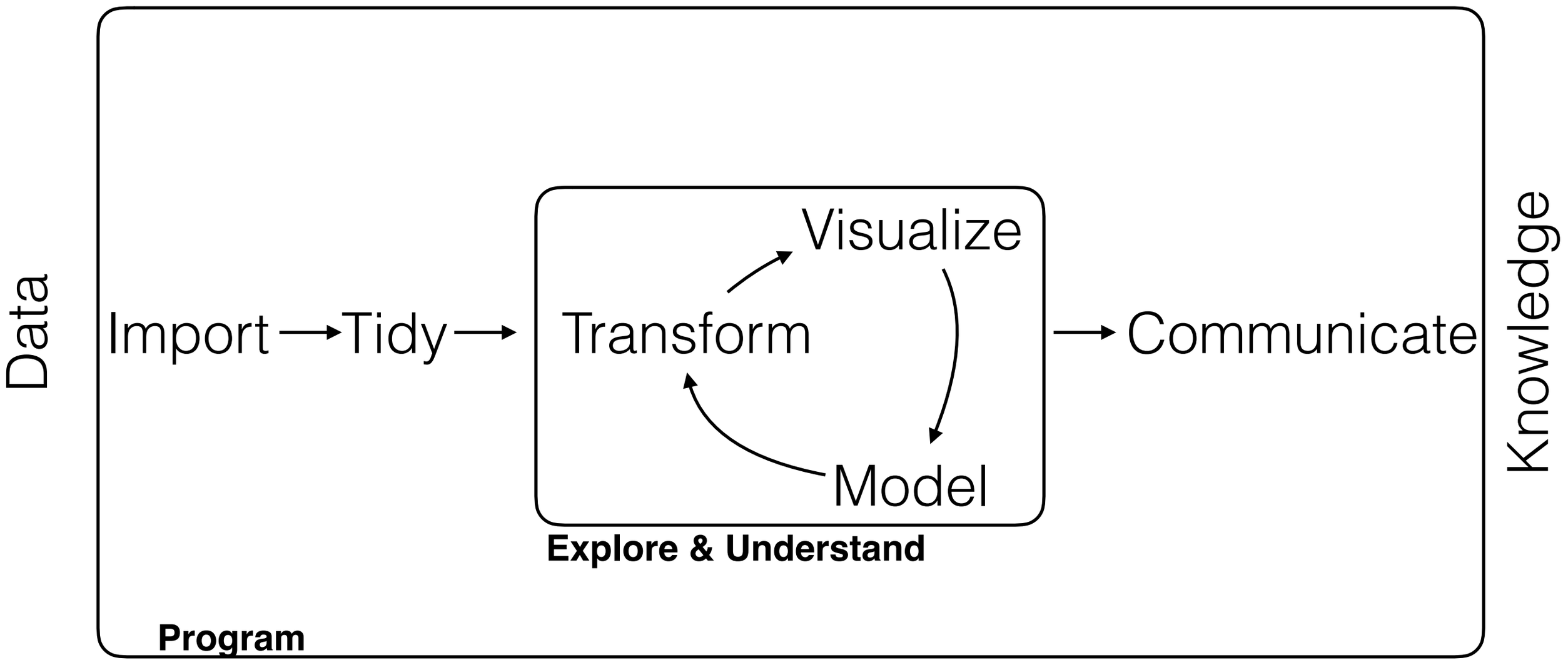
Visualize Data

Strategic and Competitive Intelligence

Filippo Chiarello

filippochiarello.90@gmail.com

What you will learn



Adapt from: "R for Data Science", H. Wickham, 2017

Grammar?

Grammar

Linguistic Definition → the **set of rules** governing how **elements** of phrases interact in any given natural language**.**

Grammar

Linguistic Definition → the **set of rules** governing how **elements** of phrases interact in any given natural language**.**

Generic Definition → ?

Grammar

Linguistic Definition → the **set of rules** governing how **elements** of phrases interact in any given natural language**.**

Generic Definition → the **set of rules** governing how **elements** interacts

Grammar

Linguistic Definition → the **set of rules** governing how **elements** of phrases interact in any given natural language**.**

Generic Definition → the **set of rules** governing how **elements** interacts

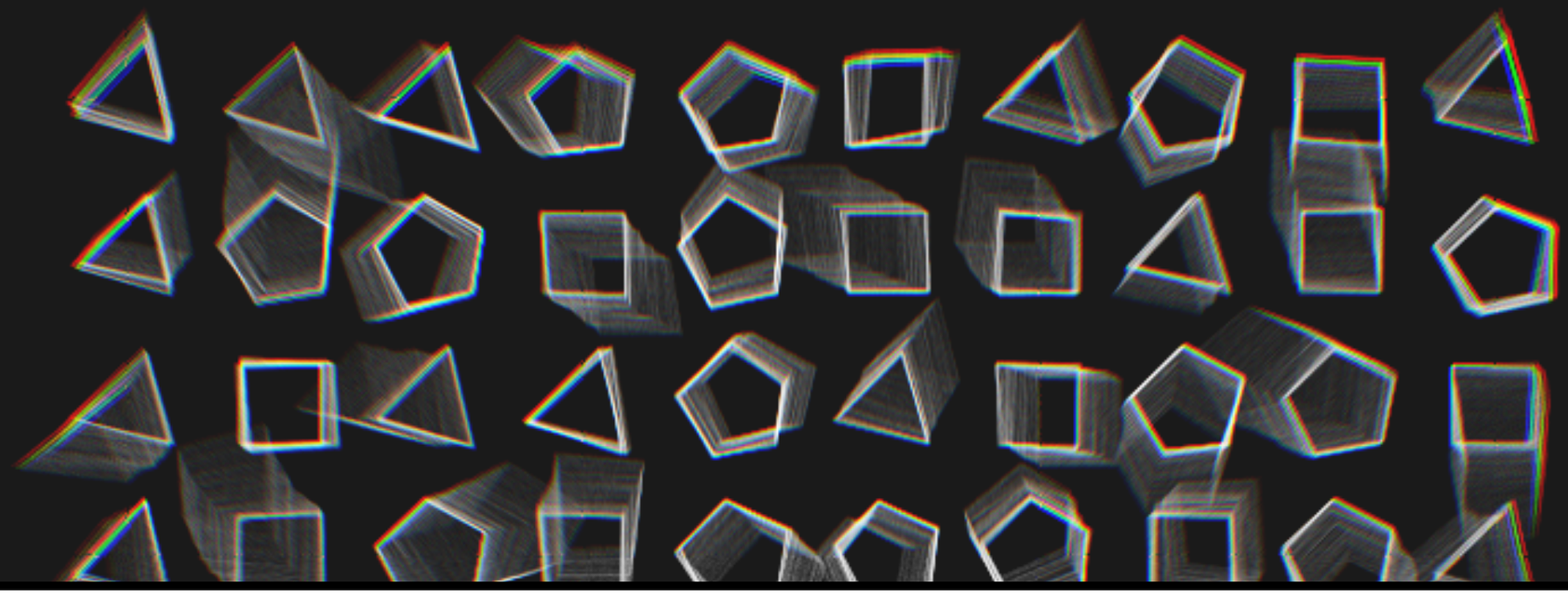
Graph Definition → ?

Grammar

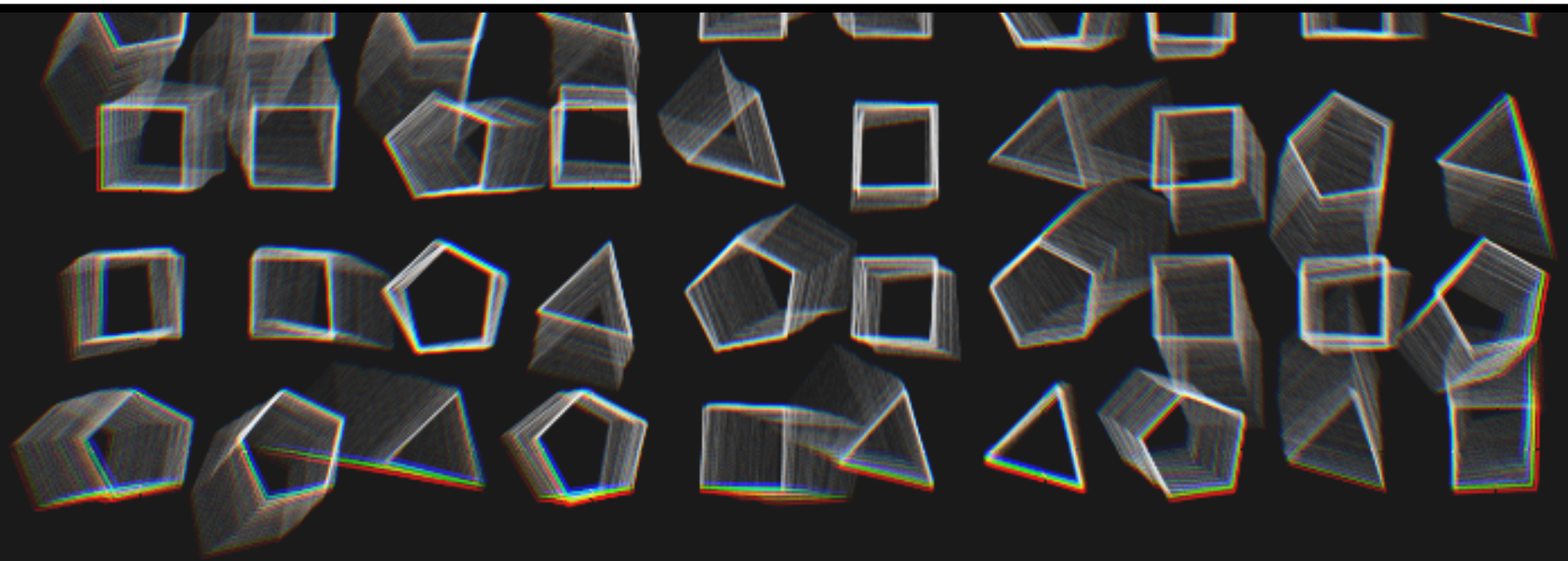
Linguistic Definition → the **set of rules** governing how **elements** of phrases interact in any given natural language**.**

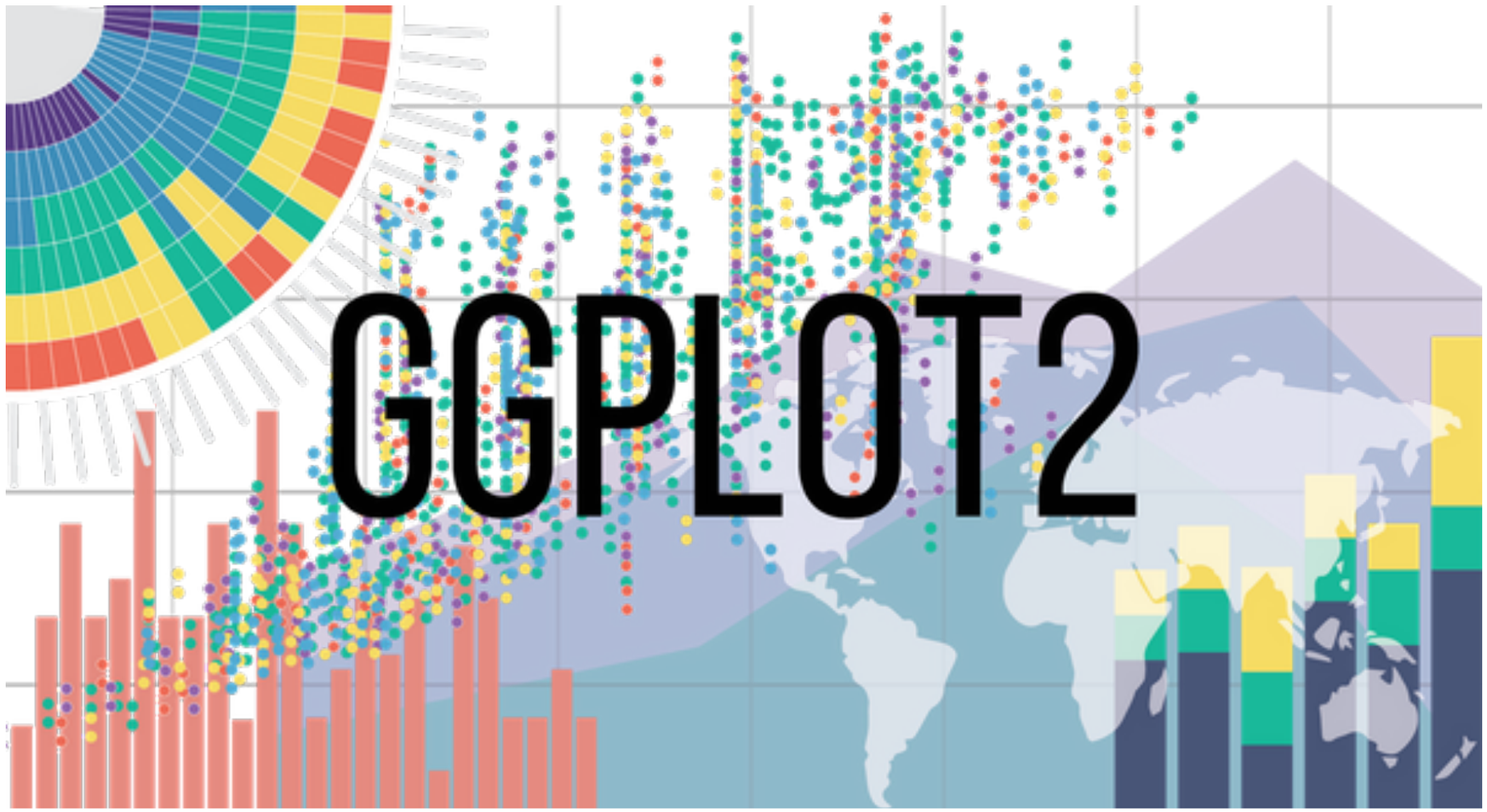
Generic Definition → the **set of rules** governing how **elements** interacts

Graph Definition → the **set of rules** governing how **elements** of graphs interacts in any given graph type

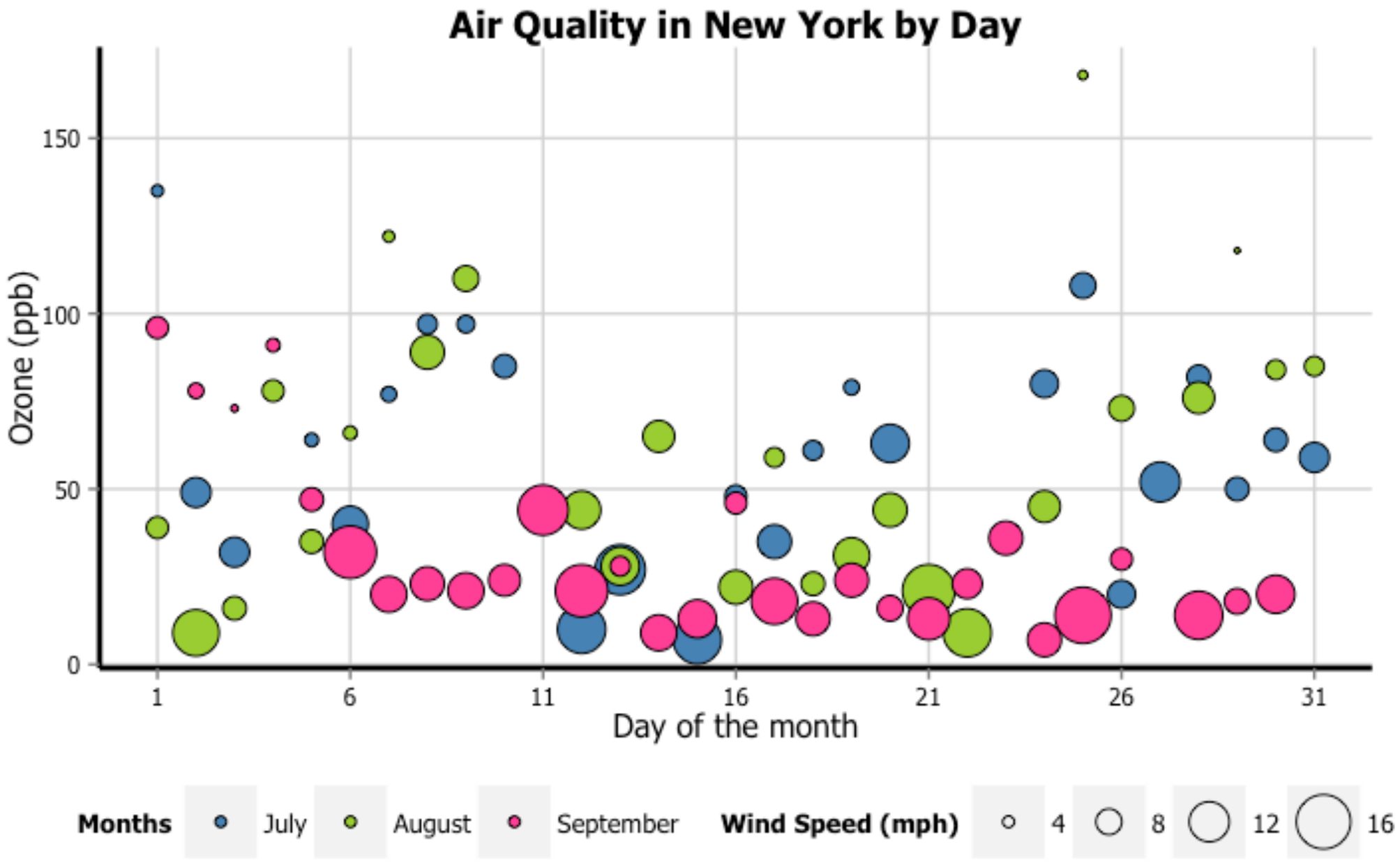


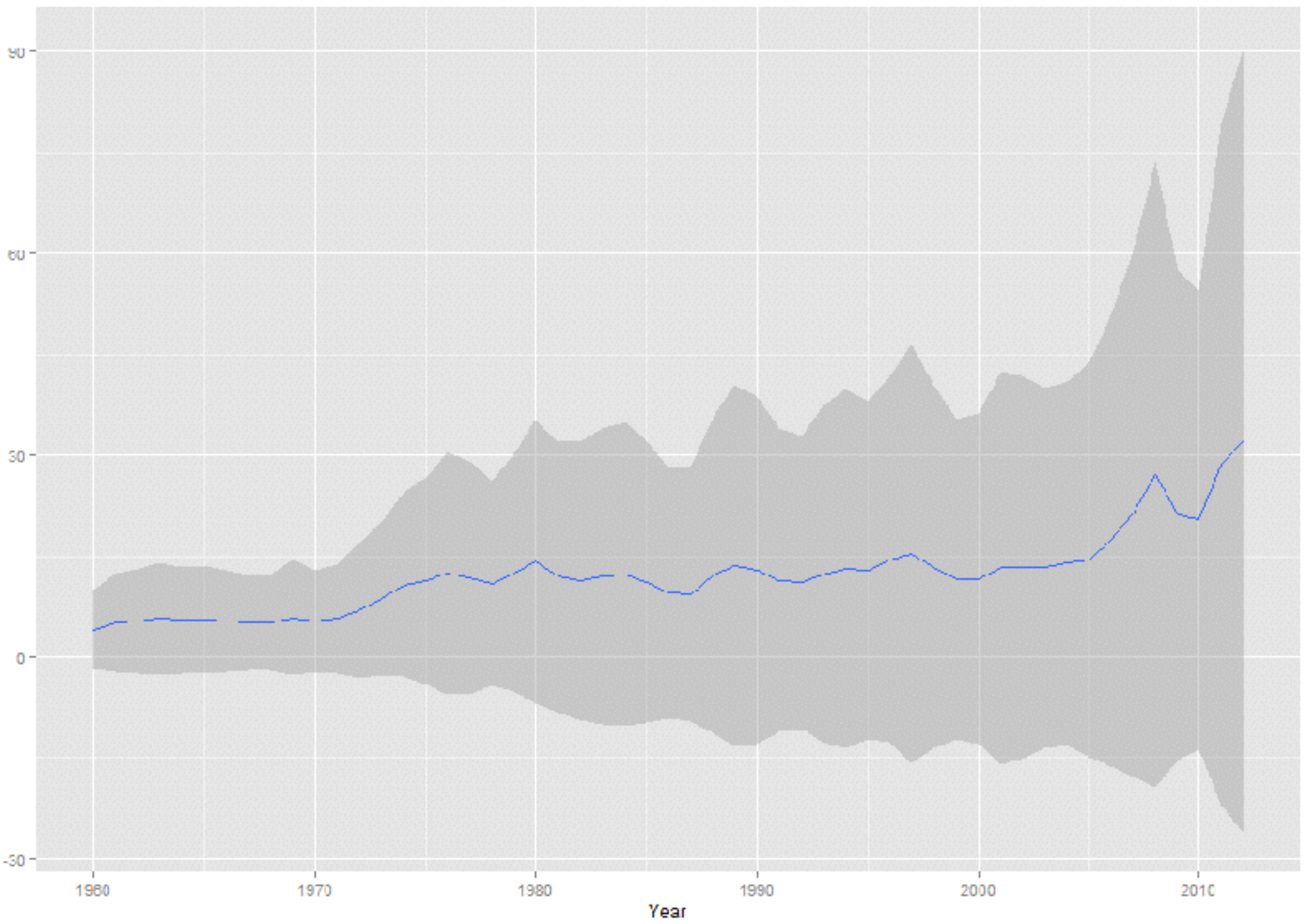
Which are the elements of a graph?

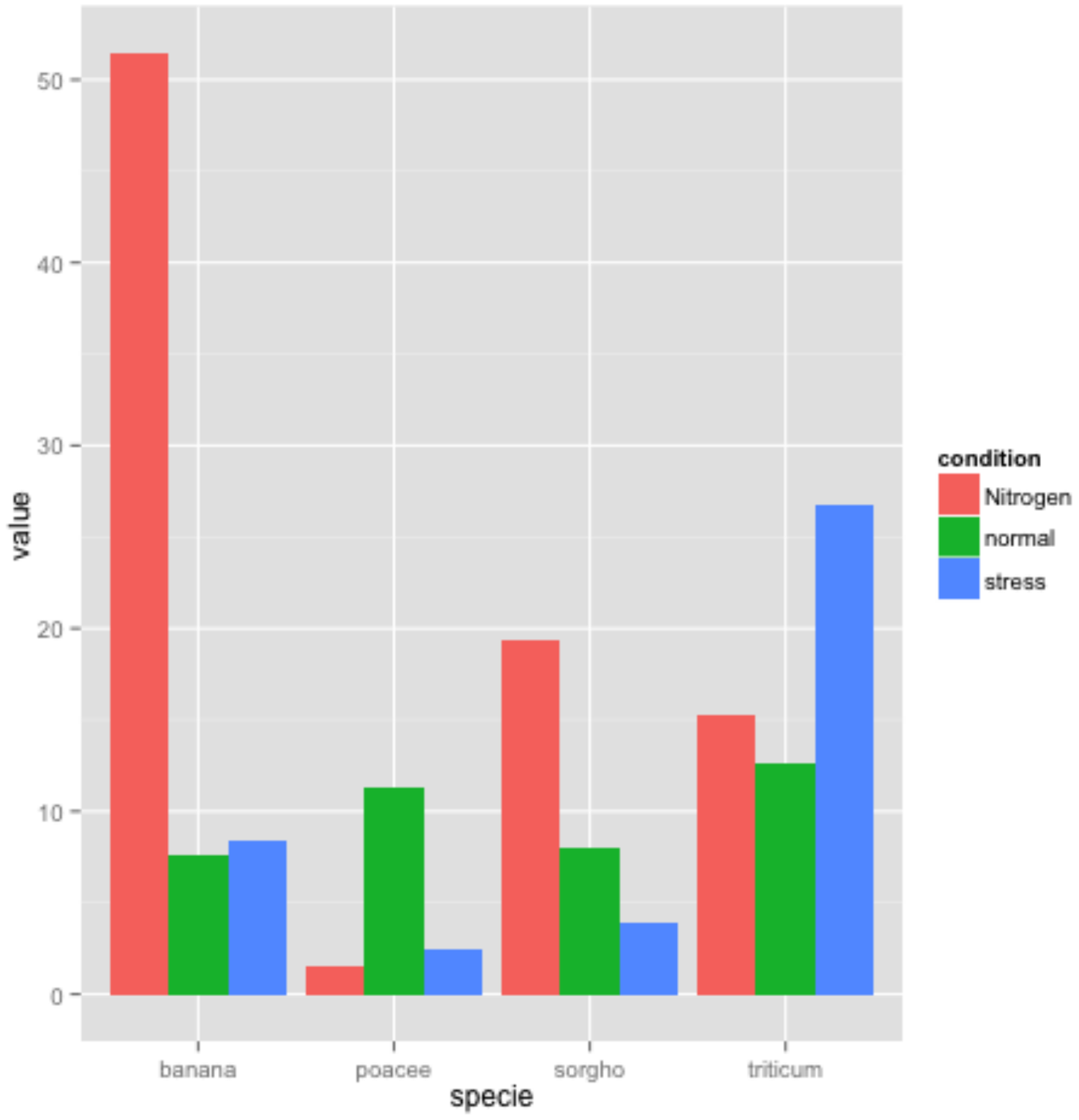


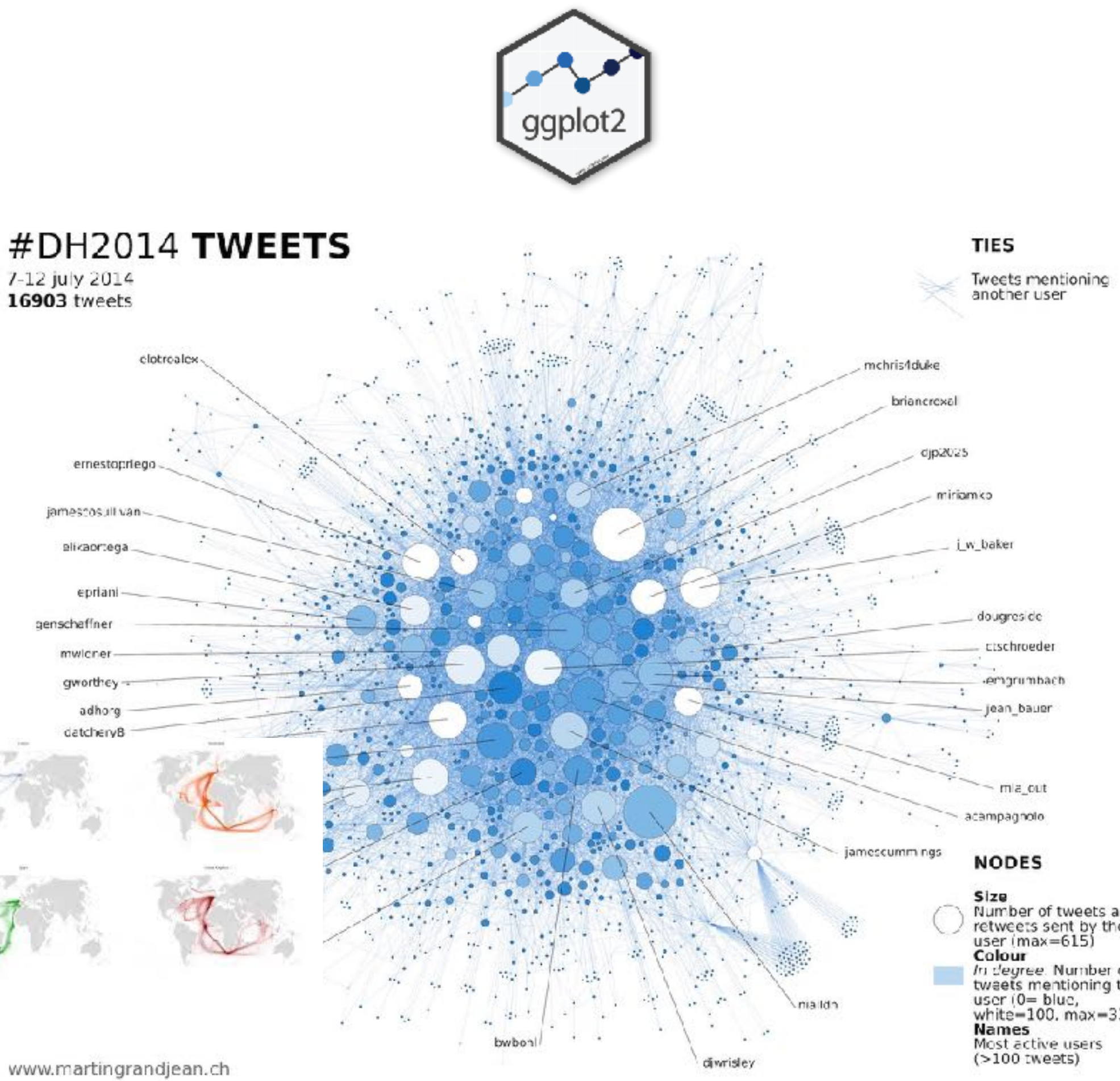


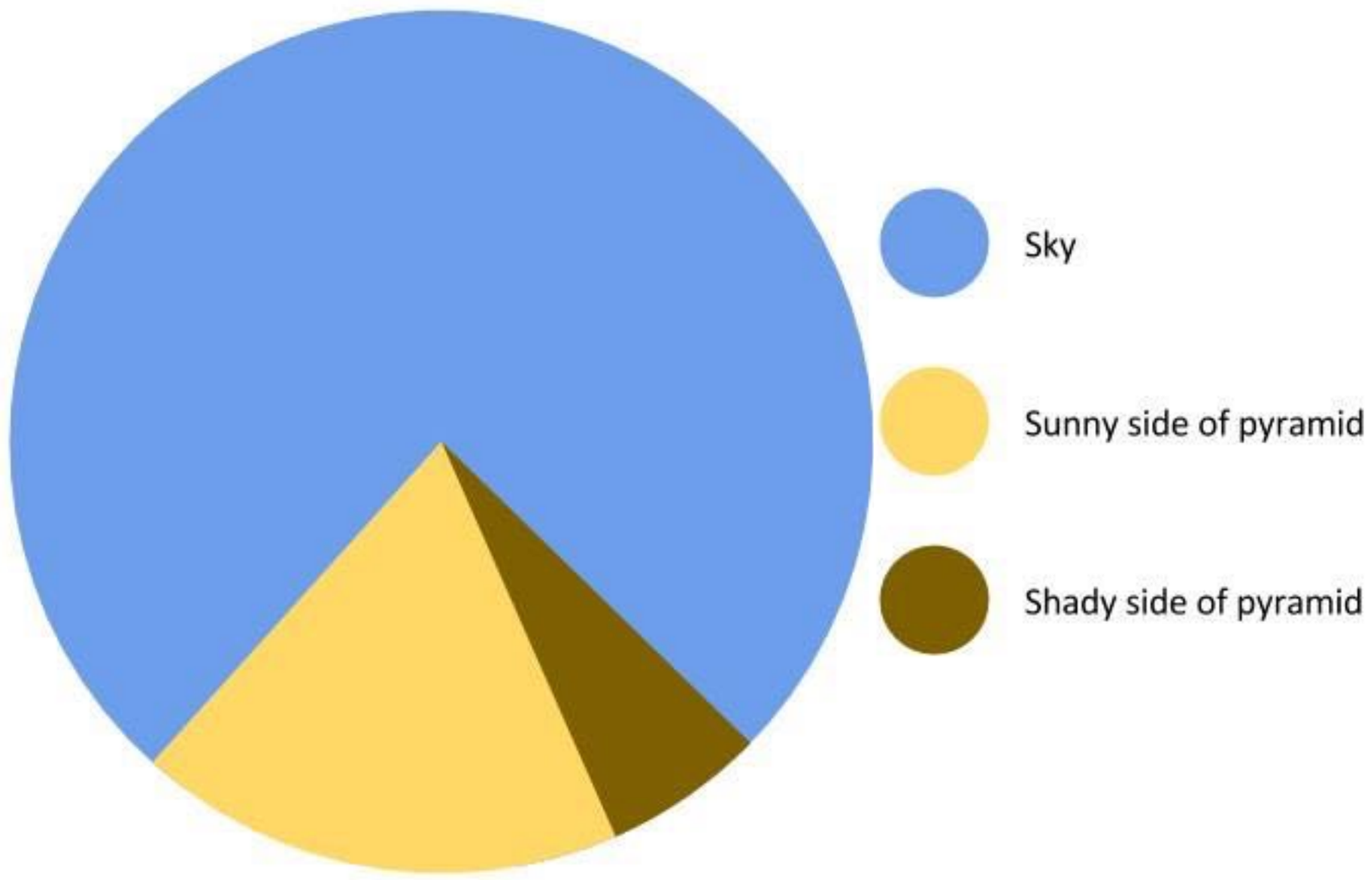
*ggplot2 implements the **grammar of graphics**, a coherent system for describing and building graphs.*

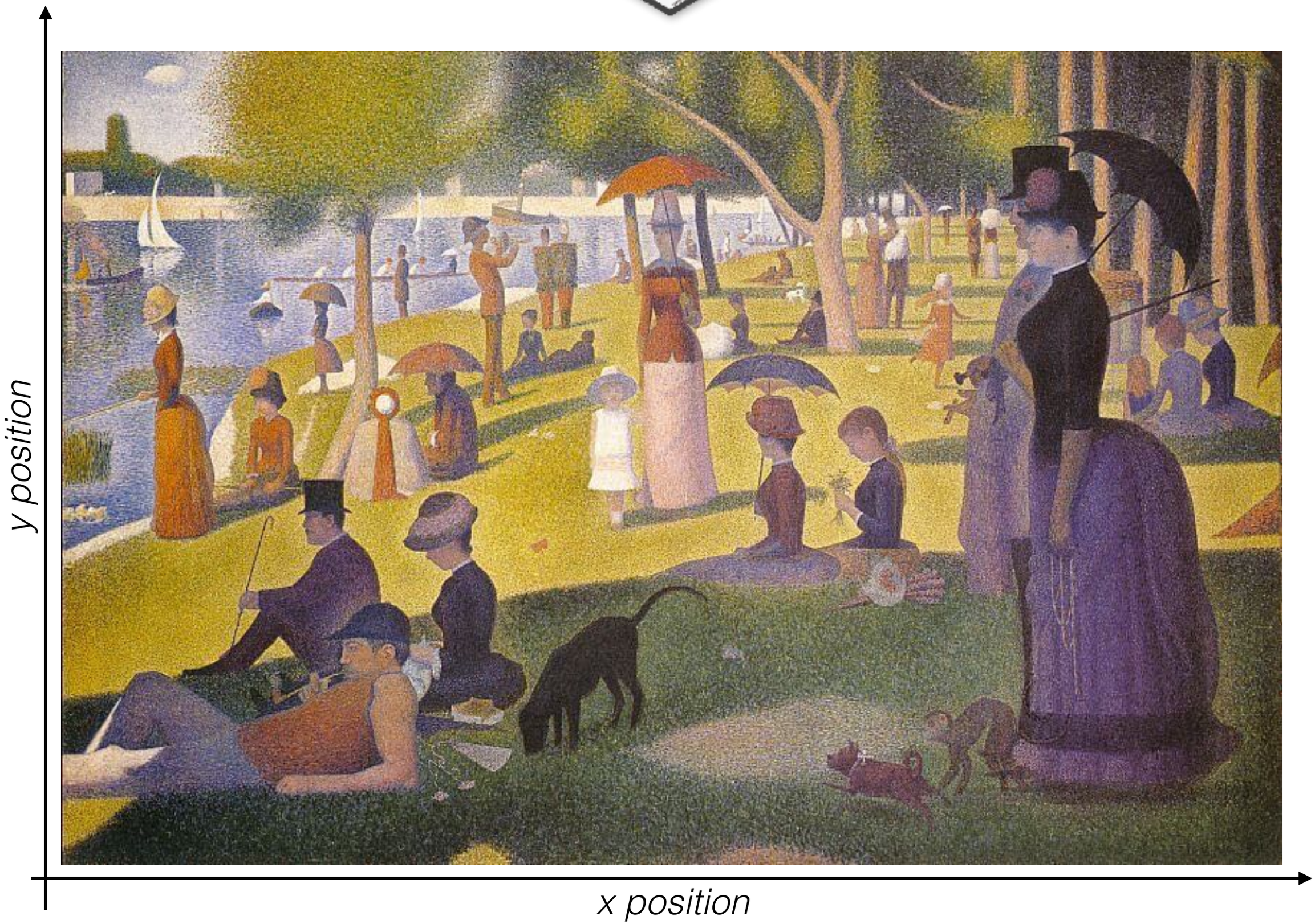

Air Quality in New York by Day
Ozone (ppb)
150
100
50
0
1
6
11
16
21
26
31
Day of the month
Months
July
August
September
Wind Speed (mph)
4
8
12
16

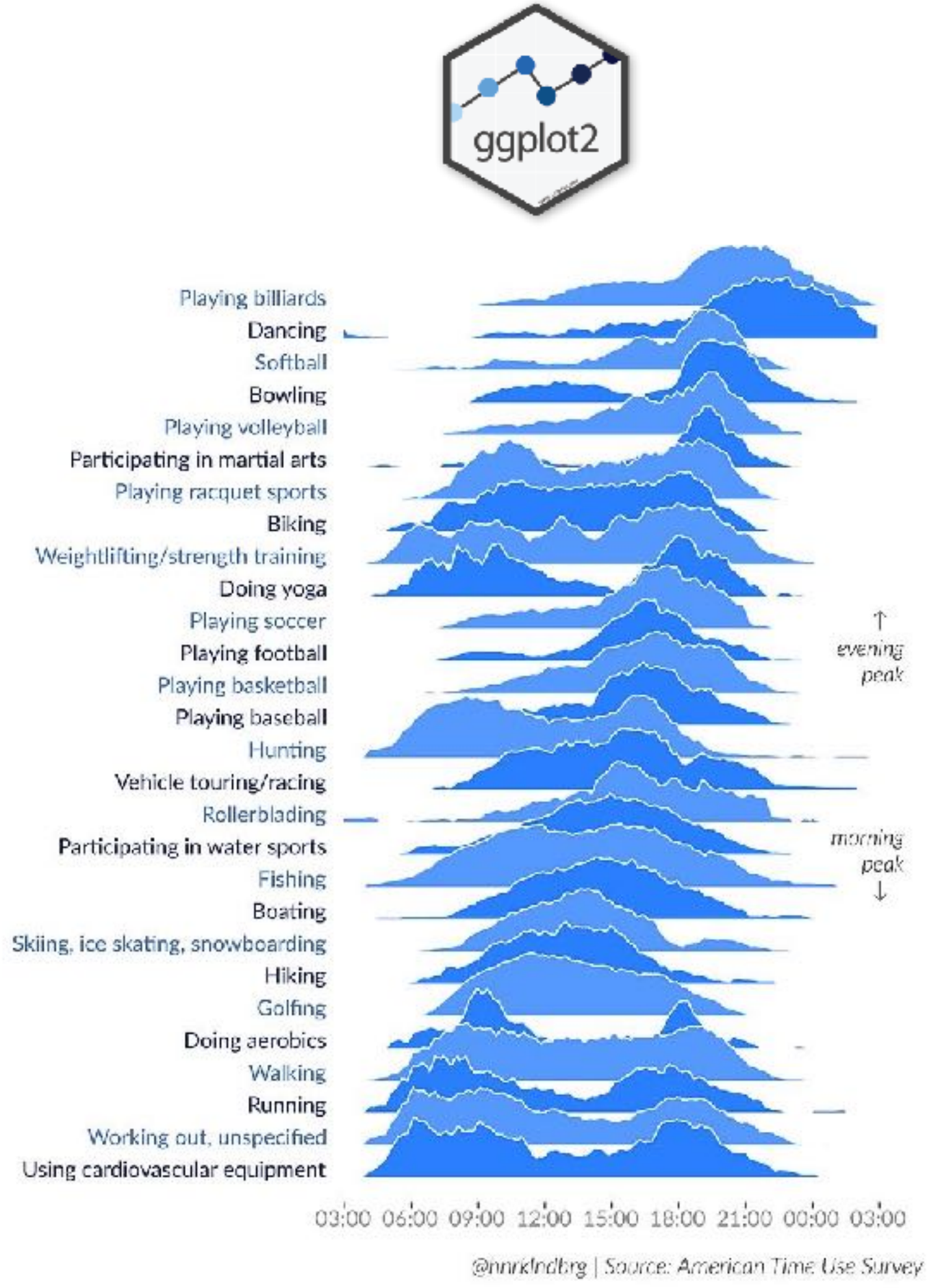

90
60
30
0
-30
value
1960
1970
1980
1990
2000
2010
Year

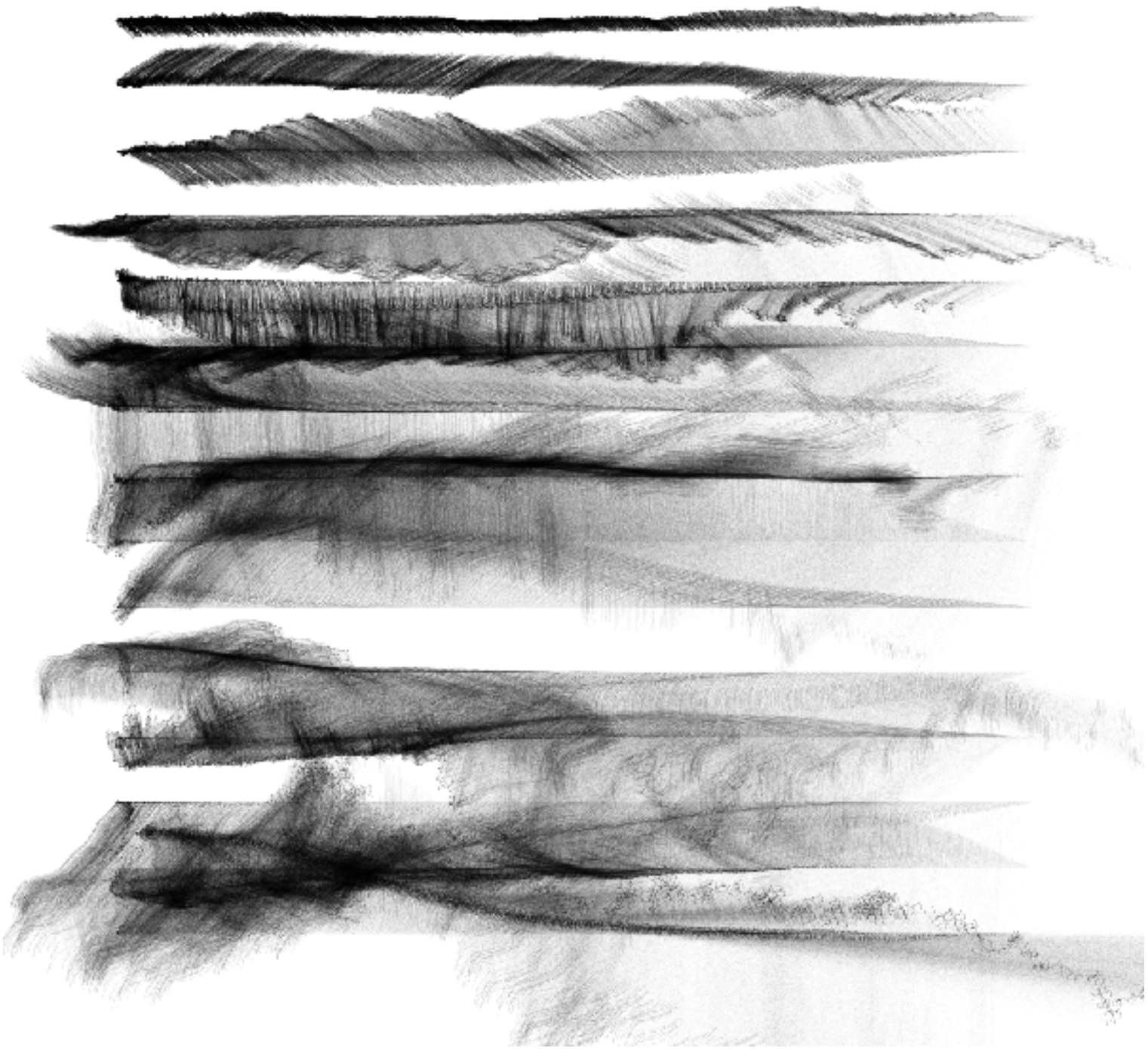

50
40
30
20
10
0
value
banana
poacee
sorgho
triticum
specie
condition
Nitrogen
normal
stress


#DH2014 TWEETS
7-12 july 2014
16903 tweets
TIES
Tweets mentioning another user
clotrealex
ernestoprlego
jamescosullivan
alixaortega
epriani
genscheffner
mwicner
gworthey
adhorg
datcheryB
mchris4duke
briancrexal
cjp2025
miriamko
j_w_baker
dougrosdc
ctschroeder
emgrumbach
jean_bauer
mia_out
acampagnolo
jamescummings
mialdo
bwbond
diwrisley
NODES
Size
Number of tweets and retweets sent by the user (max=615)
Colour
In degree. Number of twaets mentioning the user (0= blue, white=100, max=333)
Names
Most active users (>100 tweets)
www.martingrandjean.ch


Sky
Sunny side of pyramid
Shady side of pyramid


y position
x position


ggplot2
Playing billiards
Dancing
Softball
Bowling
Playing volleyball
Participating in martial arts
Playing racquet sports
Biking
Weightlifting/strength training
Doing yoga
Playing soccer
Playing football
Playing basketball
Playing baseball
Hunting
Vehicle touring/racing
Rollerblading
Participating in water sports
Fishing
Boating
Skiing, ice skating, snowboarding
Hiking
Golfing
Doing aerobics
Walking
Running
Working out, unspecified
Using cardiovascular equipment
↑
evening
peak
morning
peak
↓
03:00 06:00 09:00 12:00 15:00 18:00 21:00 00:00 03:00
@nnrklndbrg | Source: American Time Use Survey



Elements of Graphs

A	B	C	D
2	3	4	a
1	2	1	a
4	5	15	b
9	10	80	b

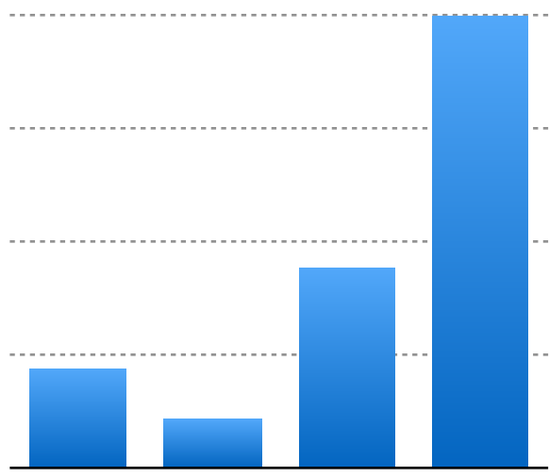
Element: Data

Elements of Graphs

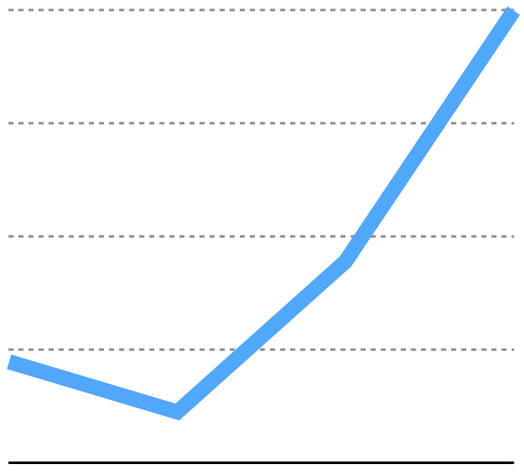
A	C	D
2	4	a
1	1	a
4	15	b
9	80	b

Element: Aesthetics

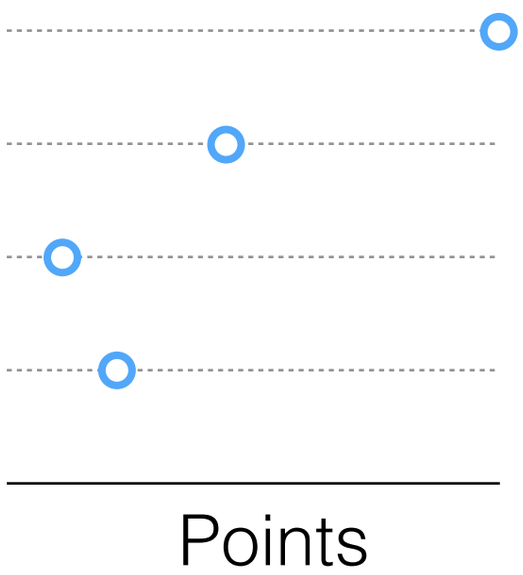
Elements of Graphs



Bars



Lines

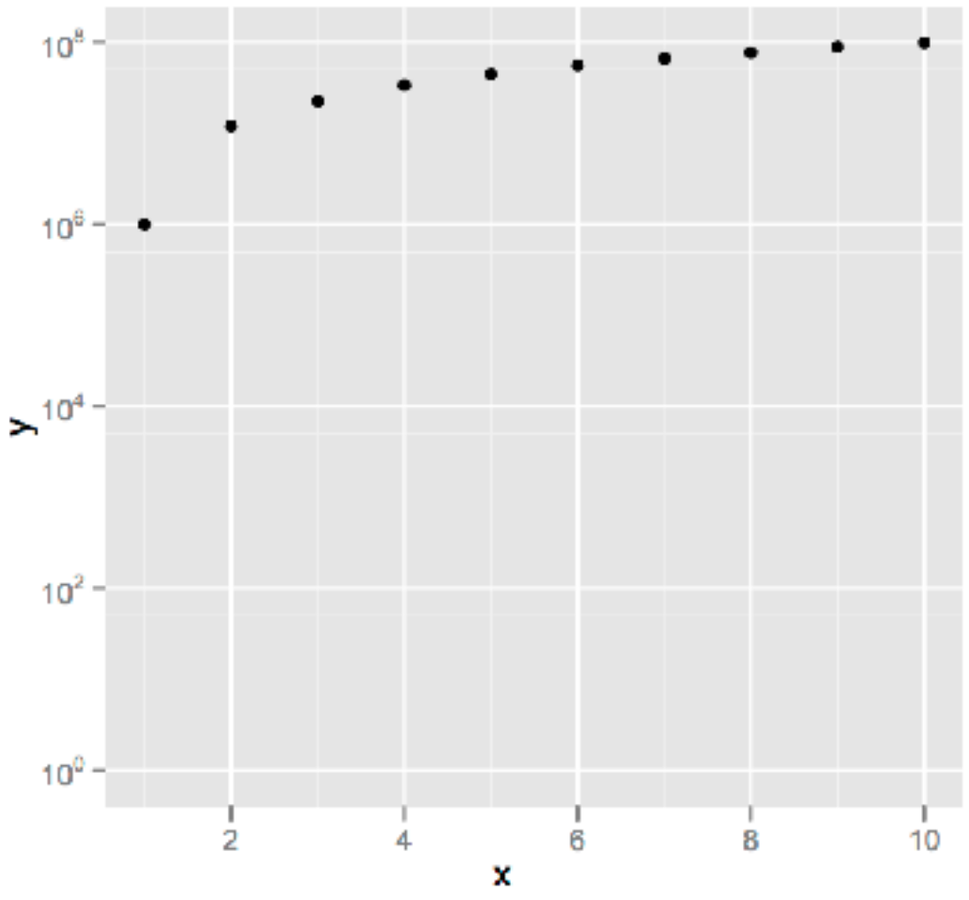


Points

Many others…

Element: geometric object

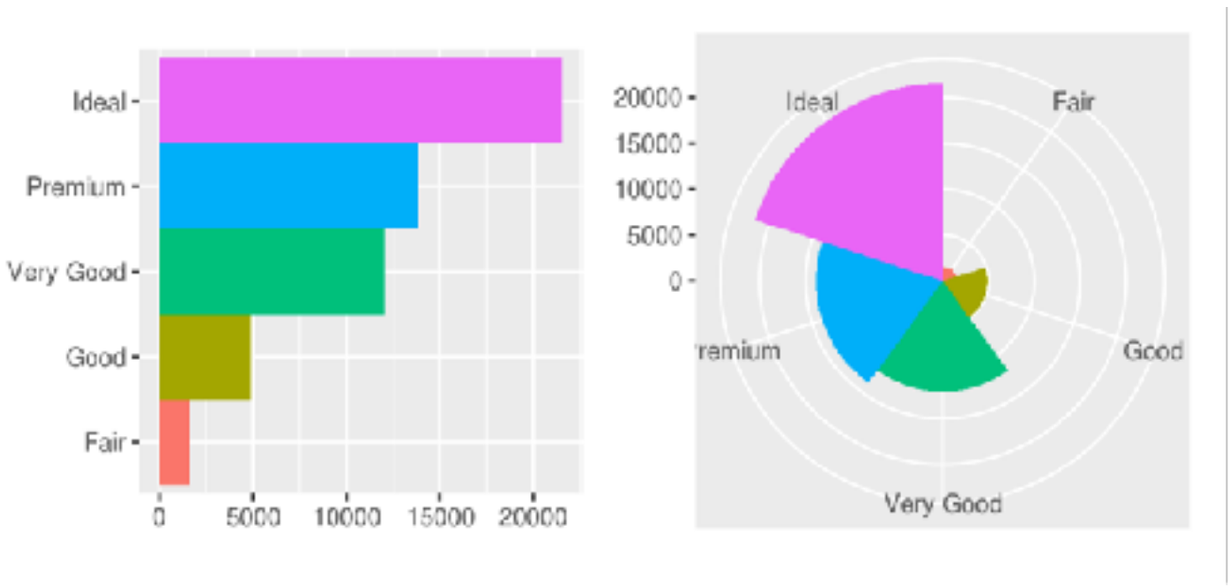
Elements of Graphs



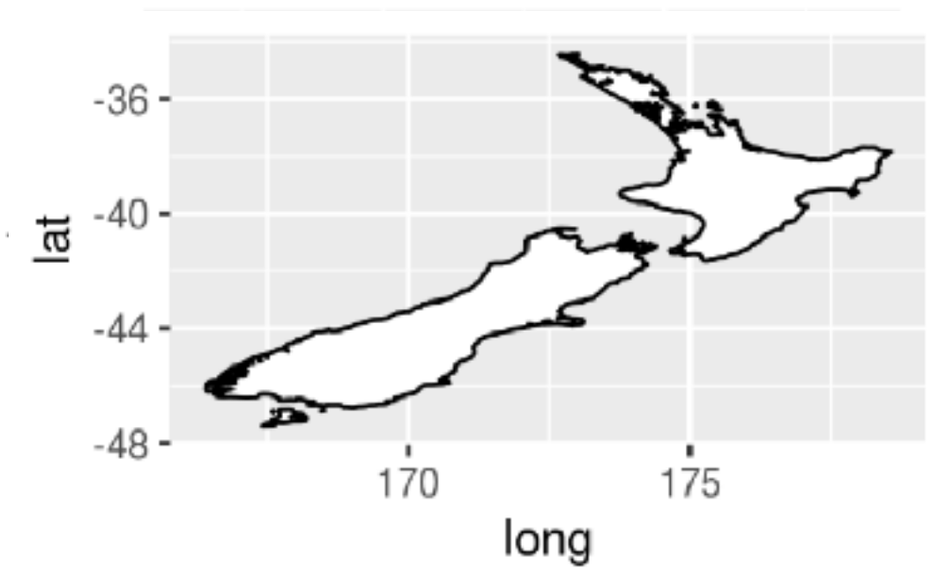
Logarithm

Element: scales

Elements of Graphs



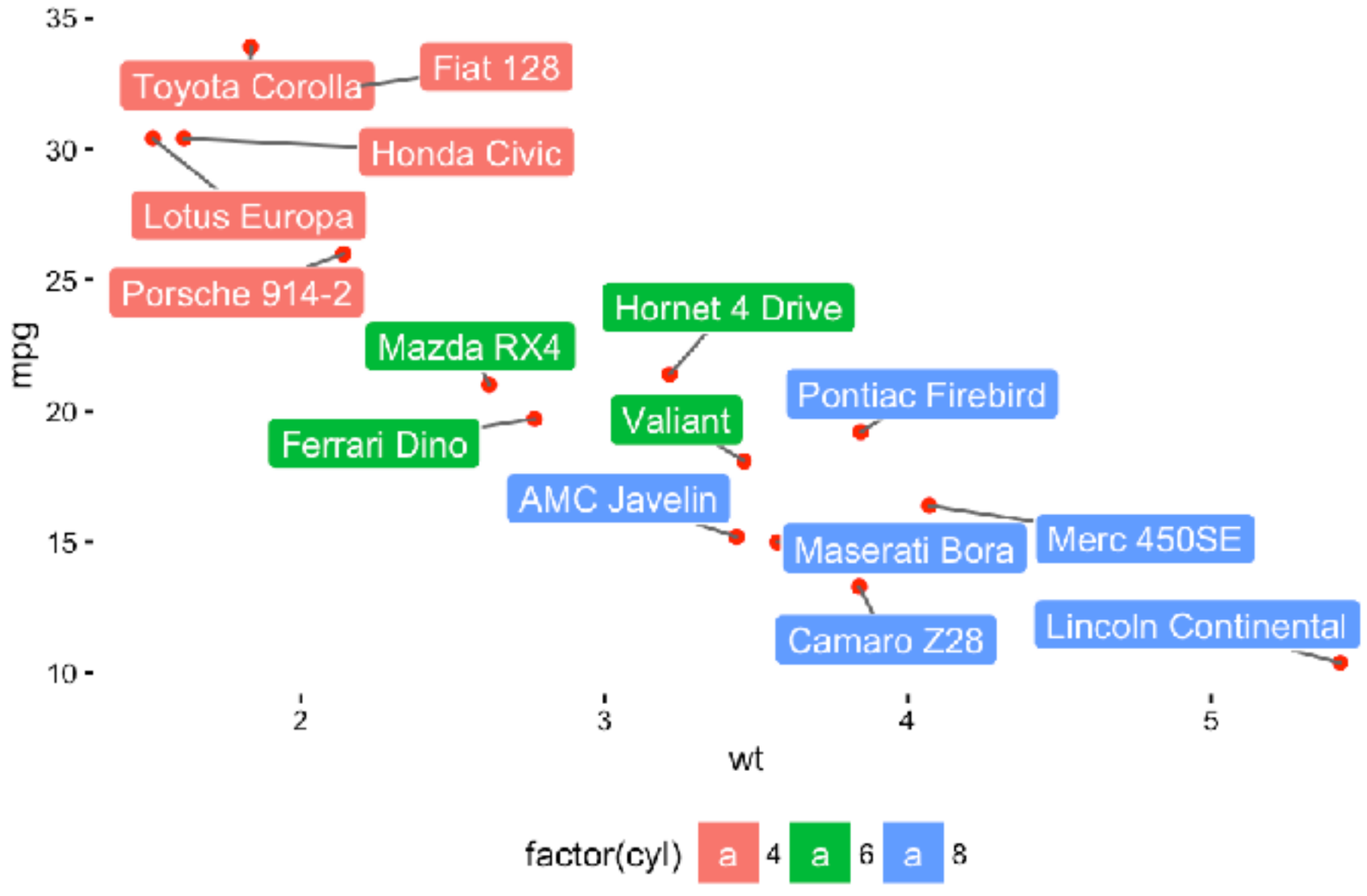
Polar



Geographic

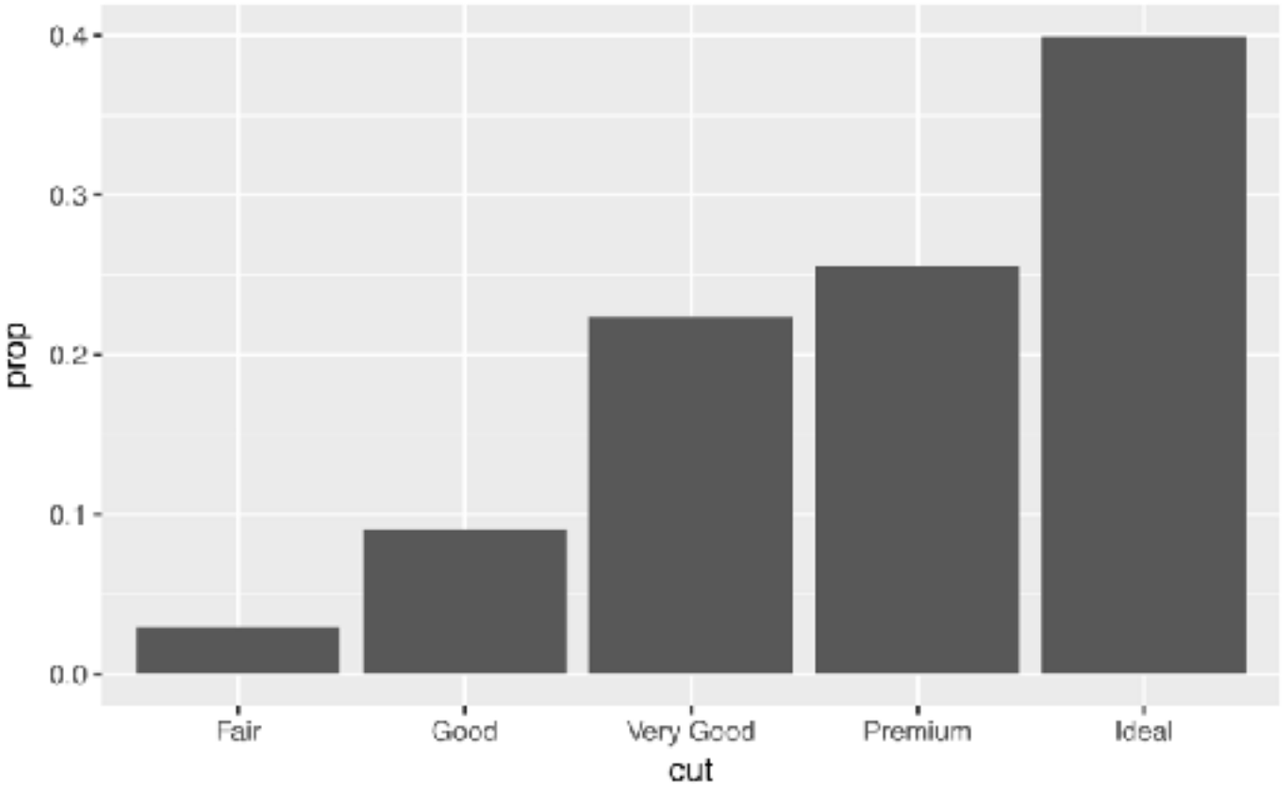
Element: Coordinates systems

Elements of Graphs

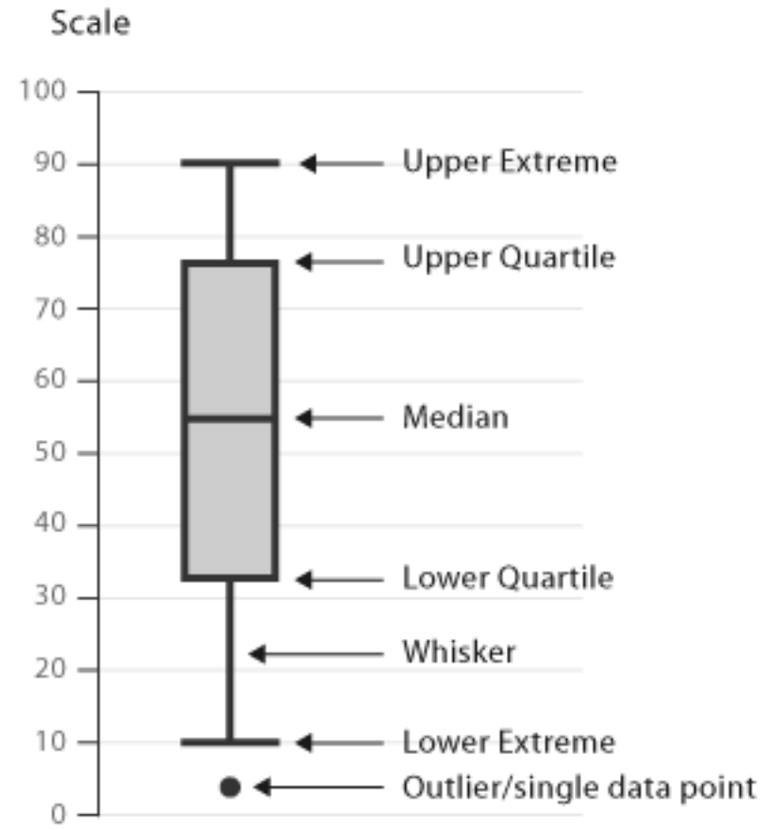


Element: plot annotations

Elements of Graphs



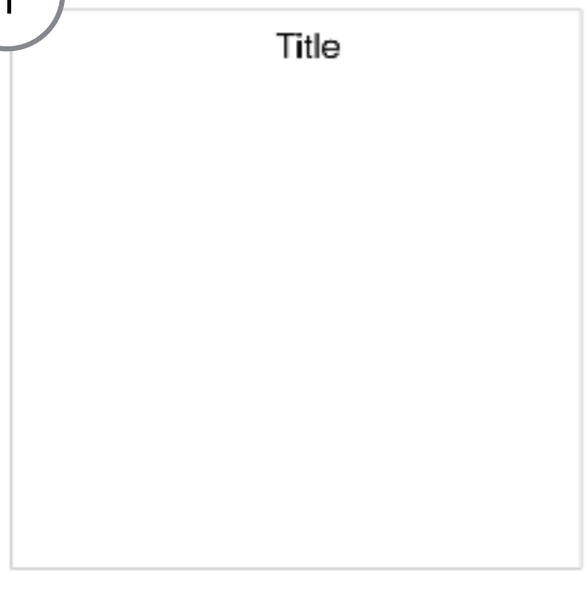
Bins



Element: Statistical transformation

Make the graph

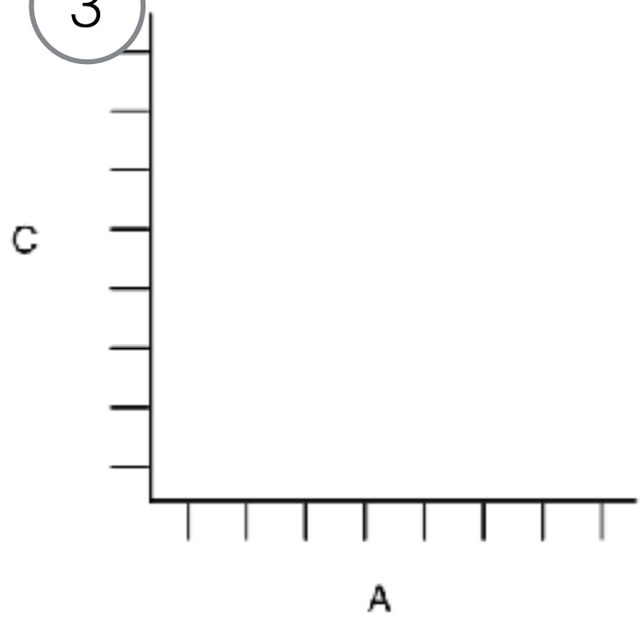
1



2

A	B	C	D
2	3	4	a
1	2	1	a
4	5	15	b
9	10	80	b

3

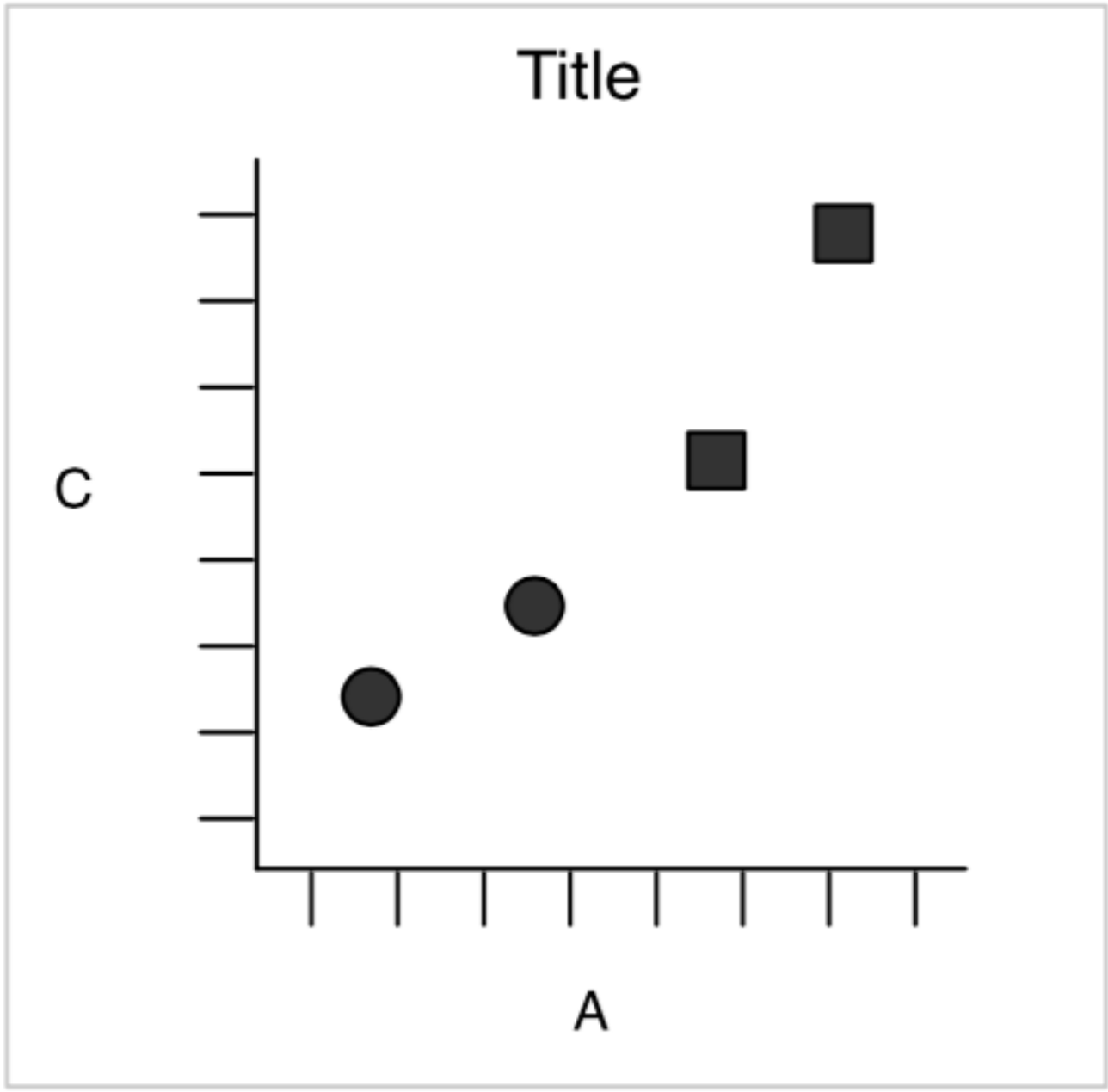


4

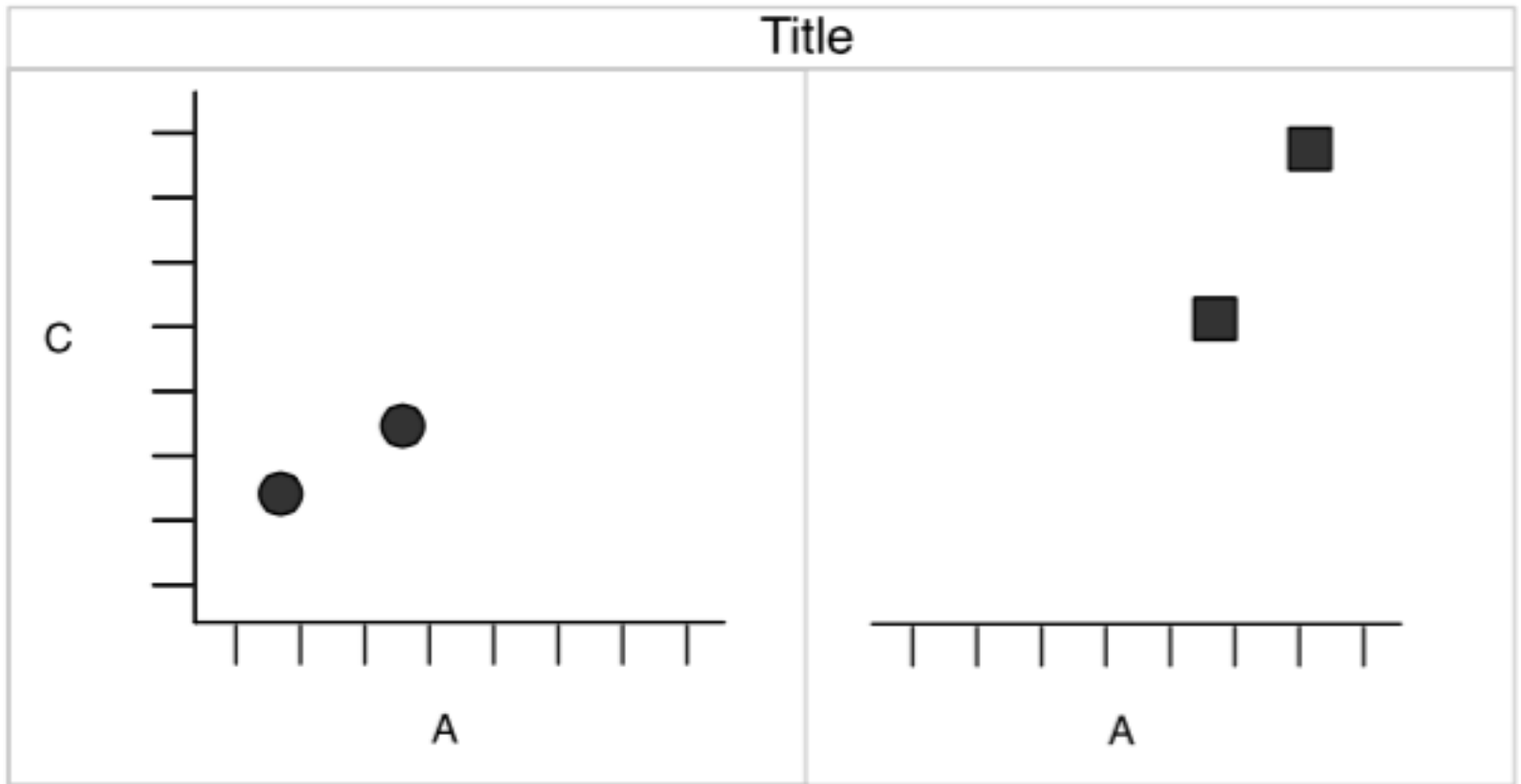
5

A	C	D
2	4	a
1	1	a
4	15	b
9	80	b

Result



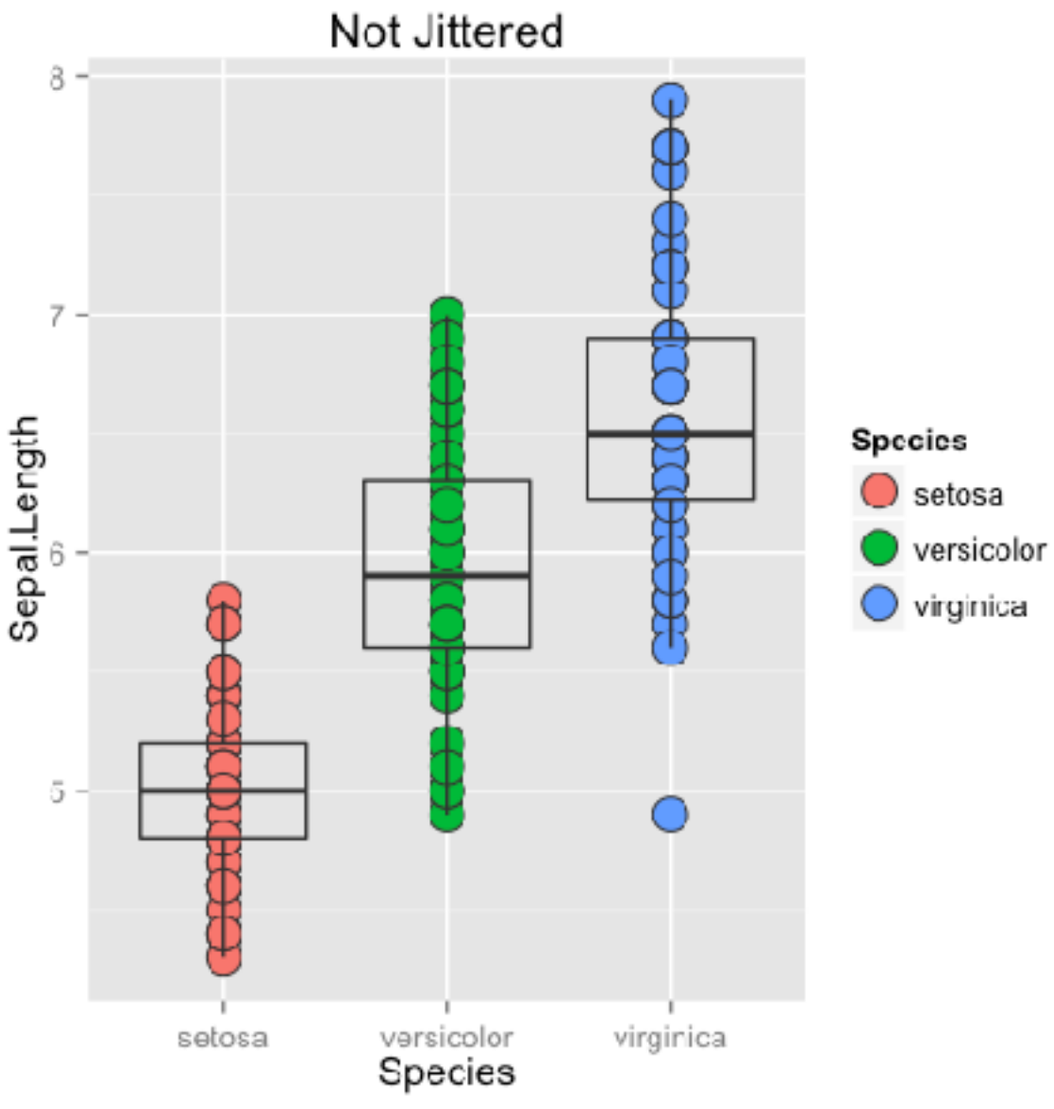
Elements of Graphs

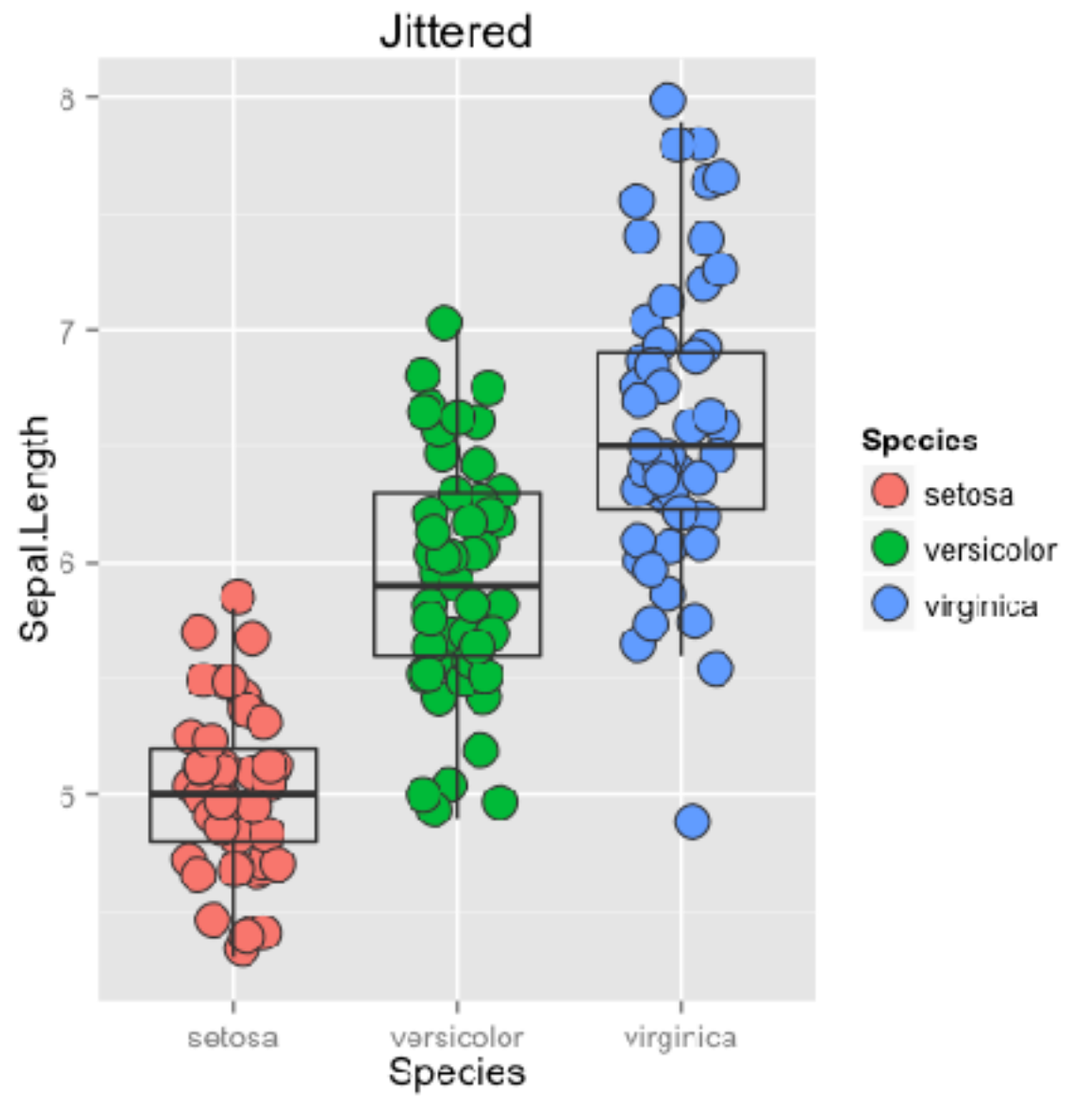


Facets by variable D

Element: Facets

Elements of Graphs





Element: position adjustment

The Grammar

A plot may have multiple **layers**. The layered grammar defines the components of a plot as:

1- A default **dataset**[1] and set of mappings from variables to **aesthetics**[2]
2- One or more layers, with each layer having one **geometric object**[3], one **statistical transformation**[4], one **position adjustment**[5], **plot annotations**[6](optionally, a different dataset and set of aesthetic mappings)
3- One **scale**[7] for each aesthetic mapping used
4- A **coordinate system**[8]
5- The **facet**[9] specification

Exercise

VisualizeData.R

Common problems

- Misplaced character

- Matching of (and “. Check the left-hand of your console:

- Put the + in the wrong place:

```
ggplot(data = mpg)
+ geom_point(mapping = aes(x = displ, y = hwy))
```

Common solutions

- Check the left-hand of your console. If it's a +, it means that R doesn't think you've typed a complete expression. In this case it's usually easy to start from scratch again by pressing ESCAPE

- You can get help about R function by running ?function_name in the console