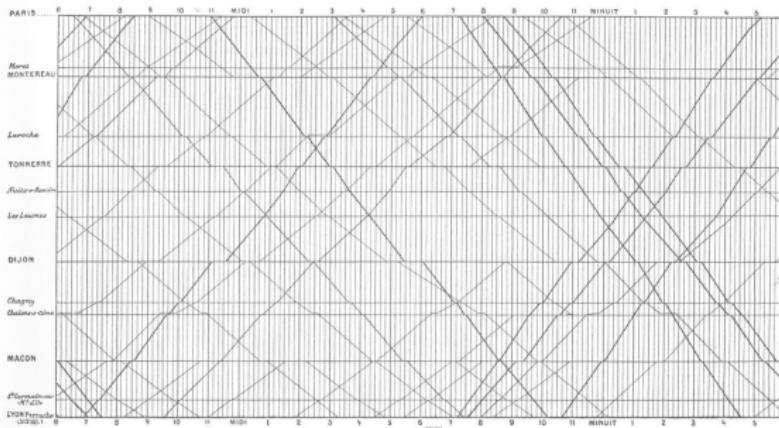
Jacques Bertin, "Sémiologie Graphique", 1973.

Jacques Bertin (1918-2010)

- *sémiologie graphique!*
- Systematic system of sign for information transmission.

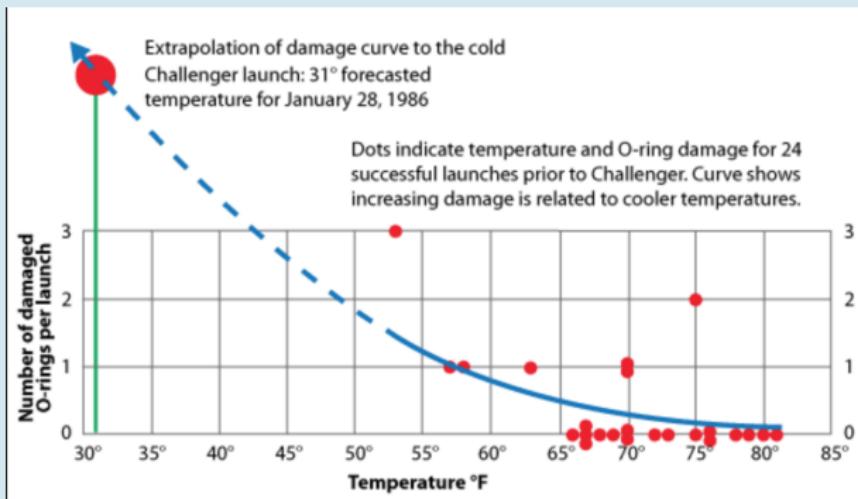
Edward Tufte (1942-)

- Probably the most widely known works on data visualization.
- Highly compressed, elegant, and informative data, as expressed in dense printed graphics.
- Importance of *beauty* aspect...



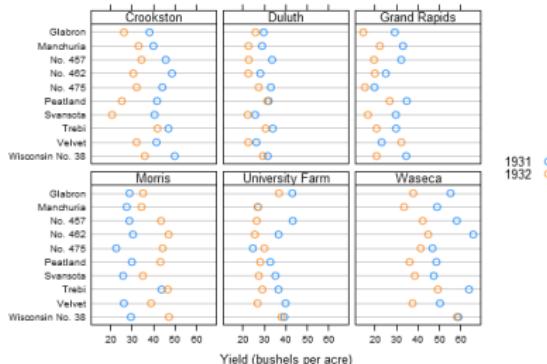
E.J. Marey (1885)

Challenger corrected!



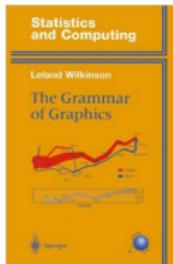
Small Multiples





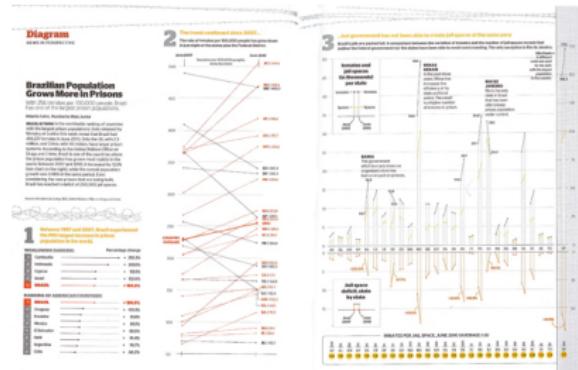
William Cleveland (1943-)

- His Elements of Graphing Data and Visualizing Data pioneered systematic considerations of legibility
- Cleveland is particularly known for promoting the dot plot as an alternative to bars and pies.
- The dot plot provides clarity and easy comparison of data.
- Cleveland also pioneered Trellis graphics that emphasizes comparison of multiple panels of data.



Leland Wilkinson (1945-)

- Its Grammar of Graphics was extremely influential in thinking about graphics:
 - Grammar means "rules for art and science"
 - Specifies rules both mathematical and aesthetic
 - Earlier graph producers focused on aesthetics of static content
 - Dynamic graphics and scientific visualization, by contrast, require sophisticated designs to enable brushing, drill-down, zooming, linking
 - The Grammar of Graphics is easily adapted to this approach
- ggplot2 (Hadley Wickham) is inspired by this formalism!



Alberto Cairo

- Data Journalism / Importance of storytelling!
- *The functional art* : An introduction to Information Graphics and Visualization, the communication of facts and data by means of charts, graphs, maps, and diagrams.
- *The truthful art* : Explains how to transform elementary principles of data and scientific reasoning into tools that you can use in daily life to interpret data sets and extract stories from them.

Outline

Principles



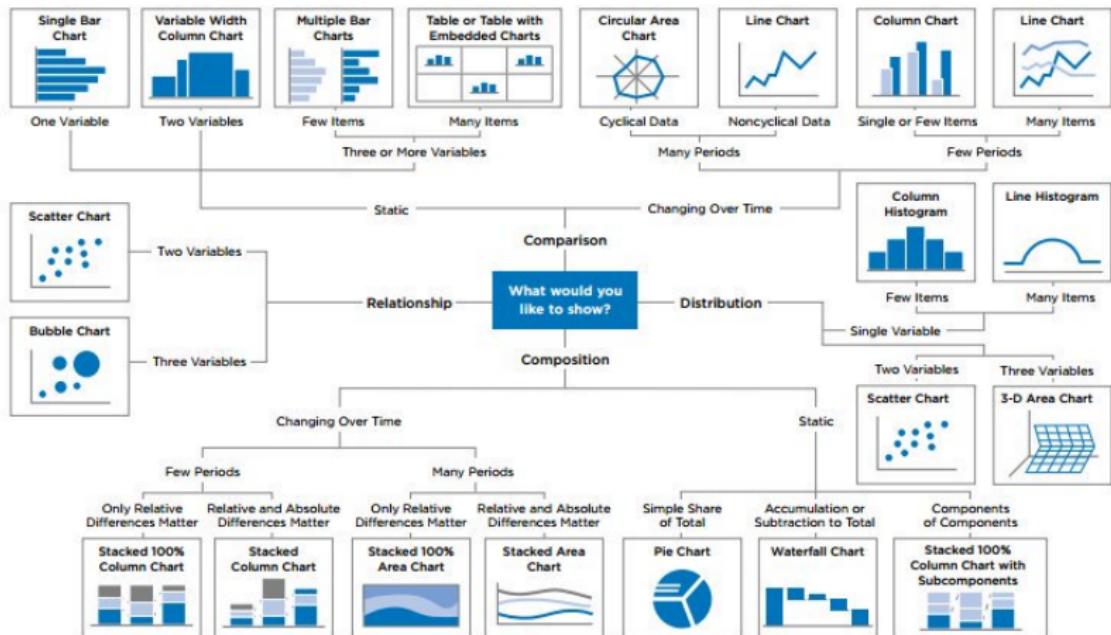
- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

How to Pick the Right Chart?

Principles



SELECTING THE APPROPRIATE CHART FOR STRATEGY PRESENTATIONS



How to Build a Graph?

Principles



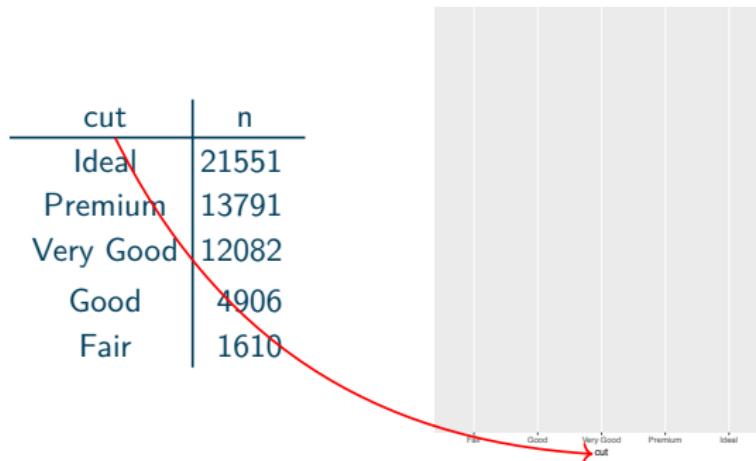
| cut | n |
|-----------|-------|
| Ideal | 21551 |
| Premium | 13791 |
| Very Good | 12082 |
| Good | 4906 |
| Fair | 1610 |

Bar Plot Example

- Start from the **data** and a blank canvas.

How to Build a Graph?

Principles

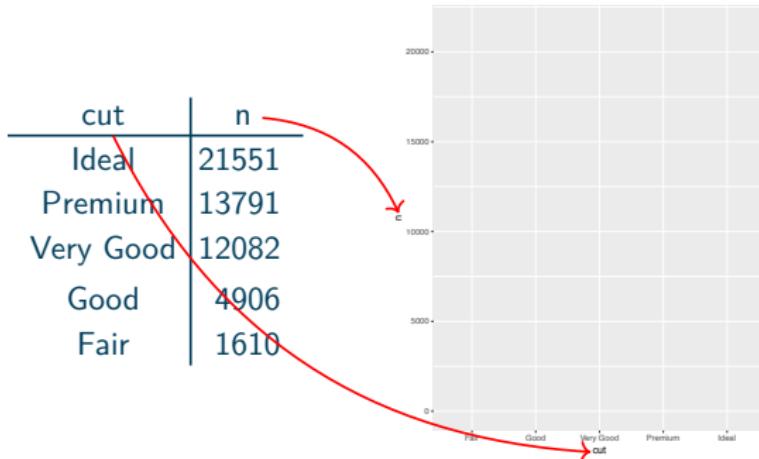


Bar Plot Example

- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.

How to Build a Graph?

Principles

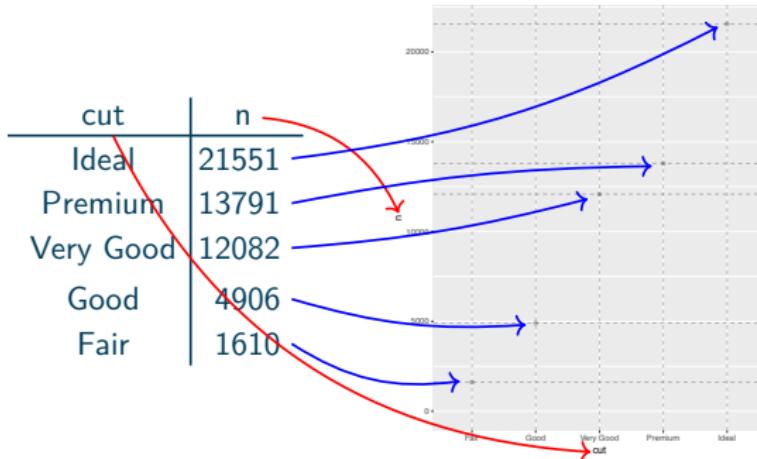


Bar Plot Example

- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.

How to Build a Graph?

Principles

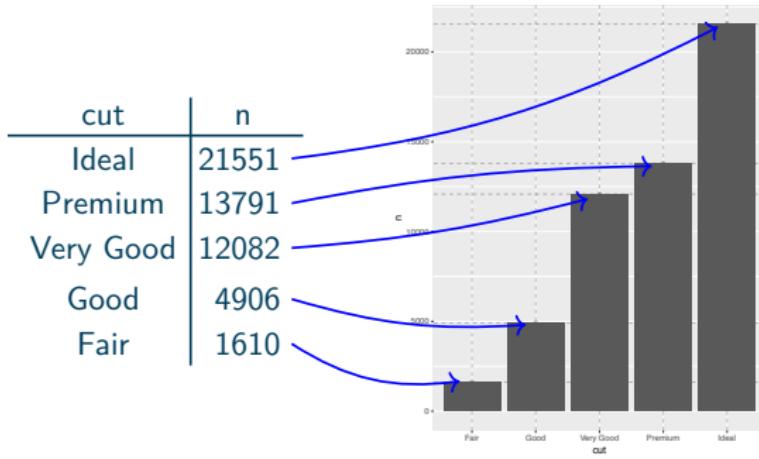


Bar Plot Example

- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.
- Place the data with a **meaningful** mapping.

How to Build a Graph?

Principles



Bar Plot Example

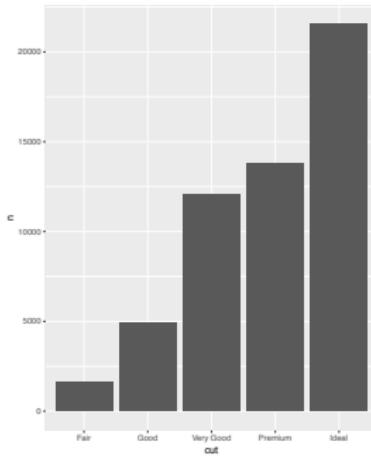
- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.
- Place the data with a **meaningful** mapping.
- Draw the graph with a certain **geometry**.

How to Build a Graph?

Principles



| cut | n |
|-----------|-------|
| Ideal | 21551 |
| Premium | 13791 |
| Very Good | 12082 |
| Good | 4906 |
| Fair | 1610 |



Bar Plot Example

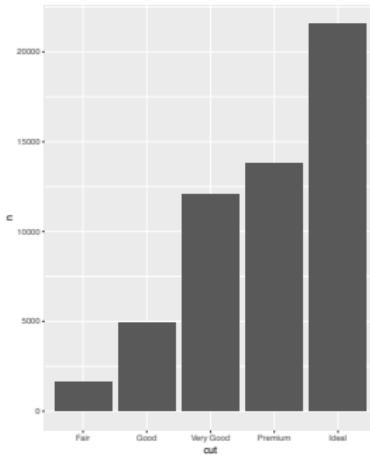
- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.
- Place the data with a **meaningful** mapping.
- Draw the graph with a certain **geometry**.

How to Build a Graph?

Principles



| cut | n |
|-----------|-------|
| Ideal | 21551 |
| Premium | 13791 |
| Very Good | 12082 |
| Good | 4906 |
| Fair | 1610 |

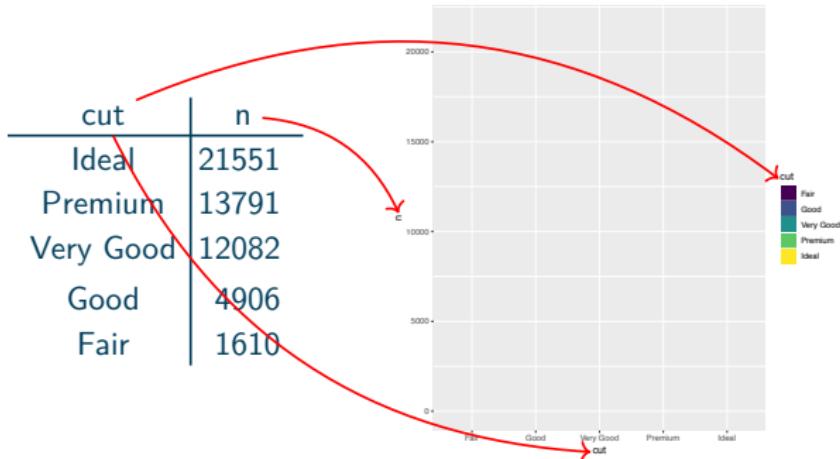


Bar Plot Example

- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.
- Place the data with a **meaningful** mapping.
- Draw the graph with a certain **geometry**.
- Systematic way of describing a graph.

How to Build a Graph?

Principles



Bar Plot Example

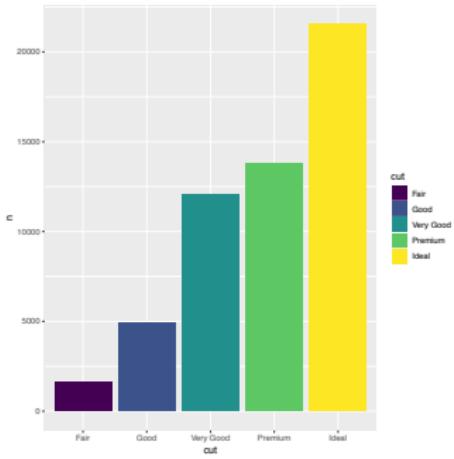
- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.
- Place the data with a **meaningful** mapping.
- Draw the graph with a certain **geometry**.
- Systematic way of describing a graph.

How to Build a Graph?

Principles



| cut | n |
|-----------|-------|
| Ideal | 21551 |
| Premium | 13791 |
| Very Good | 12082 |
| Good | 4906 |
| Fair | 1610 |

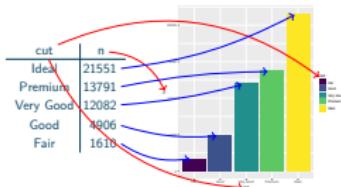


Bar Plot Example

- Start from the **data** and a blank canvas.
- Map **variables** to **axis**.
- Place the data with a **meaningful** mapping.
- Draw the graph with a certain **geometry**.
- Systematic way of describing a graph.

Grammar of Graphics

Principles



| | |
|------------------------------------|-------------|
| Describes all the non-data ink | Theme |
| Plotting space for the data | Coordinates |
| Statistical models & summaries | Statistics |
| Rows and columns of sub-plots | Facets |
| Shapes used to represent the data | Geometries |
| Scales onto which data is mapped | Aesthetics |
| The actual variables to be plotted | Data |

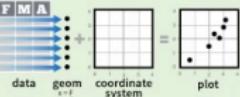


Wilkinson

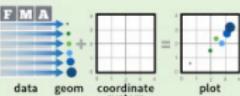
- **DATA** - weighting, reshaping, counting, bootstrapping
- **VARIABLES** - transform, sort, log, rank., resid., quant.
- **STATISTICS** - statistic preprocessing of data
- **AESTHETICS** - mapping between position/color/size/... and variable
- **SCALES** - nominal, ordinal, interval, ratio...
- **GEOMETRY** - line, area, etc., along with modifiers like jitter and dodge
- **COORDINATES** - refers to the coordinate system of the graph (cartesian, polar, etc.)
- **FACETS** - subgroups, multiway tables
- **GUIDES** - legends, axes, color scales, keys

ggplot2 cheatsheet

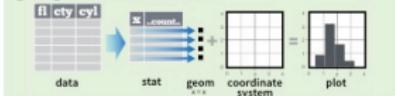
ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same few components: a **data** set, a set of **geoms**—visual marks that represent data points, and a **coordinate system**.



To display data values, map variables in the data set to aesthetic properties of the geom like **size**, **color**, and **x** and **y** locations.



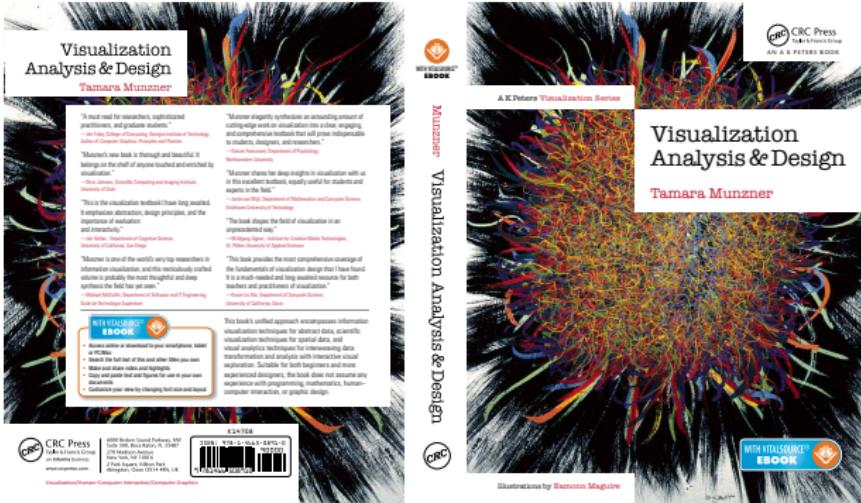
Some plots visualize a **transformation** of the original data set. Use a **stat** to choose a common transformation to visualize, e.g. `a + geom_bar(stat = "bin")`



- ggplot2 (Hadley Wickham) is inspired by this formalism!

Visualization Analysis and Design

Principles



Visualization Analysis and Design

- Book from Tamara Munzner
- Published by CRC Press in 2014
- Supplementary slides *taken* from her slidedesk!

Outline

Classical Graphs



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Outline

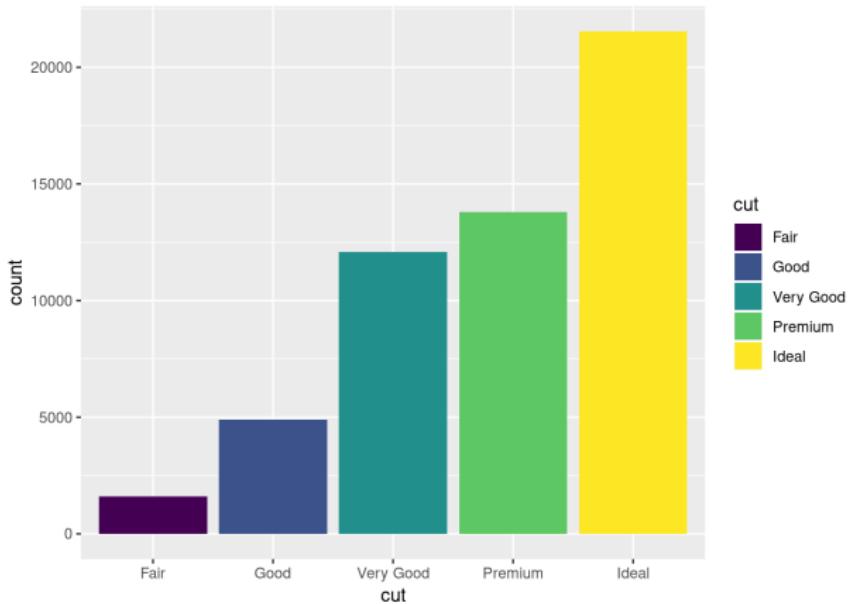
Classical Graphs



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Quantities - Bar

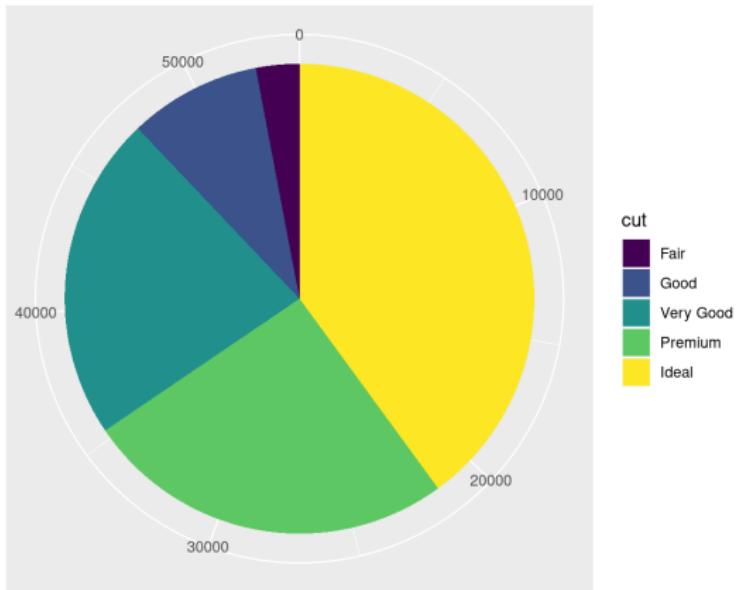
Classical Graphs



- Adapted to counts and quantities.

Proportion - Pie

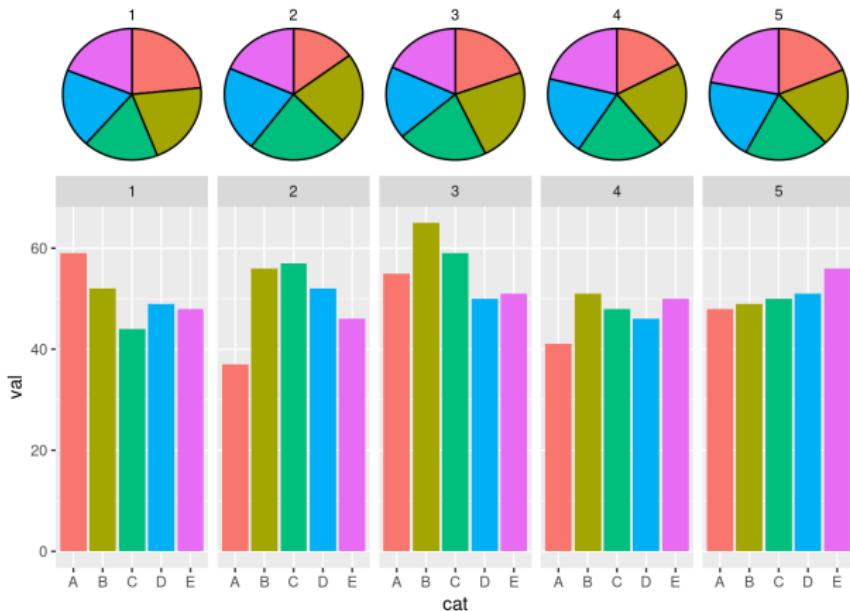
Classical Graphs



- Should not be used for comparison...

Proportion - Pie

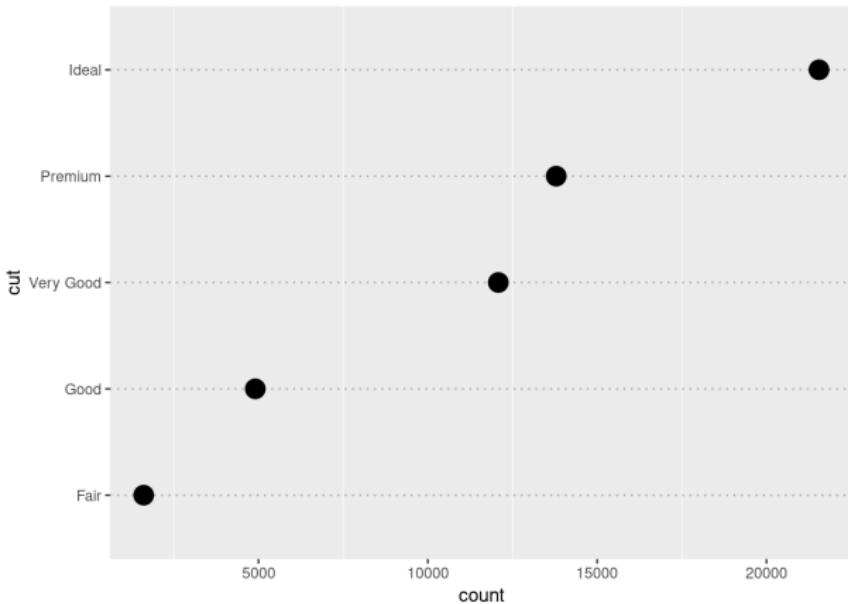
Classical Graphs



- Bar plots more efficient for comparison.
- Pie plots more efficient for global proportions!

Quantities - Cleveland Dot Plots

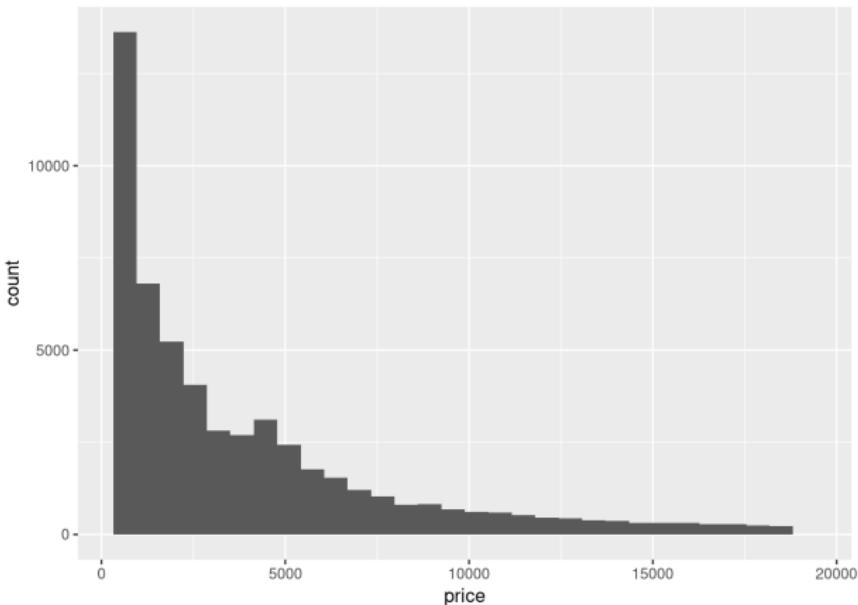
Classical Graphs



- Less *ink*, more pleasant...

Distribution - Histogram and density

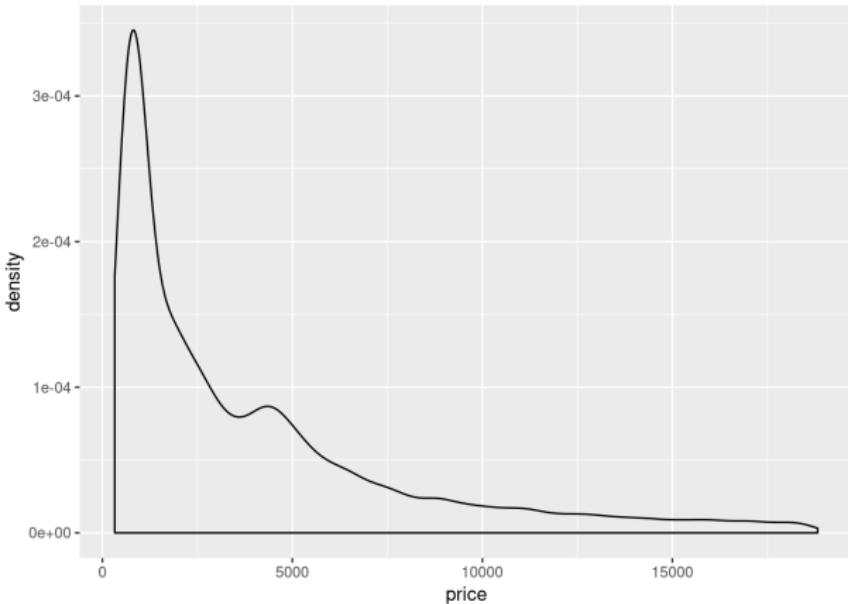
Classical Graphs



- Easily interpretable
- Adapted to continuous variable.

Distribution - Histogram and density

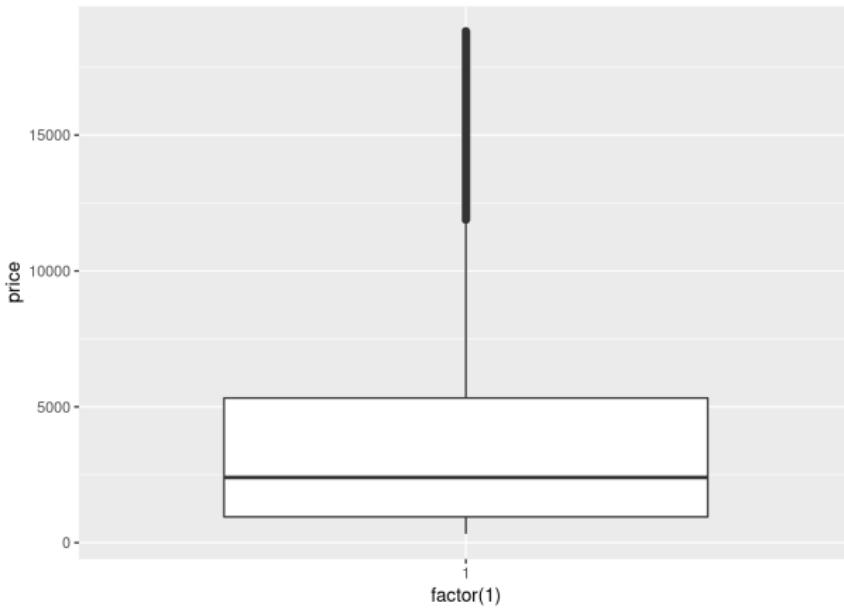
Classical Graphs



- Regularized view...

Distribution - Box and Whiskers

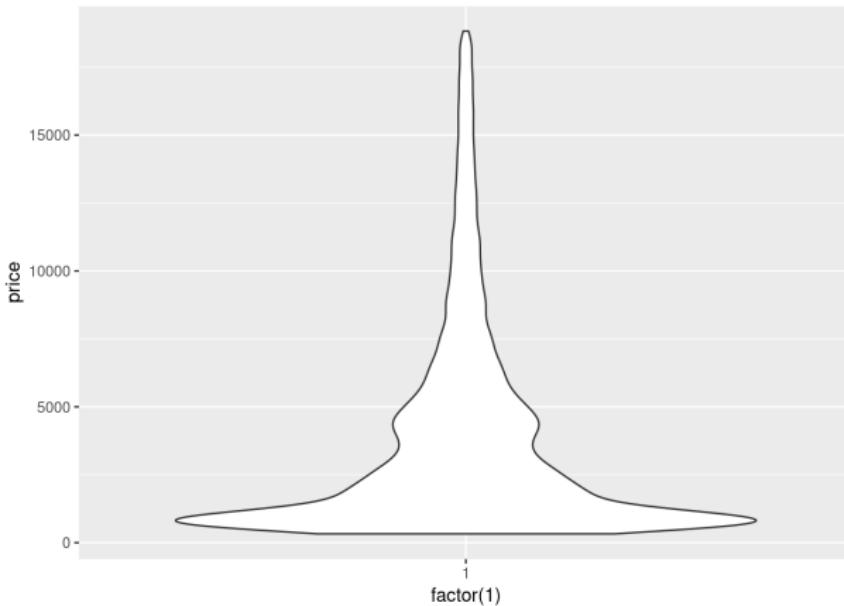
Classical Graphs



- Most classical representation after pie...

Distribution - Violin plot

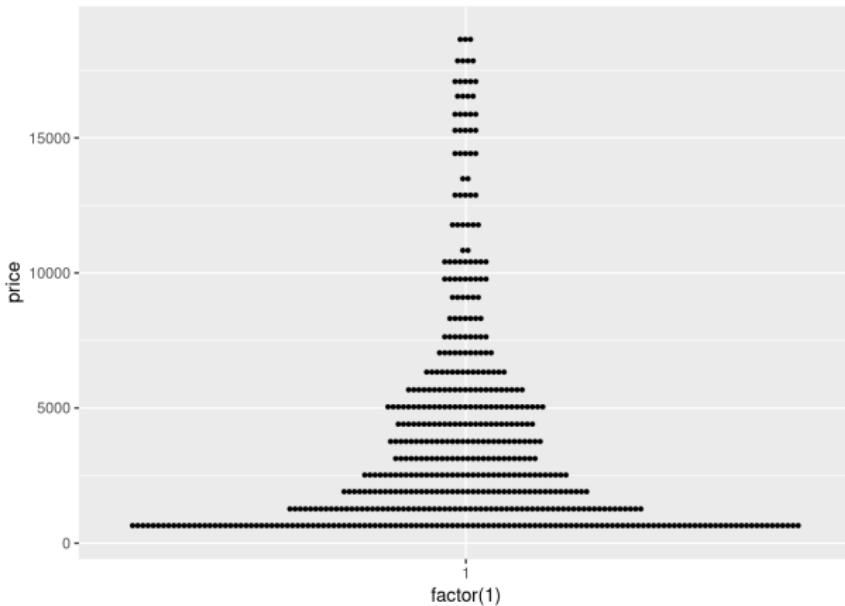
Classical Graphs



- Combined box plot and density estimation.

Distribution - Dot plots

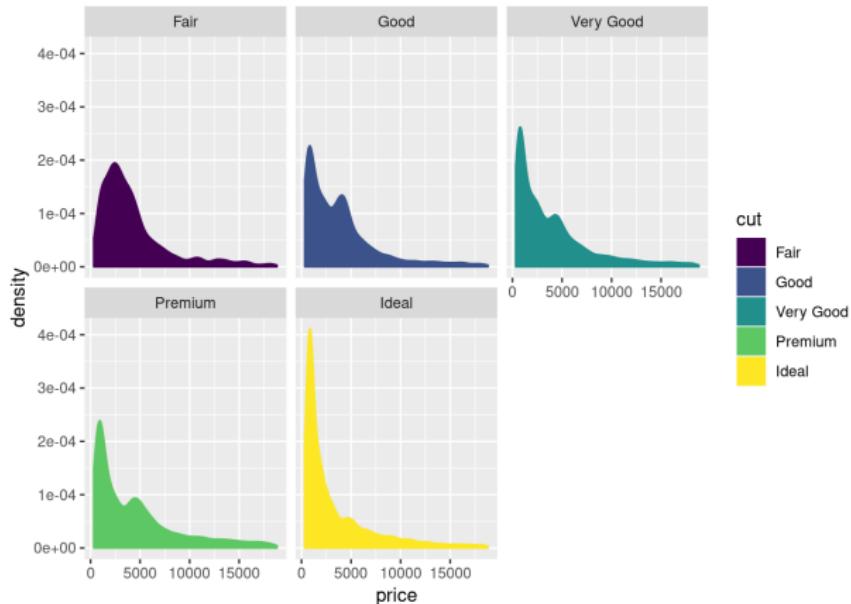
Classical Graphs



- Combined binning and individuals...

Grouping

Classical Graphs



- Key to construct complex representation.

Outline

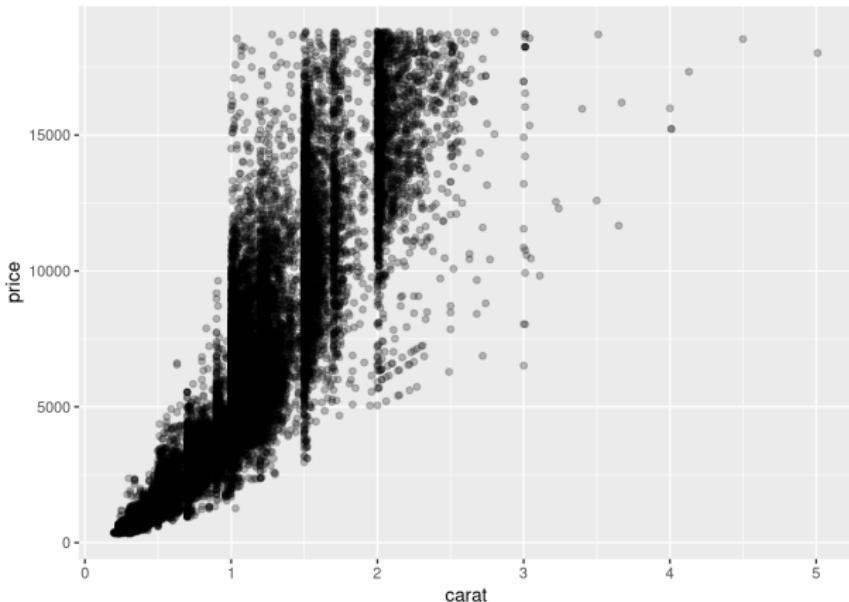
Classical Graphs



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Relation - Scatter Plot

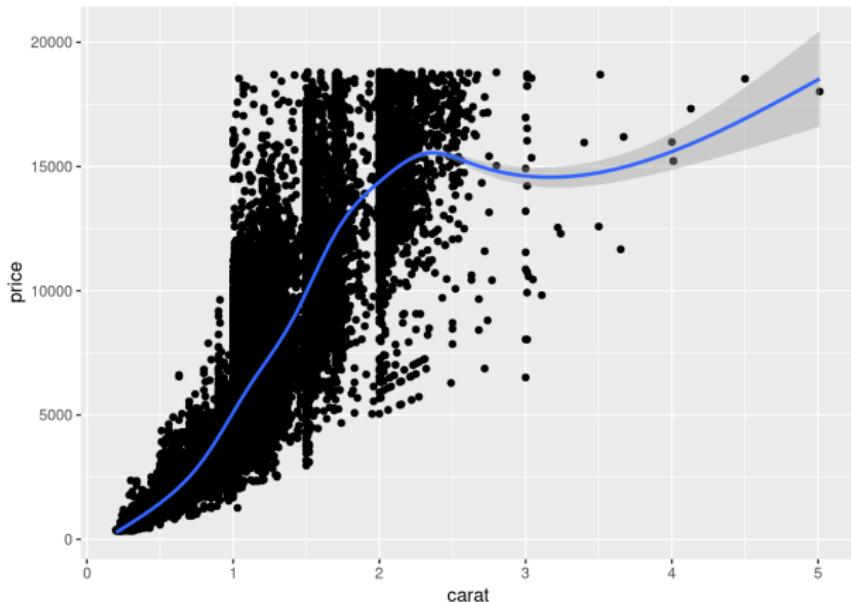
Classical Graphs



- Used to visualize the relationship between two variables.

Relation - Smoothing

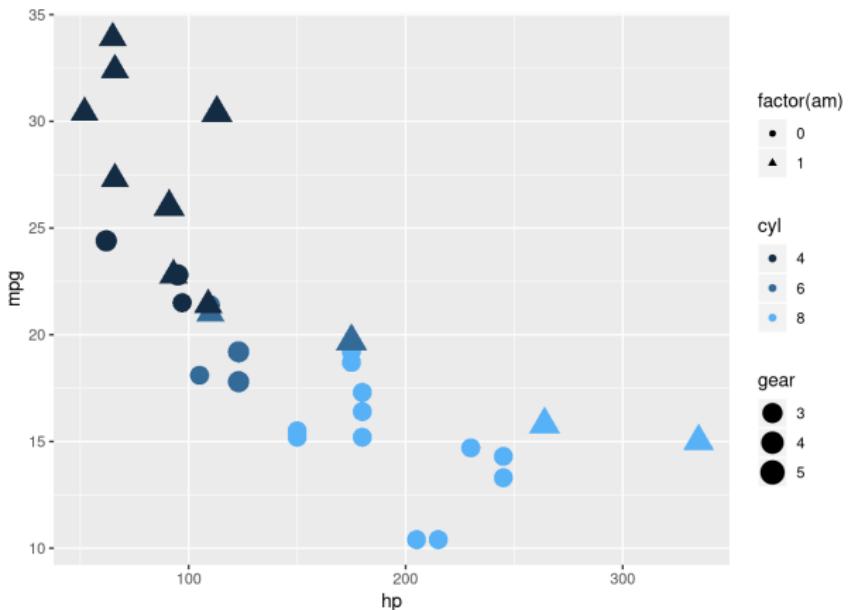
Classical Graphs



- Strong visual help.

Relation - Symbols

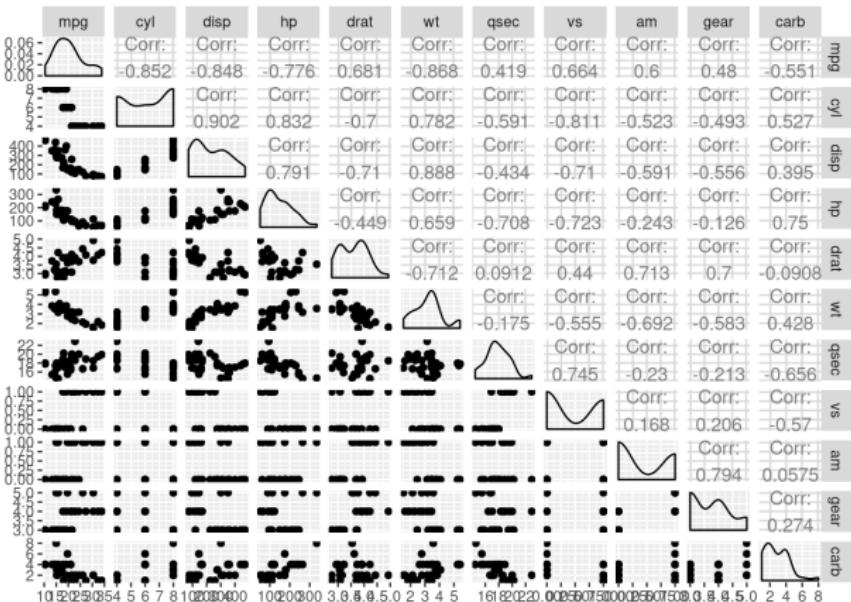
Classical Graphs



- Good idea to augment the information density...
- but can lead to too much complexity.

Grouping - Scatter Plot Matrix

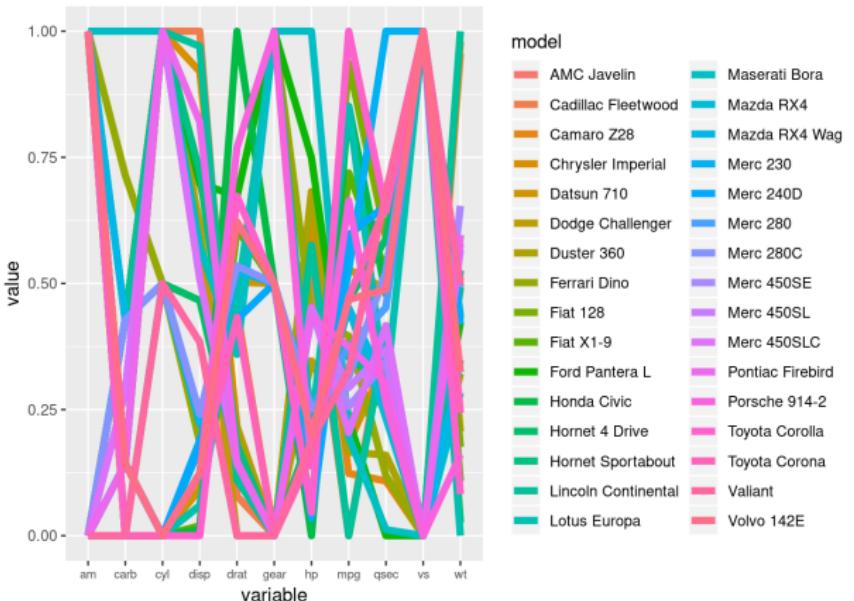
Classical Graphs



- Gather all the dependencies...

HighD - Parallel Coordinates / Radar Plot

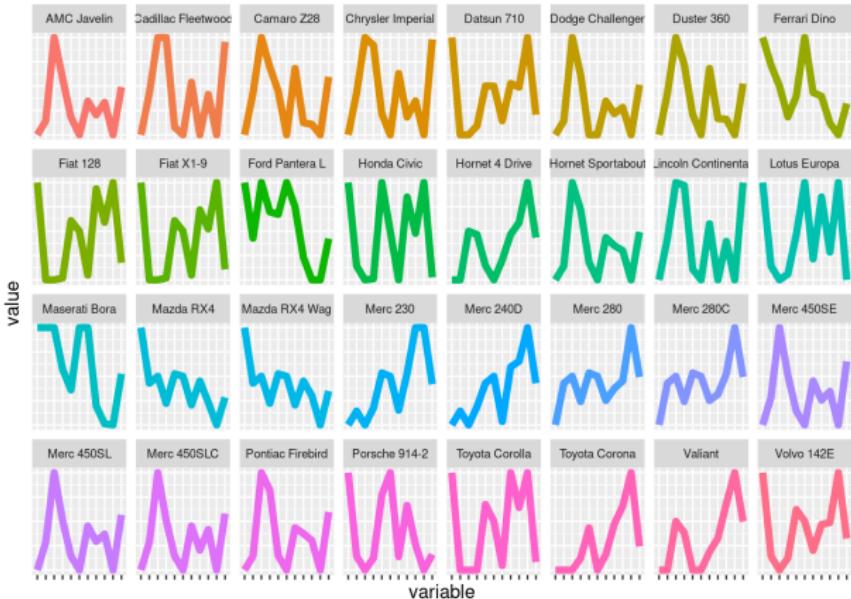
Classical Graphs



- Clever ideas to visualize groups.

HighD - Parallel Coordinates / Radar Plot

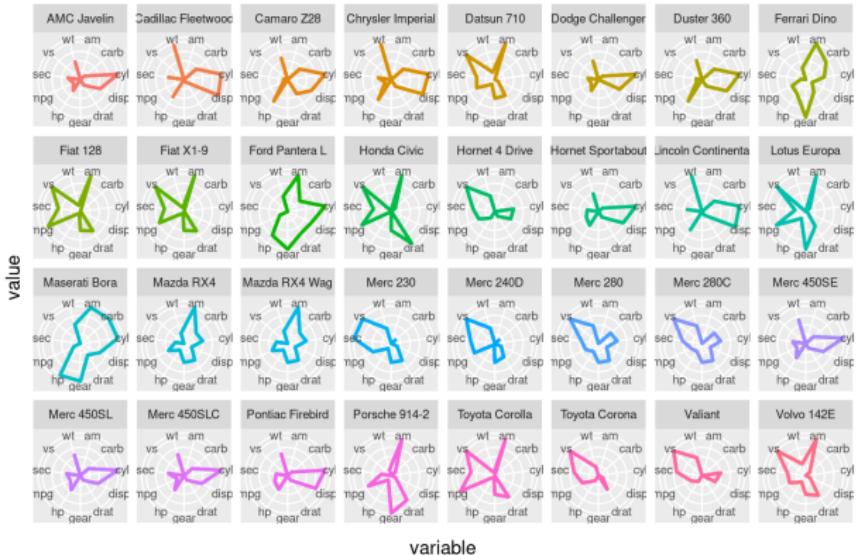
Classical Graphs



- Clever ideas to visualize groups.
- Example of small multiples

HighD - Parallel Coordinates / Radar Plot

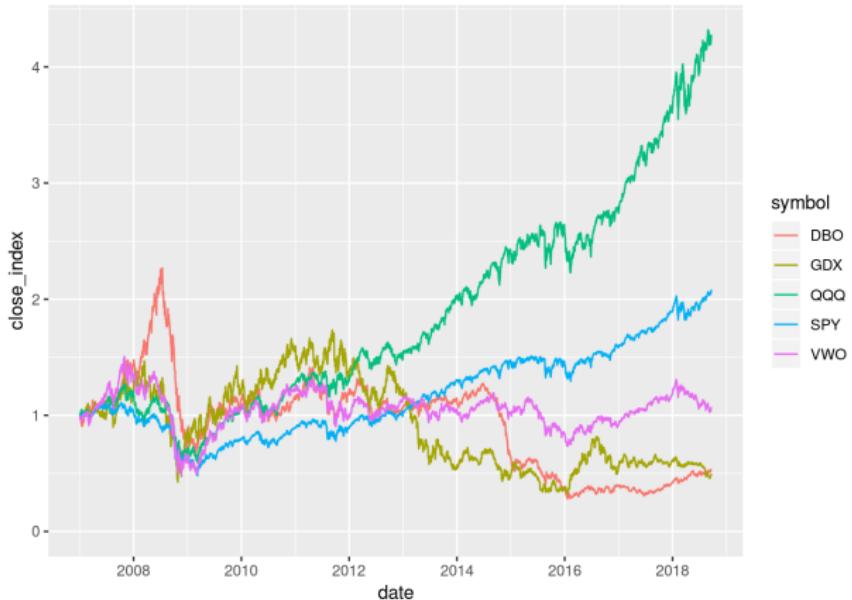
Classical Graphs



- Clever ideas to visualize groups.
- Example of small multiples

Evolution - Time series

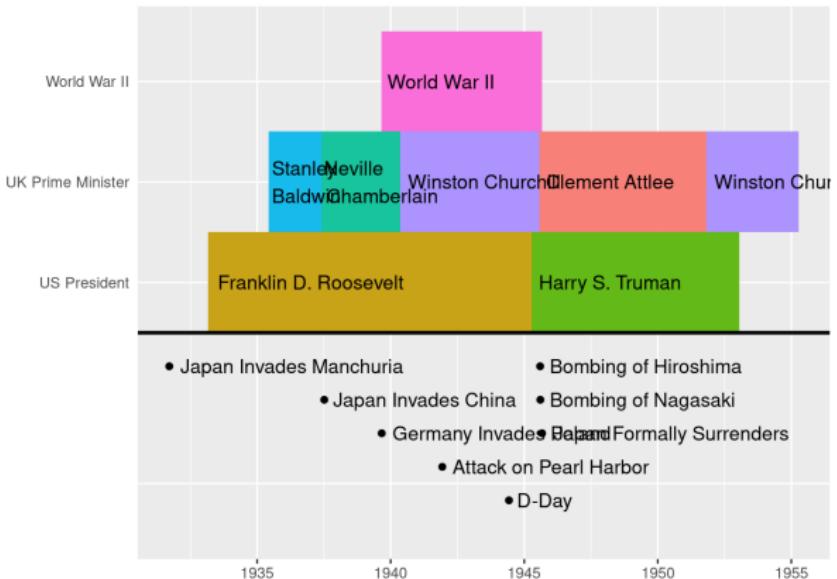
Classical Graphs



- Order makes lines pertinent...
- Columns can also be used.

Evolution - Timeline

Classical Graphs



- Is this really a plot?

Outline

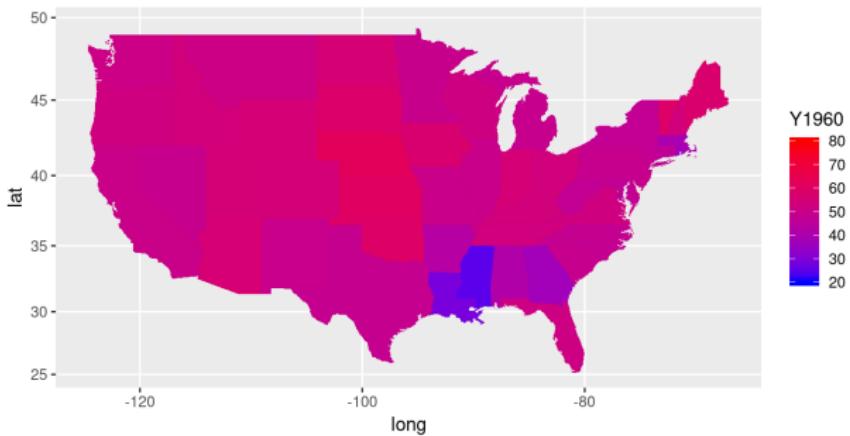
Classical Graphs



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs**
 - Univariate
 - Multivariate
 - Maps**
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Map - Choroplets

Classical Graphs



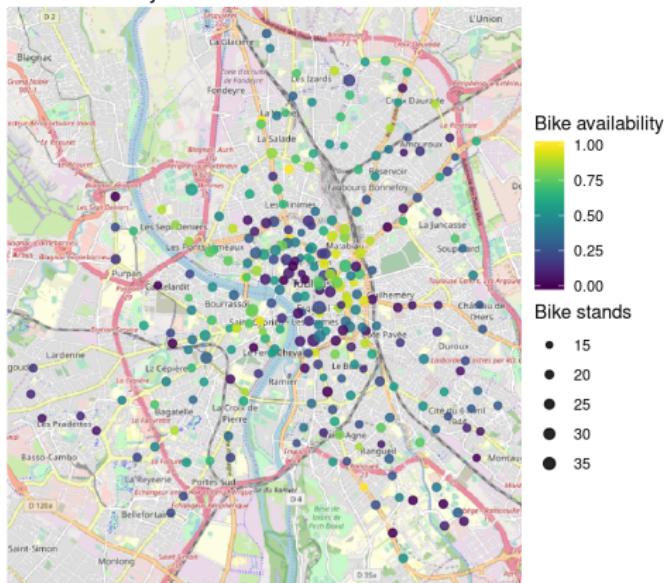
- Strong visual impact!

Map - Symbols

Classical Graphs



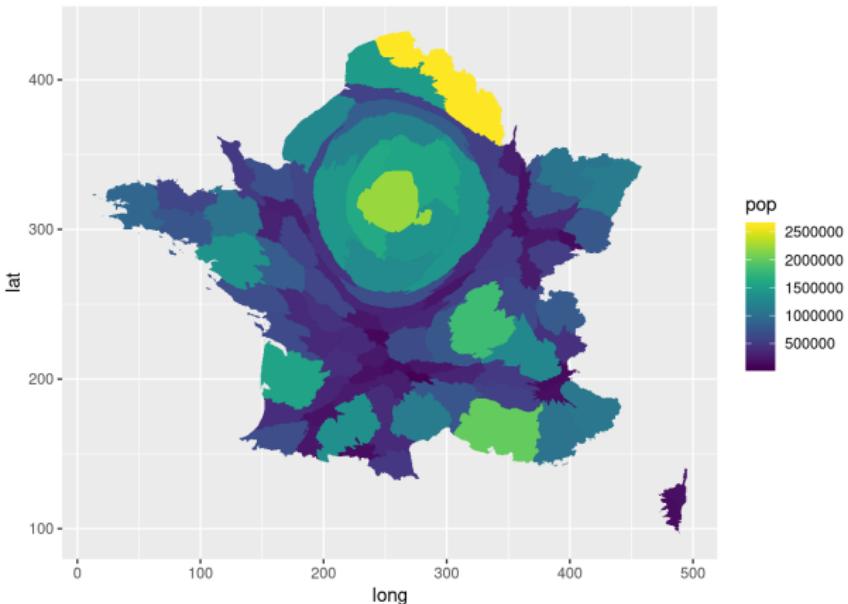
Bike availability



- Same ideas than decoration
- Could be extended to quite complex decorations...

Map - Cartograms

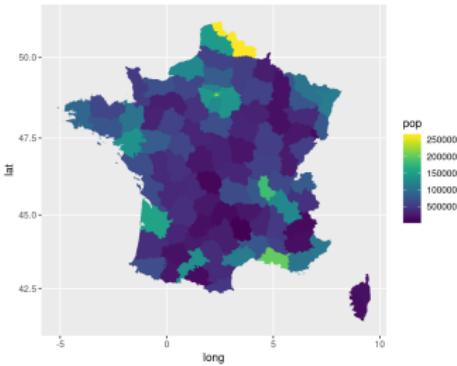
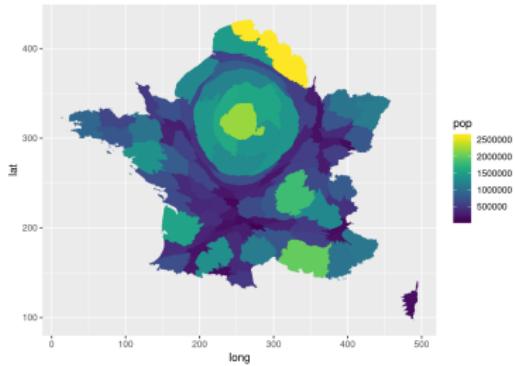
Classical Graphs



- Mainly useful when the reference is known.

Map - Cartograms

Classical Graphs



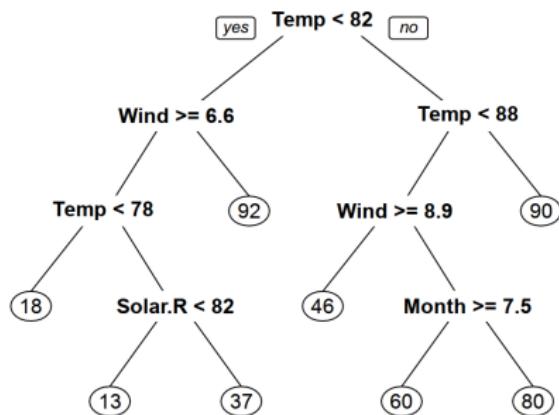
Outline

Classical Graphs



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - **Hierarchy**
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

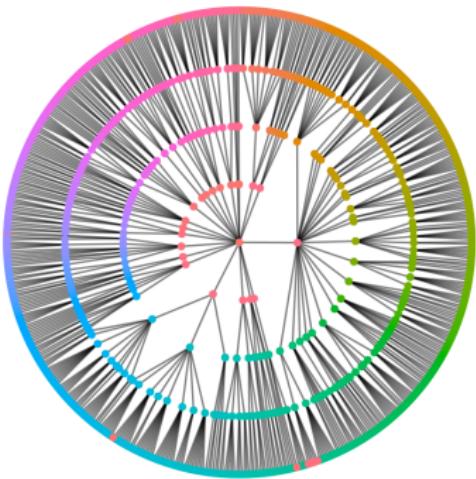
Hierarchy - Trees



- Often used in classification...

Hierarchy - Tree Graph

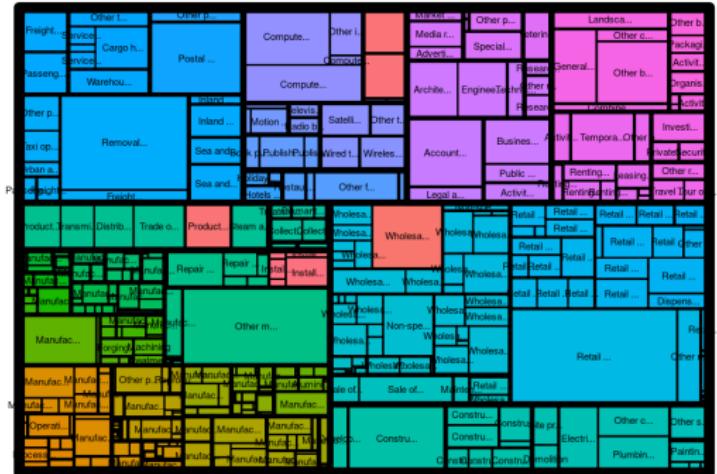
Classical Graphs



- Polar variant.

Hierarchy - Tree Map

Classical Graphs



Outline

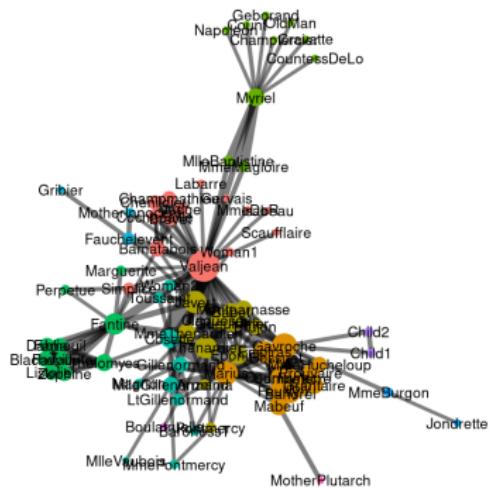
Classical Graphs



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Network - Planar Layout

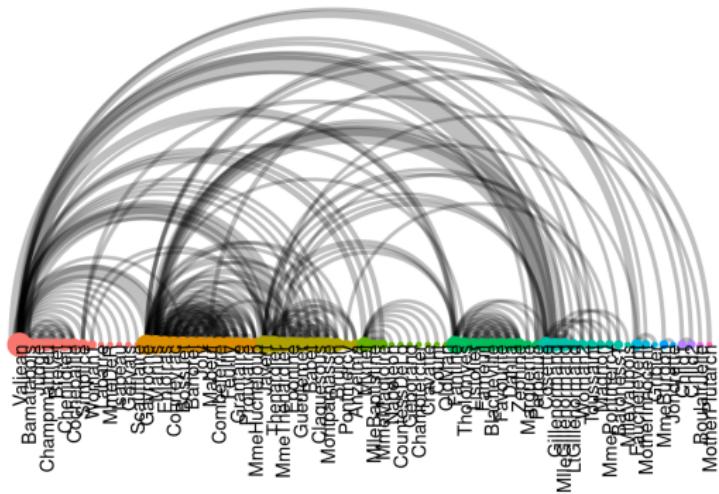
Classical Graphs



- Many possible layouts.

Network - Arc Diagram

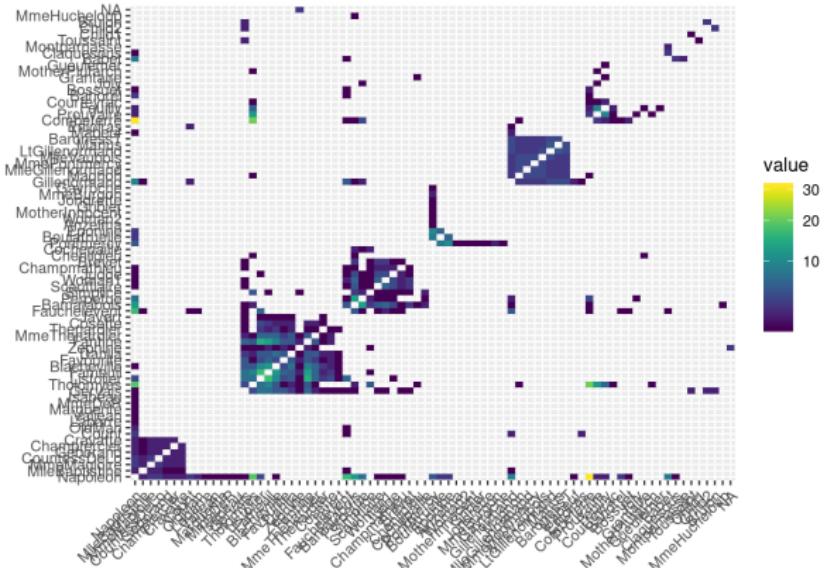
Classical Graphs



- Very different layout...

Network - Matrix View

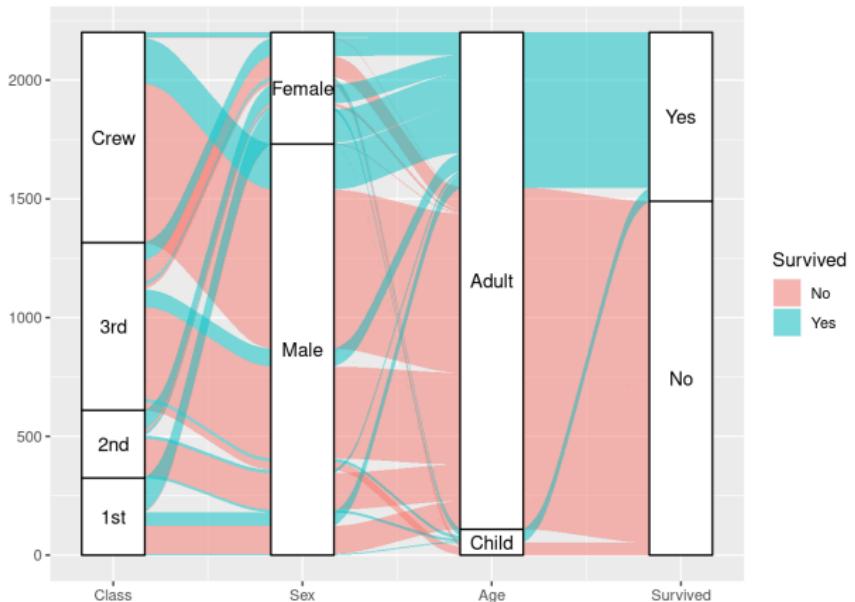
Classical Graphs



- Adjacency matrix visualization.

Network - Flow

Classical Graphs



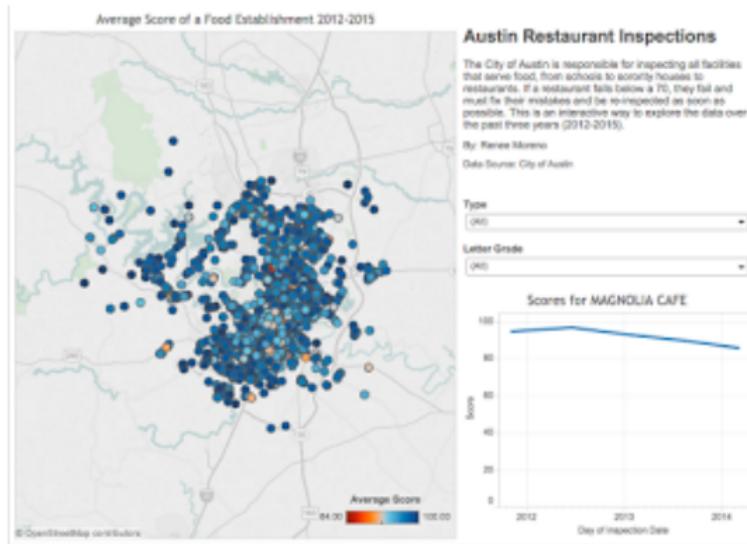
- Vertice oriented visualization.

Outline

- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 **Interactivity and Dynamic Display**
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

Interactivity And Dynamic Display

Interactivity and
Dynamic Display



From static to dynamic

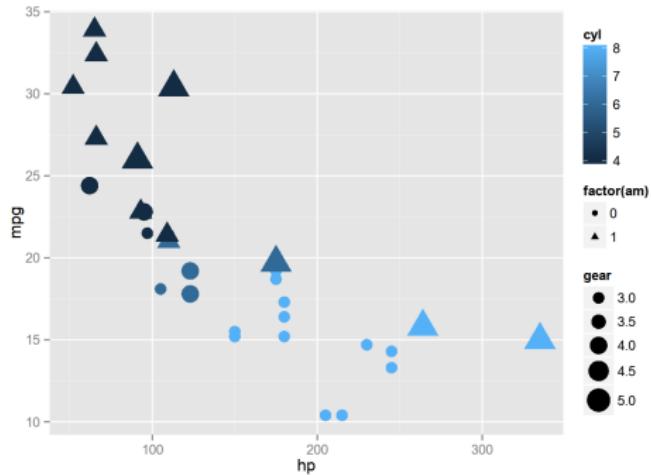
- Technology evolution: from paper to screen/mouse.
- Two directions:
 - Animation: use of time.
 - Interactivity: user interaction.

Animation

- Adapted to 1D mapping...
- Easily deployed (movie or animated picture)

Interactivity

Interactivity and
Dynamic Display

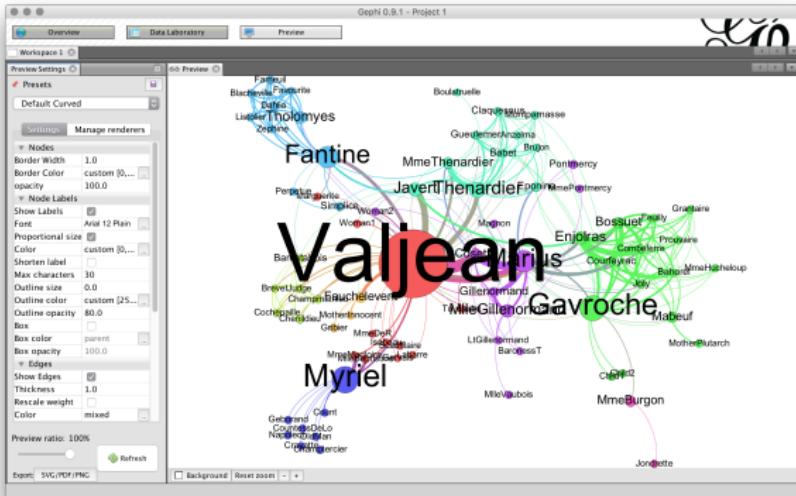


Interactivity: true dynamism

- More requirement.
- Solutions:
 - Standalone app (Gephi, Tulip...),
 - Javascript visualization libraries (d3.js, bokeh, plotly...)
 - Client/server infrastructure (Shiny, flask...)

Standalone App

Interactivity and Dynamic Display

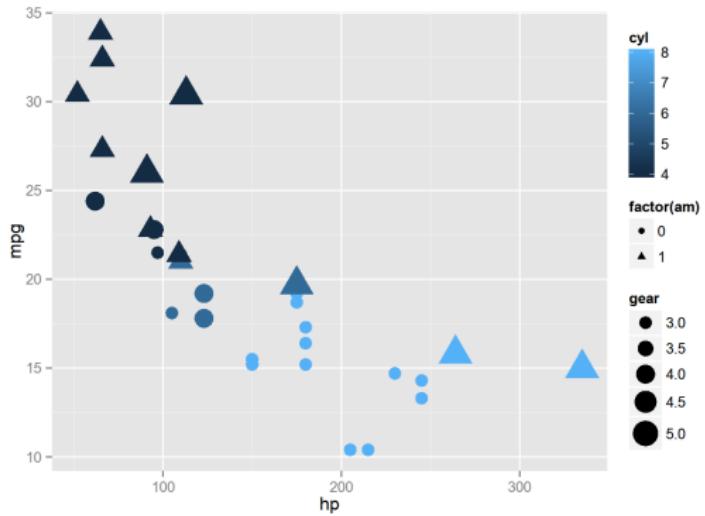


Dedicated software

- Gephi/Tulip for graphs
- Tableau
- ...

Javascript Based Interaction

Interactivity and
Dynamic Display

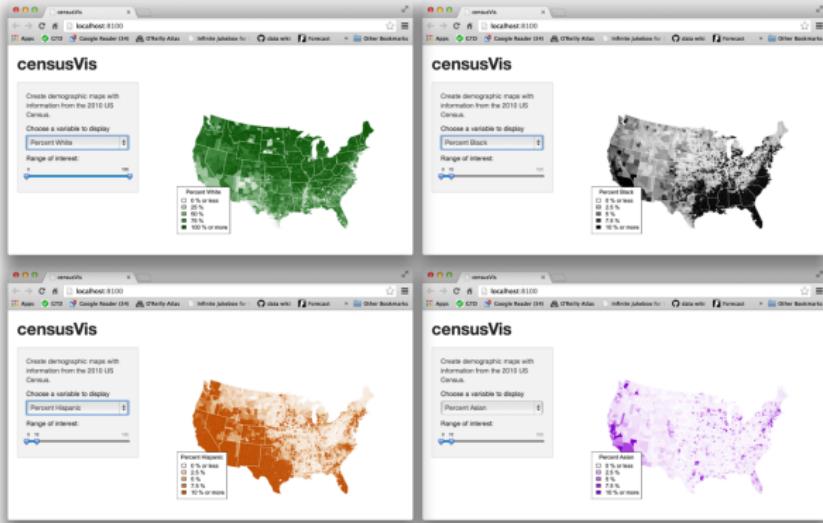


Javascript based frameworks

- Local (lightweight) computation.
- Examples:
 - Tooltip, Zoom, Brushing...
 - Linked panels...

Client/Server

Interactivity and
Dynamic Display



Client/Server Approach

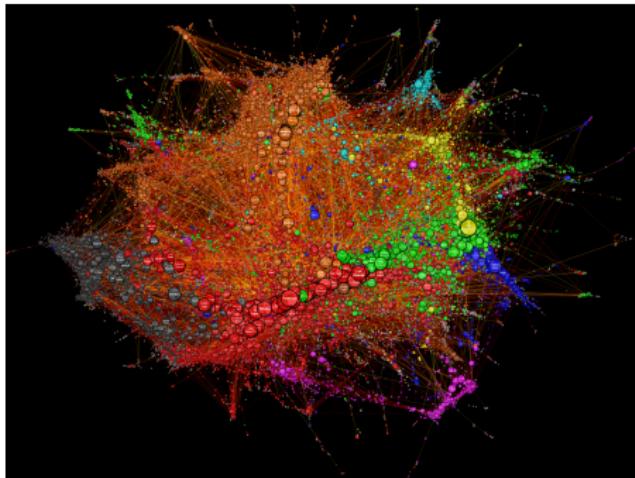
- Shiny, Flask...
- Visualization on the client.
- Computation in the server.

Outline

Big Data



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data**
- 9 References
- 10 Miscellaneous
 - Visualization Principle

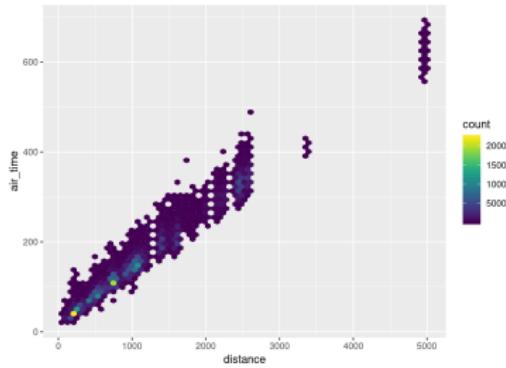
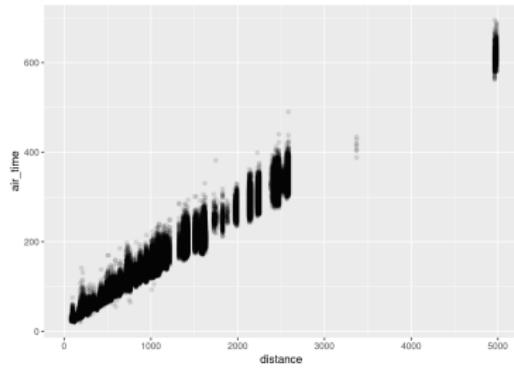


More data points than pixels!

- Even if the processing possible, it is almost impossible to visualize faithfully the data!
- Summarization/selection required:
 - Grouping by categories or binning,
 - Small multiples,
 - Interactive selection.

Binning

Big Data



- Binning \sim 2D histogram

Outline

References



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 Miscellaneous
 - Visualization Principle

References

References



L. Wilkinson.
The Grammar of Graphics.
Springer, 2011



H. Wickham.
ggplot2: Elegant Graphics for Data Analysis.
Springer, 2016



A. Cairo.
The Functional Art: An introduction to information graphics and visualization.
New Riders, 2012



A. Cairo.
The Truthful Art: data, charts and maps for communication.
New Riders, 2016



T. Munzner.
Visualization Analysis and Design.
CRC Press, 2014



K. Healy.
Data Visualization: A Practical Introduction.
Princeton University Press, 2018

Outline

Miscellaneous



- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 **Miscellaneous**
 - **Visualization Principle**

Outline

Miscellaneous

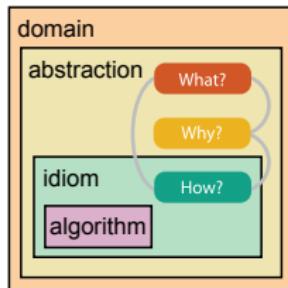


- 1 Introduction
- 2 Table Wrangling
- 3 Visualization
 - Introduction
 - Bad Examples
- 4 Historical Milestones
- 5 Principles
- 6 Classical Graphs
 - Univariate
 - Multivariate
 - Maps
 - Hierarchy
 - Networks
- 7 Interactivity and Dynamic Display
- 8 Big Data
- 9 References
- 10 **Miscellaneous**
 - Visualization Principle

Nested model: Four levels of vis design

- *domain situation*
 - who are the target users?
- *abstraction*
 - translate from specifics of domain to vocabulary of vis
 - **what** is shown? **data abstraction**
 - **why** is the user looking at it? **task abstraction**
- *idiom*
 - **how** is it shown?
 - **visual encoding** idiom: how to draw
 - **interaction** idiom: how to manipulate
- *algorithm*
 - efficient computation

[A Nested Model of Visualization Design and Validation.
Munzner. *IEEE TVCG* 15(6):921-928, 2009
(Proc. InfoVis 2009).]



[A Multi-Level Typology of Abstract Visualization Tasks
Brehmer and Munzner. *IEEE TVCG* 19(12):2376-2385,
2013 (Proc. InfoVis 2013).]

Threats to validity differ at each level

Domain situation

You misunderstood their needs

Data/task abstraction

You're showing them the wrong thing

Visual encoding/interaction idiom

The way you show it doesn't work

Algorithm

Your code is too slow

What?

Why?

How?

What?

Datasets

④ Data Types

- Items
- Attributes
- Links
- Positions
- Grids

④ Data and Dataset Types

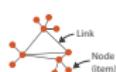
| Tables | Networks & Trees | Fields | Geometry | Clusters, Sets, Lists |
|------------|------------------|-----------|-----------|-----------------------|
| Items | Items (nodes) | Grids | Items | Clusters, Sets, Lists |
| Attributes | Links | Positions | Positions | Items |

④ Dataset Types

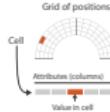
- Tables



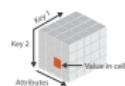
- Networks



- Fields (Continuous)



- Multidimensional Table



- Trees



- Geometry (Spatial)



④ Dataset Availability

- Static



- Dynamic



Attributes

④ Attribute Types

- Categorical



- Ordered

- Ordinal



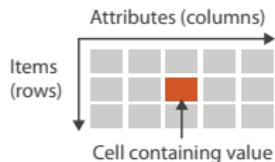
- Quantitative



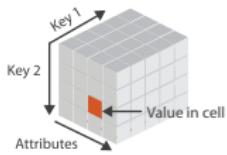
Three major datatypes

Dataset Types

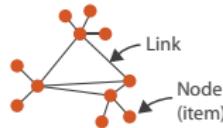
→ Tables



→ Multidimensional Table



→ Networks

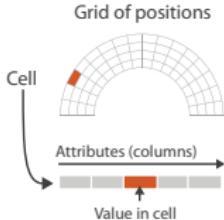


→ Trees



→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)

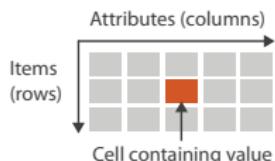


- visualization vs computer graphics
 - geometry is design decision

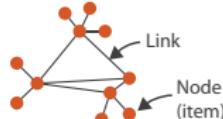
Types: Datasets and data

→ Dataset Types

→ Tables

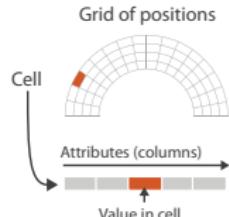


→ Networks



→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)



→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative

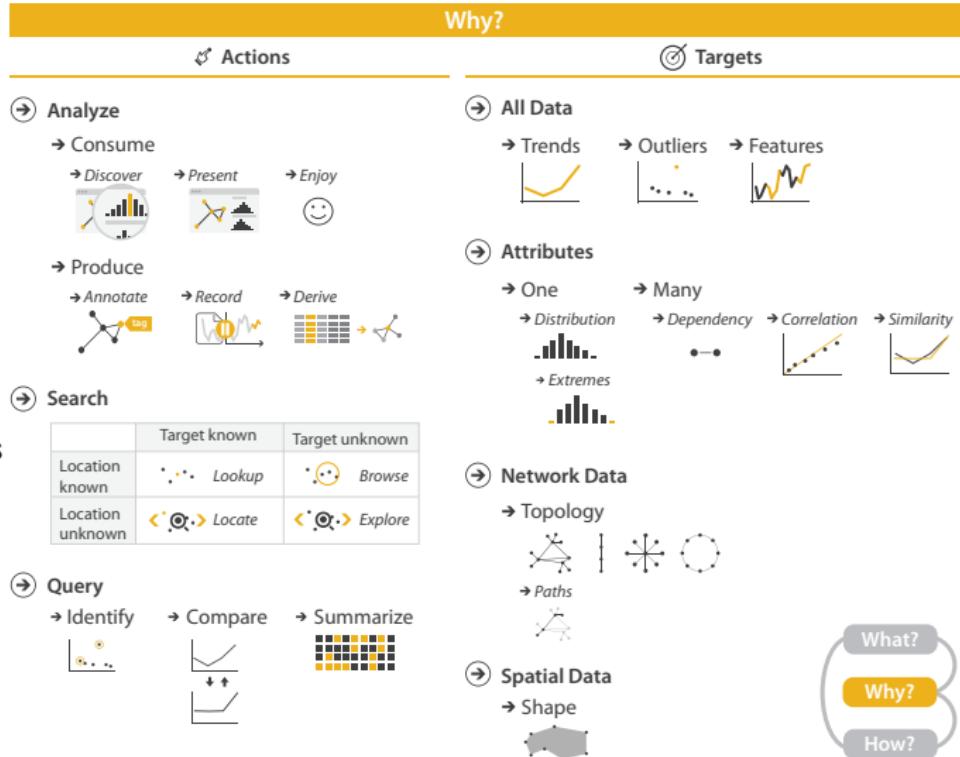


What?

Why?

How?

- {action, target} pairs
 - discover distribution
 - compare trends
 - locate outliers
 - browse topology



What?

Why?

How?

Actions: Analyze, Query

- analyze
 - consume
 - discover vs present
 - aka explore vs explain
 - enjoy
 - aka casual, social
 - produce
 - annotate, record, derive
 - query
 - how much data matters?
 - one, some, all
 - independent choices

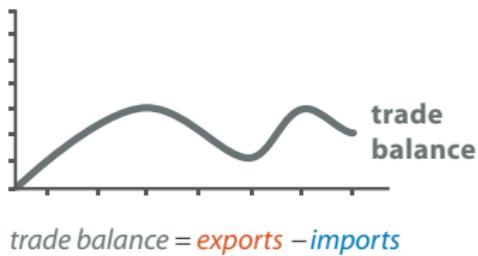


Derive: Crucial Design Choice

- don't just draw what you're given!
 - decide what the right thing to show is
 - create it with a series of transformations from the original dataset
 - draw that
- one of the four major strategies for handling complexity



Original Data



$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived Data

Targets

→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One

→ Distribution



→ Extremes



→ Many

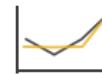
→ Dependency



→ Correlation



→ Similarity



→ Network Data

→ Topology



→ Paths

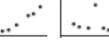


→ Spatial Data

→ Shape



How?

| Encode | Manipulate | Facet | Reduce |
|---|--|--|--|
| <p>⊕ Arrange → Express → Separate  </p> <p>⊕ Order → Align  </p> <p>⊕ Use </p> | <p>⊕ Map from categorical and ordered attributes → Color → Hue → Saturation → Luminance    → Size, Angle, Curvature, ...    → Shape + ● ■ ▲  → Motion <i>Direction, Rate, Frequency, ...</i> </p> | <p>⊕ Change  </p> <p>⊕ Select </p> <p>⊕ Navigate </p> | <p>⊕ Juxtapose </p> <p>⊕ Partition </p> <p>⊕ Superimpose </p> <p>⊕ Filter  → Aggregate  → Embed </p> |

What?

Why?

How?

How to encode: Arrange space, map channels

Encode

④ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



④ Map

from categorical and ordered attributes

→ Color



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

Direction, Rate, Frequency, ...



Definitions: Marks and channels

- marks

- geometric primitives

⇒ Points



⇒ Lines



⇒ Areas



- channels

- control appearance of marks

⇒ Position

→ Horizontal



→ Vertical



→ Both



⇒ Color



⇒ Shape



⇒ Tilt



⇒ Size

→ Length



→ Area

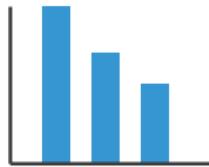


→ Volume



Encoding visually with marks and channels

- analyze idiom structure
 - as combination of marks and channels



1:
vertical position

mark: line



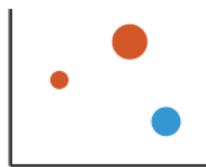
2:
vertical position
horizontal position

mark: point



3:
vertical position
horizontal position
color hue

mark: point



4:
vertical position
horizontal position
color hue
size (area)

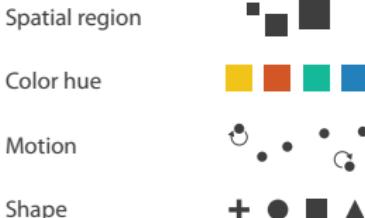
mark: point

Channels: Rankings

④ Magnitude Channels: Ordered Attributes



⑤ Identity Channels: Categorical Attributes



- **expressiveness principle**
 - match channel and data characteristics
- **effectiveness principle**
 - encode most important attributes with highest ranked channels

Four strategies to handle complexity

→ Derive



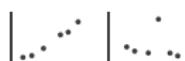
Manipulate

④ Change



Facet

④ Juxtapose



Reduce

④ Filter



- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view

④ Select



④ Partition



④ Aggregate



④ Navigate



④ Superimpose

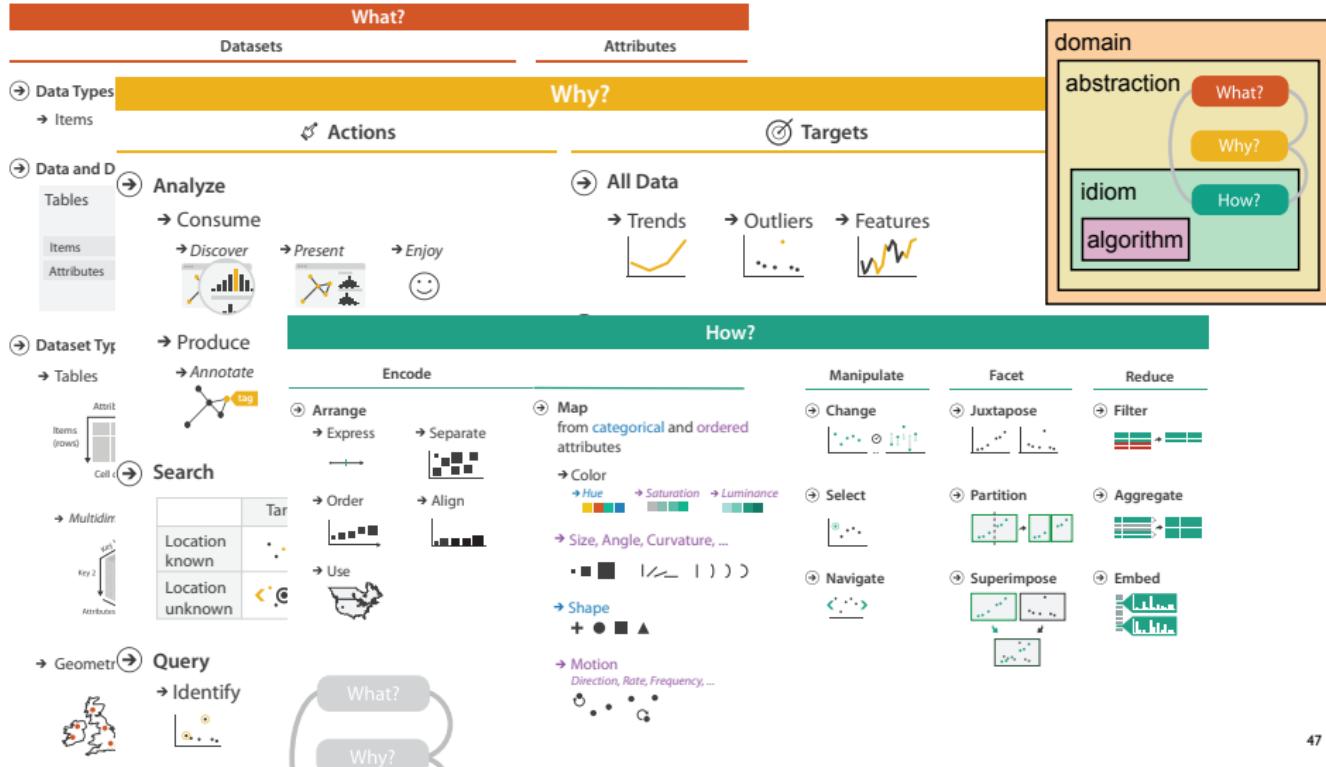


④ Embed

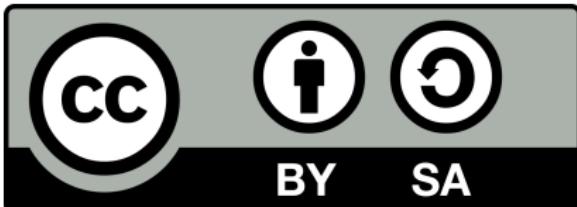


more at:

Visualization Analysis and Design.
Munzner. AK Peters Visualization Series, CRC Press, 2014.



Licence and Contributors



Creative Commons Attribution-ShareAlike (CC BY-SA 4.0)

- You are free to:
 - **Share:** copy and redistribute the material in any medium or format
 - **Adapt:** remix, transform, and build upon the material for any purpose, even commercially.
- Under the following terms:
 - **Attribution:** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
 - **ShareAlike:** If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
 - **No additional restrictions:** You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Contributors

- S. Boucheron, A.K. Fermin, S. Gaiffas, A. Guilloux, Ch. Keribin, E. Le Pennec, E. Matzner, E. Scornet and X Exed team.