LEADING Boot Camp

Information Visualization

Erjia Yan

Introduction

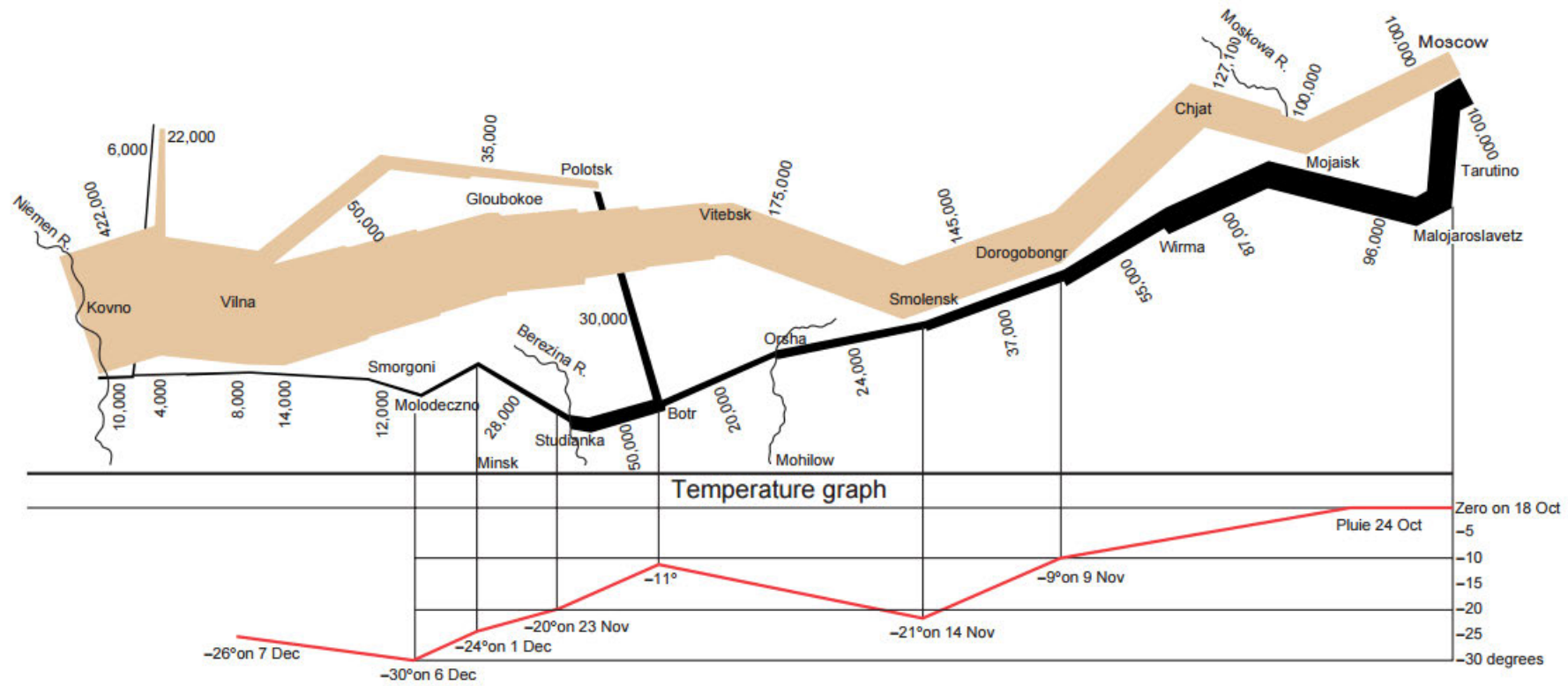Section I

FIGURE 1 | Losses suffered by the Grande Armée during the Russian Campaign.[c]
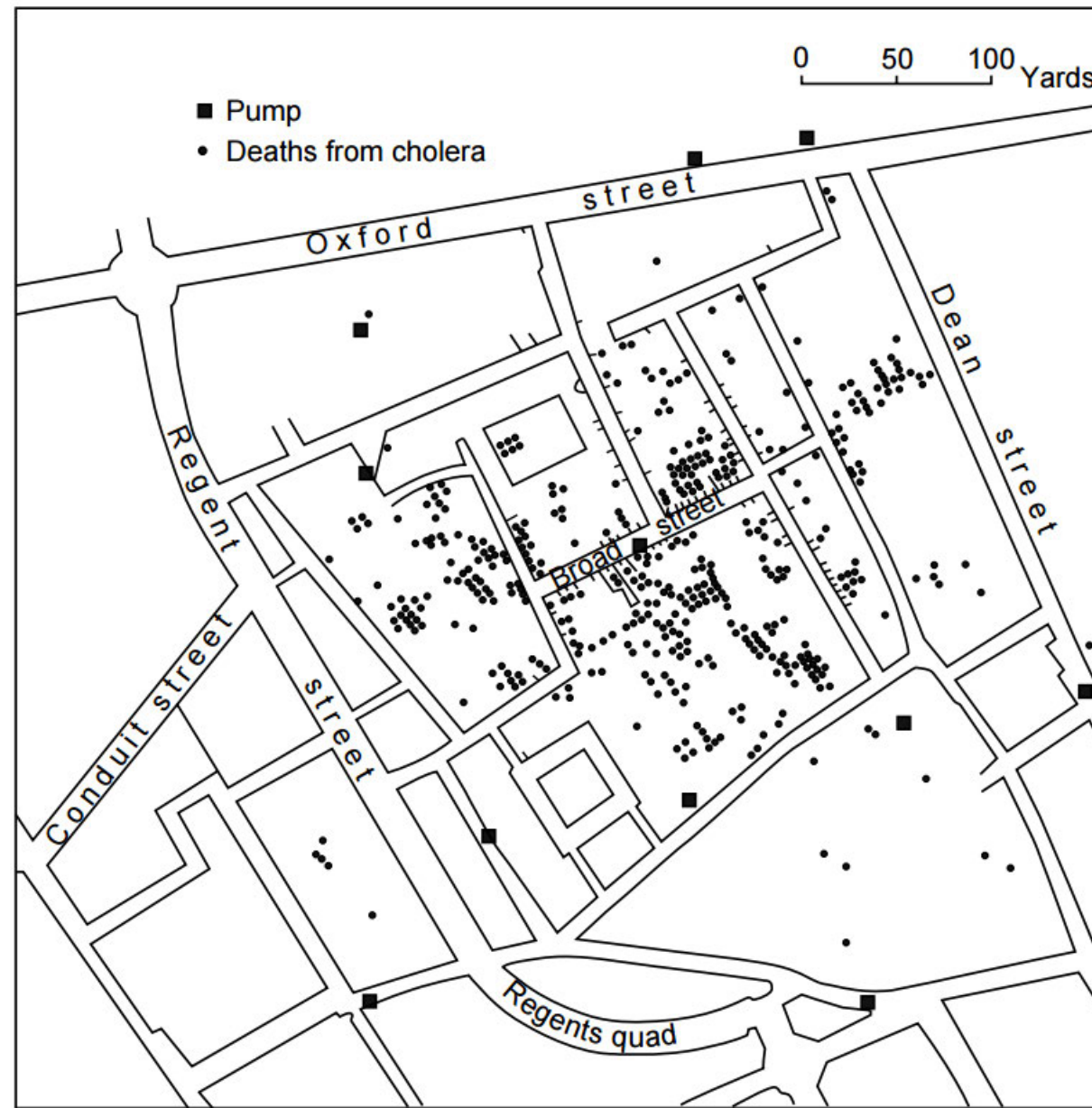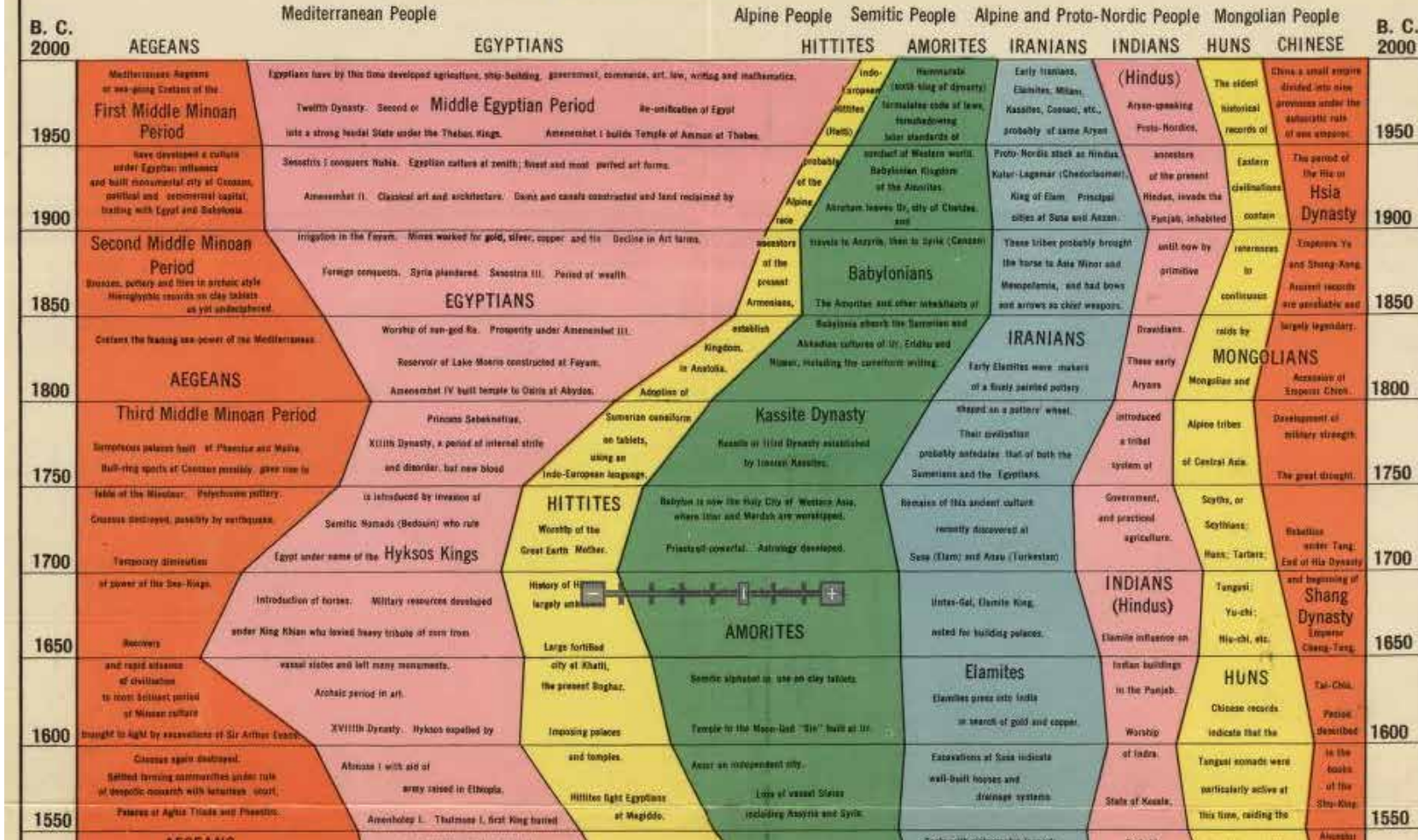
FIGURE 2 | John Snow's dot map of cholera deaths.[d]

The HISTOMAP

FOUR THOUSAND YEARS OF WORLD HISTORY

RELATIVE POWER OF CONTEMPORARY STATES, NATIONS AND EMPIRES

	Mediterranean People		Alpine People	Semitic People	Alpine and Proto-Nordic People		Mongolian People		
B. C. 2000	**AEGEANS**	**EGYPTIANS**	**HITTITES**	**AMORITES**	**IRANIANS**	**INDIANS**	**HUNS**	**CHINESE**	**B. C. 2000**

1950 — Mediterranean Aegeans or sea-going Cretans of the **First Middle Minoan Period** — Egyptians have by this time developed agriculture, ship-building, government, commerce, art, law, writing and mathematics. Twelfth Dynasty. Second or **Middle Egyptian Period**. Re-unification of Egypt into a strong feudal State under the Theban Kings. Amenemhet I builds Temple of Ammon at Thebes. — Hammurabi (sixth king of dynasty) formulates code of laws, foreshadowing later standards of — Early Iranians, Elamites, Mitani, Kassites, Cossaei, etc., probably of same Aryan — **(Hindus)** Aryan-speaking Proto-Nordics. — The oldest historical records of — China a small empire divided into nine provinces under the autocratic rule of nine emperors.

1900 — have developed a culture under Egyptian influence and built monumental city of Cnossus, political and commercial capital, trading with Egypt and Babylonia. — Sesostris I conquers Nubia. Egyptian culture at zenith; finest and most perfect art forms. Amenemhet II. Classical art and architecture. Dams and canals constructed and land reclaimed by — conduct of Western world. Babylonian Kingdom of the Amorites. Abraham leaves Ur, city of Chaldea, and — Proto-Nordic stock as Hindus. Kutur-Lagamar (Chedorlaomer) King of Elam. Principal cities of Susa and Anzan. — ancestors of the present Hindus, invade the Punjab, inhabited — Eastern civilizations contain — The period of the Hia or **Hsia Dynasty**

1850 — **Second Middle Minoan Period**. Bronzes, pottery and tiles in archaic style. Hieroglyphic records on clay tablets as yet undeciphered. — irrigation in the Fayum. Mines worked for gold, silver, copper and tin. Decline in Art forms. Foreign conquests. Syria plundered. Sesostris III. Period of wealth. **EGYPTIANS** — ancestors of the present Armenians. — travels to Assyria, then to Syria (Canaan). **Babylonians**. The Amorites and other inhabitants of — These tribes probably brought the horse to Asia Minor and Mesopotamia, and had bows and arrows as chief weapons. — until now by primitive — renerences to continuous — Emperors Yu and Shung-Kong. Ancient records are unreliable and

1800 — Cretans the leading sea-power of the Mediterranean. **AEGEANS** — Worship of sun-god Ra. Prosperity under Amenemhet III. Reservoir of Lake Moeris constructed at Fayum. Amenemhet IV built temple to Osiris at Abydos. — establish Kingdom in Anatolia. Adoption of — Babylonia absorb the Sumerian and Akkadian cultures of Ur, Eridhu and Nippur, including the cuneiform writing. — **IRANIANS**. Early Elamites were makers of a finely painted pottery — Dravidians. These early Aryans. — raids by **MONGOLIANS**. Accession of Empress Chien. — largely legendary.

1750 — **Third Middle Minoan Period**. Sumptuous palaces built at Phaestus and Mallia. Bull-ring sports at Cnossus possibly give rise to — Princess Sebeknofrue. XIIIth Dynasty, a period of internal strife and disorder, but new blood — Sumerian cuneiform on tablets, using an Indo-European language. — **Kassite Dynasty**. Kassite or IIIrd Dynasty established by Iranian Kassites. — shaped on a potters' wheel. Their civilization probably antedates that of both the Sumerians and the Egyptians. — introduced a tribal system of Central Asia. — Mongolian and Alpine tribes of Central Asia. — Development of military strength. The great drought.

1700 — fable of the Minotaur. Polychrome pottery. Cnossus destroyed, possibly by earthquake. Temporary diminution — is introduced by invasion of Semitic Nomads (Bedouin) who rule Egypt under name of the **Hyksos Kings** — **HITTITES**. Worship of the Great Earth Mother. — Babylon is now the Holy City of Western Asia, where Ishtar and Marduk are worshipped. Priestcraft powerful. Astrology developed. — Remains of this ancient culture recently discovered at Susa (Elam) and Anau (Turkestan). — Government, and practical agriculture. — Scyths, or Scythians; Huns; Tartars; — Rebellion under Tang. End of Hia Dynasty

1650 — of power of the Sea-Kings. Recovery — Introduction of horses. Military resources developed under King Khian who levied heavy tribute of corn from — History of Hittites largely unknown. Large fortified — **AMORITES** — Unter-Gal, Elamite King, noted for building palaces. — **INDIANS (Hindus)**. Elamite influence on Niu-chi, etc. — Tangusi. Yu-chi; Niu-chi, etc. — and beginning of **Shang Dynasty**. Emperor Chang-Tang.

1600 — and rapid advance of civilization to reach foremost period of Minoan culture brought to light by excavations of Sir Arthur Evans. — vassal states and left many monuments. Archaic period in art. XVIIIth Dynasty. Hyksos expelled by — city at Khatti, the present Boghaz. Imposing palaces — Semitic alphabet in use on clay tablets. Temple to the Moon-God "Sin" built at Ur. — **Elamites**. Elamites press into India in search of gold and copper. — Indian buildings in the Punjab. Chinese records indicate that the — **HUNS**. Tai-Chia. Period described

1550 — Cnossus again destroyed. Settled farming communities under rule of despotic monarch with luxurious court. Palaces of Aghia Triada and Phaestus. **AEGEANS** — Ahmose I with aid of army raised in Ethiopia. Amenhotep I. Thutmose I, first King buried — and temples. Hittites fight Egyptians at Megiddo. — Assur an independent city. Loss of vassal States including Assyria and Syria. — Excavations at Susa indicate well-built houses and drainage systems. State of Kassia. — of India. Tangusi nomads were particularly active at this time, raiding the — in the books of the Shu-King. Alexander

Mental models

Section II

Population Share

Rur

76%
of the extreme poor
live in rural areas

The NSW Health system is…

▶ **RECRUITING MORE NURSES***

47,500 +

46,573

43,147   43,346   43,405

2008/09   2009/10   2010/11   2011/12   to March 2013
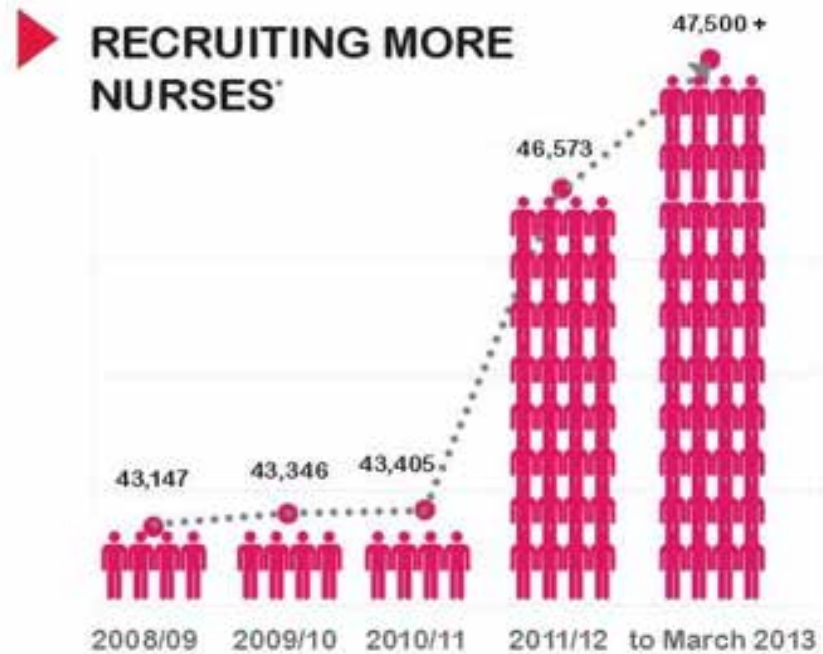
* Nursing headcount figures at June includes non casual staff and 3rd schedule

Some Important Properties

Mind-sets tend to be quick to form but resistant to change.

Once events have been perceived one way, there is a natural resistance to other perspectives.

New information is assimilated to existing images.

So, don't rush it.

Gradual, evolutionary change often goes unnoticed.

It takes more information, and more unambiguous information, to recognize an unexpected phenomenon than an expected one.

Be aware what you are expecting

Mental models are easy to form, but hard to change

The early but incorrect impression tends to persist because the amount of information necessary to invalidate a hypothesis is considerably greater than the amount of information required to make an initial interpretation.

People form impressions on the basis of very little information, but once formed, they do not reject or change them unless they obtain rather solid evidence.

Analysts might seek to limit the adverse impact of this tendency by suspending judgment for as long as possible as new information is being received.

Keeping an Open Mind

- Major intelligence failures are usually caused by failures of analysis, not failures of collection.

- Relevant information is discounted, misinterpreted, ignored, rejected, or overlooked because it fails to fit a prevailing mental model or mind-set.

Misleading visualizations...

- https://venngage.com/blog/misleading-graphs/
  1. Omitting the baseline
  2. Manipulating the Y-axis
  3. Cherry picking data
  4. Using the wrong graph
  5. Going against conventions
  6. Misleading Coronavirus graphs

Grammar of Graphics

Section III

Grammar of Graphics (I)

- Grammar of graphics deals with the question: what components are included in information visualization outputs?
- Why it's important?
  - You need to learn about the language of InfoVis.
  - It is a very important design tool.

grammar noun

gram·mar | \ ˈgra-mər \

Definition of grammar

1 **a** : the study of the classes of words, their inflections (see INFLECTION sense 3), and their functions and relations in the sentence

b : a study of what is to be preferred and what avoided in inflection (see INFLECTION sense 3) and syntax (see SYNTAX sense 1)
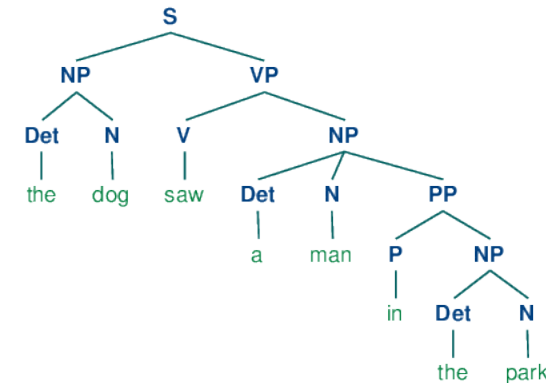
Grammar of Graphics (II)

- It was first proposed by Leland Wilkinson, an American statistician and computer scientist, in his book *The Grammar of Graphics*, published in 1999.
- It became the foundation of the ggplot2 (*gg* stands for grammar of graphics) visualization package and some other visualization projects.
  - https://ggplot2.tidyverse.org/reference/index.html
  - https://bokeh.pydata.org/en/latest/
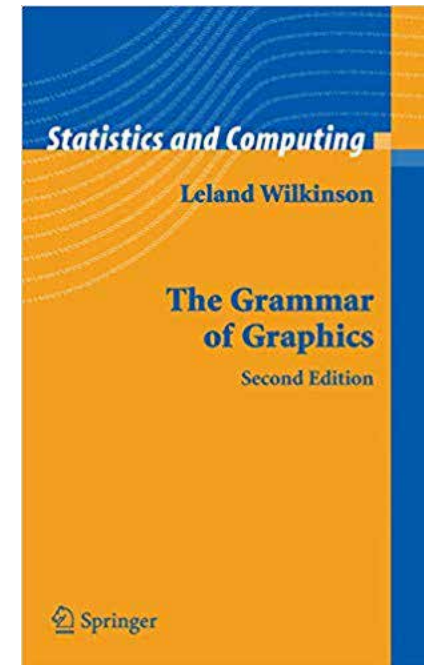
Grammar of Graphics (III)
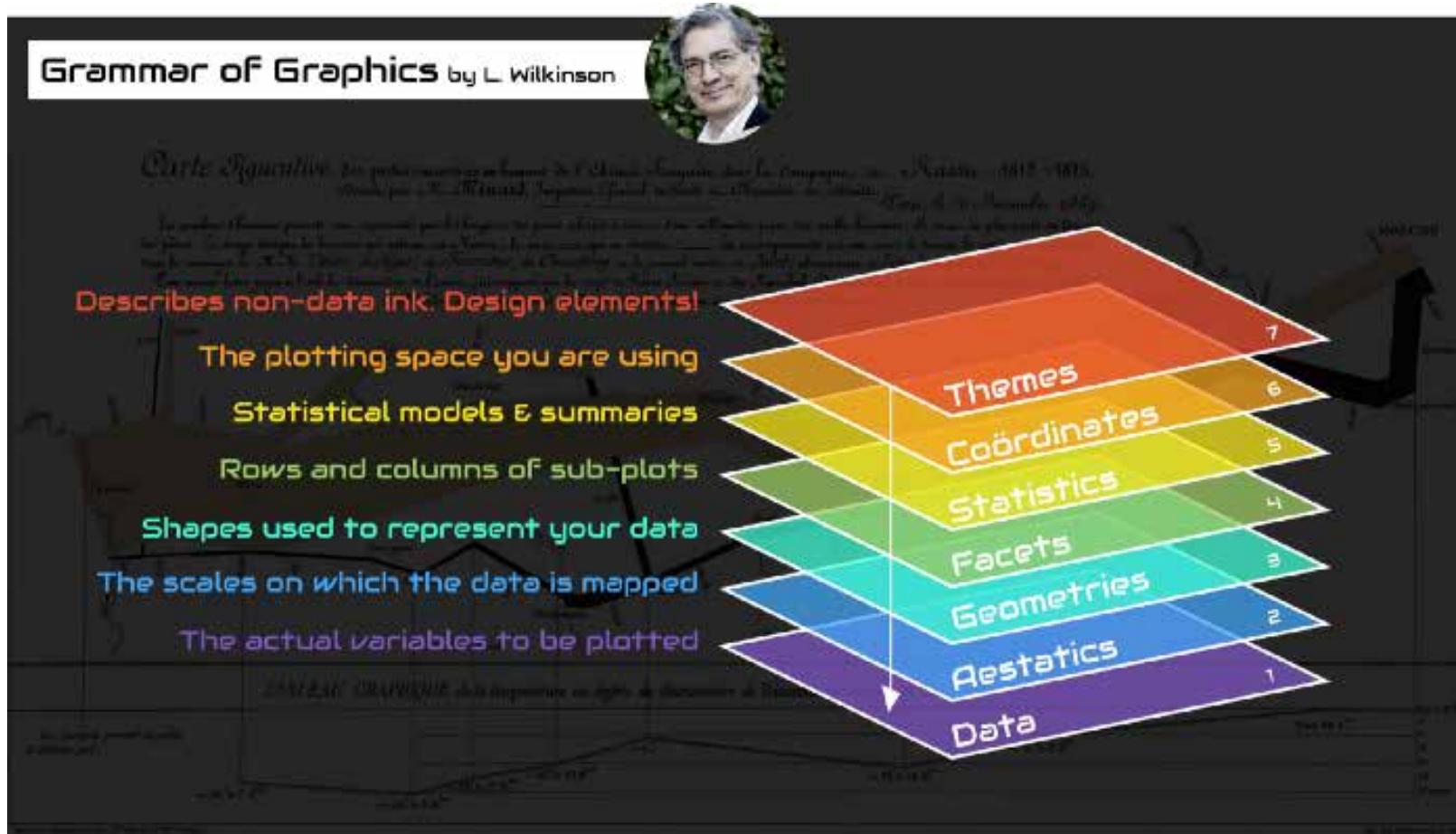
Aesthetics (I)

- Aesthetics are the shapes that we see in the graph. In general, you should consider using the following aesthetic elements (or "visual patterns"):
  - Axes
  - Color
  - Shape
  - Size
  - Transparency
- Question 2: What visual patterns can be meaningfully used in each graph type? (Answer this question based on the **ontology of visual patterns** in this week's class.)

Aesthetics (II)

- The availability and applicability of aesthetic elements are dependent upon the specific graph type.

- Some elements cannot be used or are meaningless in certain graph types. (Can you think of any examples?)
  - Is color a meaningful aesthetic element in:
    - Pie chart?
    - Bar chart?
    - Scatterplot?
  - How about size?

Aesthetics (II)

- Some aesthetics elements are more strongly connected to certain type of data.
  - Think about if the elements on the right can be used for *numeric or categorical data, or both*?
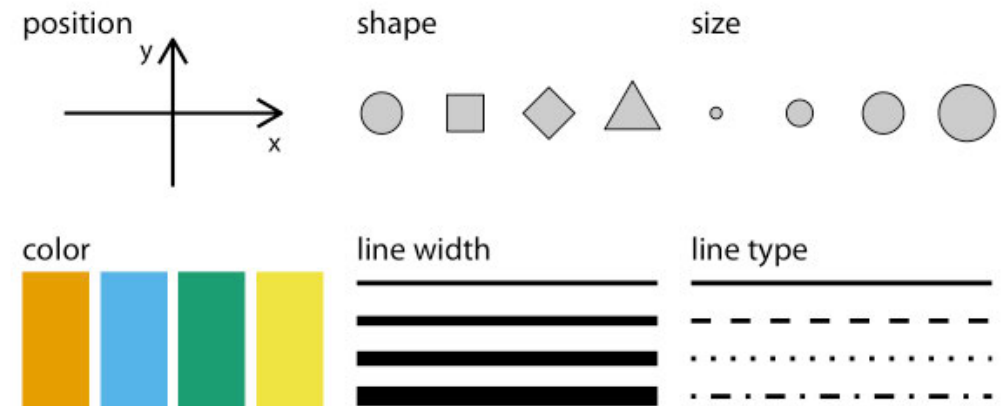
Scale

- All aesthetic elements representing a data value need to be put into a scale. Some examples of the scale includes:
  - whether the data should be represented as numeric or categorical
  - whether the numeric scale should be reversed
  - whether the numeric scale should be linear or statistically transformed (log, square root...)
  - what color palette should be used
  - what scheme of shapes should be used
  - ...

Ordinal and reversed scales

- In general, ordinal data should be plotted to categorical scales, no matter which type of aesthetic element that is used.
  - As compared, interval data (like Fahrenheit degrees) should be plotted to numeric scales.
- Reversed scales are used when:
  - Opposite graph (https://playfairdata.com/how-to-make-a-diverging-bar-chart-in-tableau/)
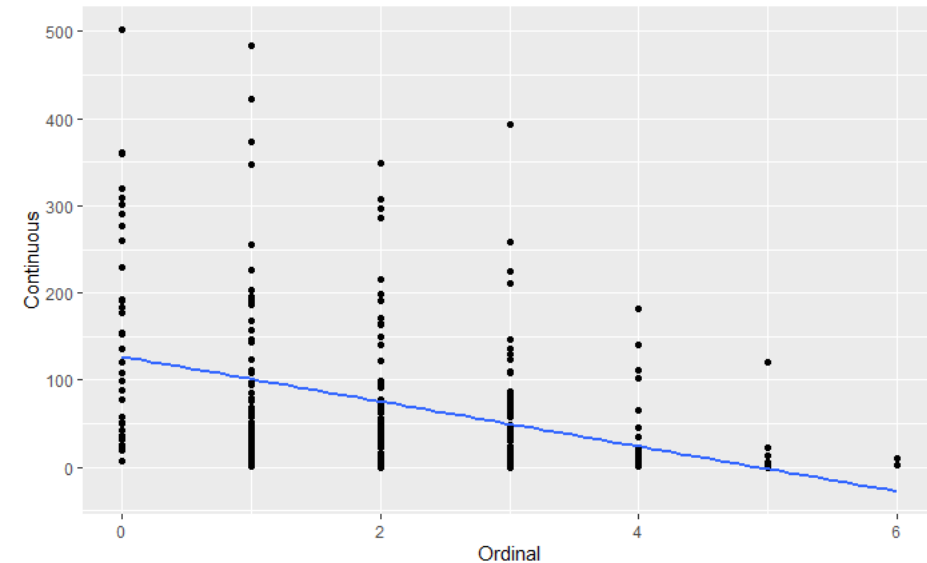
Logarithmic scale

- Logarithmic scales are often used in both axes, rather than other aesthetic elements.

- This scale should be used very carefully because it is quite difficult to be understood by the public.
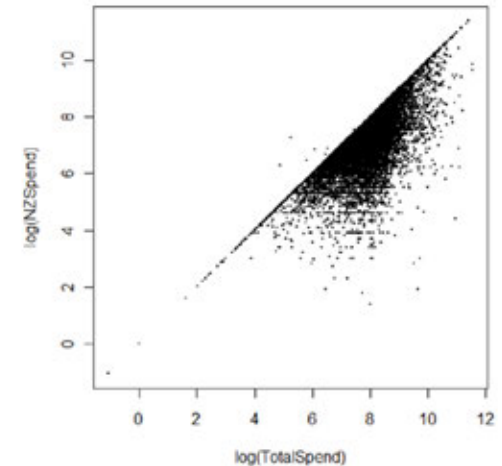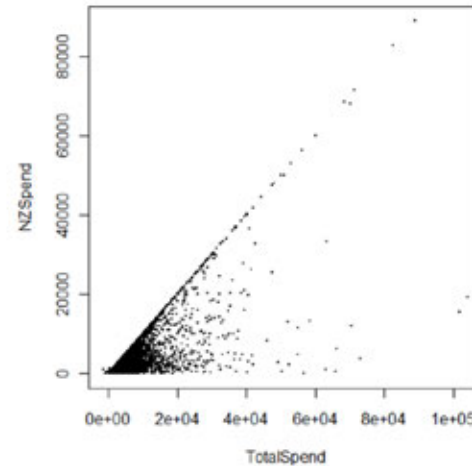
Legend of scale

- For all scales used not on axes, always use a legend!
  - An exception is when the information is obvious, in which case, think about the possibility of not using the aesthetic element.
- Even the same data point can be plotted to different types of scales.
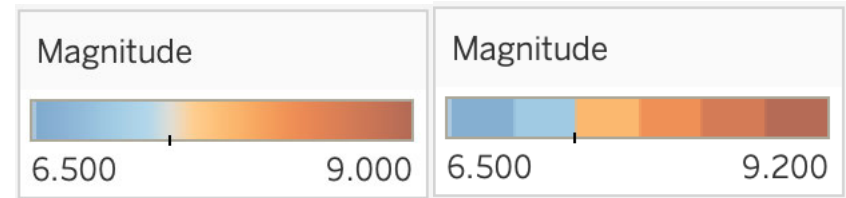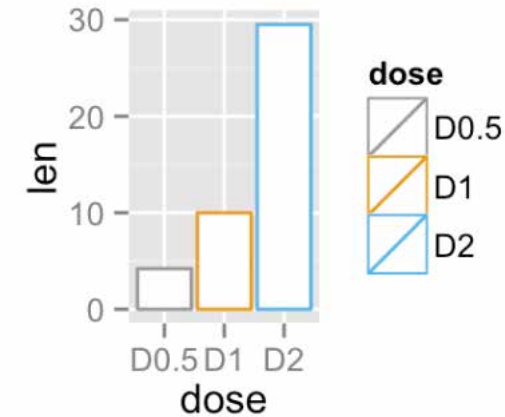
Geometrics

- Geometrics means the specific visual objects (graph types), which are the containers of visual patterns.

- It is a big decision in the visualization pipeline, but also one that is based on many other considerations (your questions, intended users...).

- No software can support all graph types, and thus it is important to understand what graph types can be created in which software.

Facet

- The idea of faceting is embedded in small multiples.

- As a layer in this framework, it is broadly applicable to nearly all graph types.
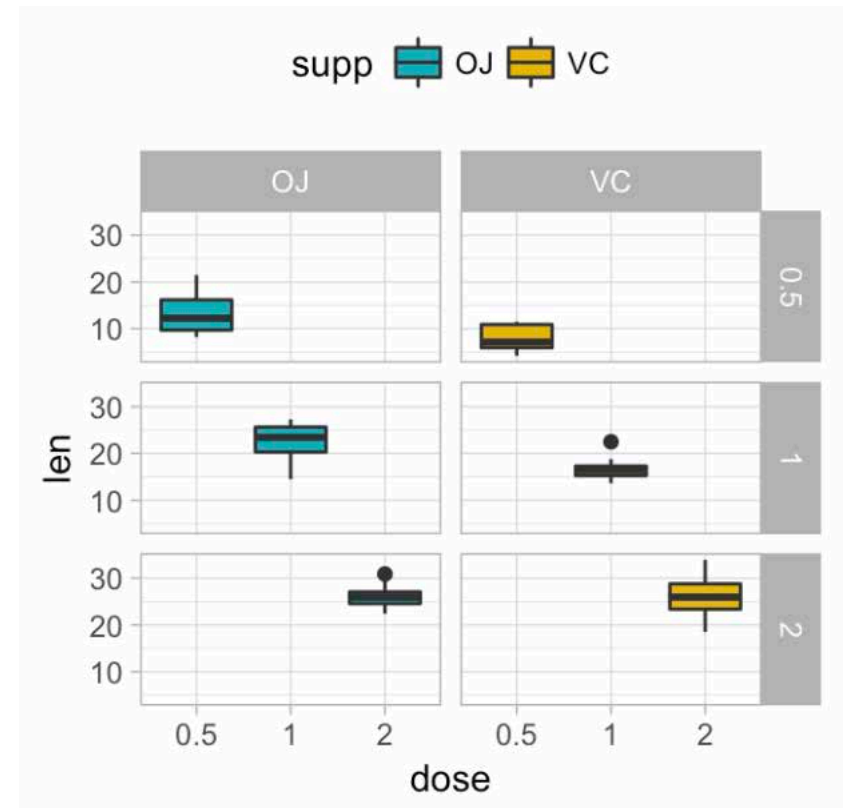
Statistical transformation

- We have talked about statistical transformation of scales, but that does not change the data points themselves.

- In many cases, we also want to transform the dataset per se to draw deeper meanings.
  - Sometimes, these procedures are black-boxed into the software.
  - A lot of transformations are embedded in the functions to create the visualization.

Group	Member
A	A1
A	A2
A	A3
B	B1
B	B2
B	B3
B	B4

Group	Member
A	3
B	4

Box plot

- Box plot is one type of graph that shows the statistical summary of data value distribution.
  - As compared to histogram, it is able to show an extra data point, the group.
- It is a useful when:
  - (1) there are too many data points in your dataset (so that they are not able to be displayed meaningfully) and
  - (2) all you need is to understand the distribution of groups.
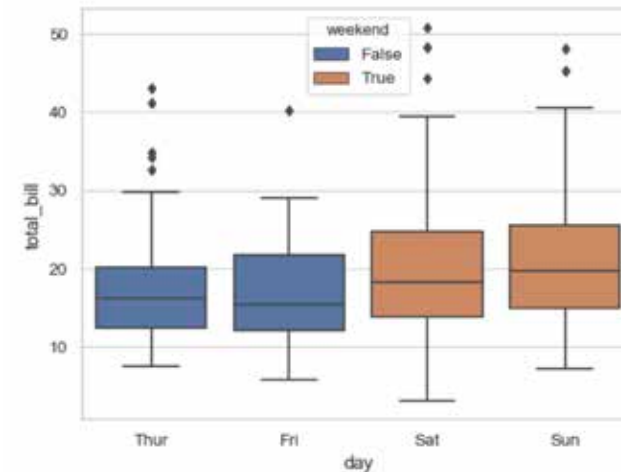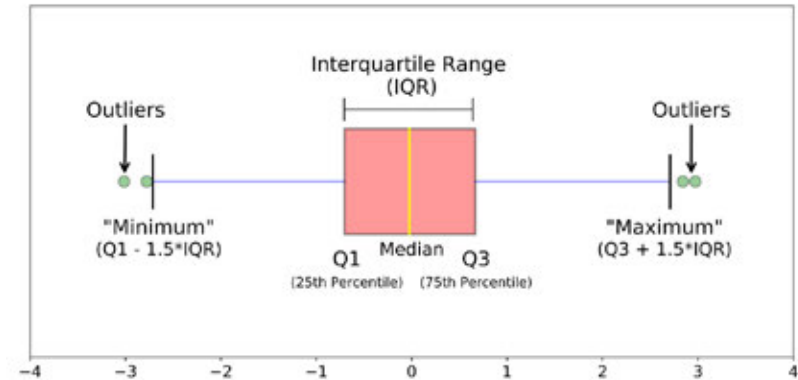
Regression (trend) line

- Different regression models can be applied to create the line to summarize the trend of the data.
  - Linear
  - Generalized linear
  - Local regression
  - …
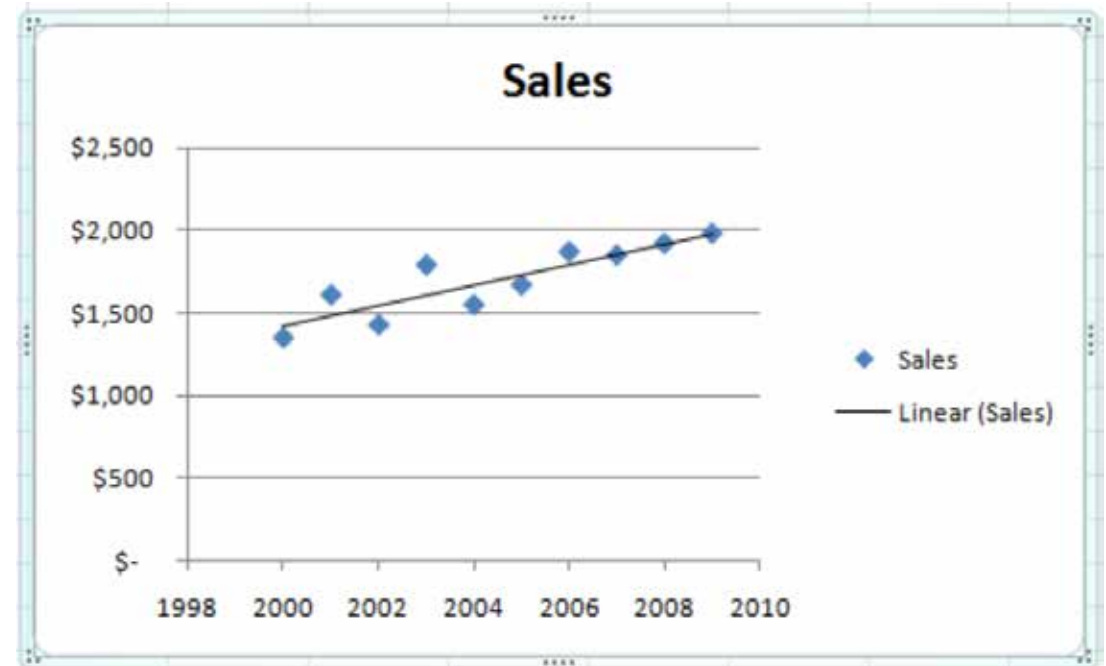
Annotation (I)

- Annotations are used so that users can get access to important information beyond the graph to understand the visualization.

- You should not assume that your readers know everything about your work. Instead, you should try to explain the most basic and/important things in clear and simple ways.

  - The amount of annotation also partly depends on whom your work is designed for.
  - But even for a visualization work designed for experts, it is still important to offer annotations for them to understand what is going on.

Annotation (II)

- You should think about the following types of annotations:
  - Title
  - Caption
    - Source of data
    - Authorship (if the visualization is not created by a very famous organization)
  - Highlights on aesthetic elements
    - Numbers
    - Shapes
    - Explanations

- Example: https://informationisbeautiful.net/visualizations/what-could-really-increase-life-expectancy-lifespan-and-longevity/

Coordinates systems

- A coordinate system in the visualization denotes the rules of how objects should be mapped to the 2-d space that we can see.

- Cartesian coordinate system: a point is referenced by two coordinates (x, y) specifying the positions along the x- and y-axes, respectively.
https://en.wikipedia.org/wiki/Cartesian_coordinate_system
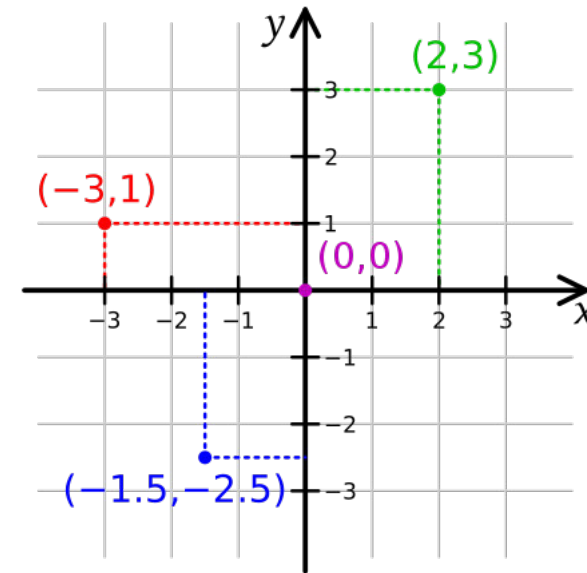
Polar coordinate system

- Polar coordinate system: a point is referenced by a distance from the reference point and an angle from a reference direction. https://en.wikipedia.org/wiki/Polar_coordinate_system
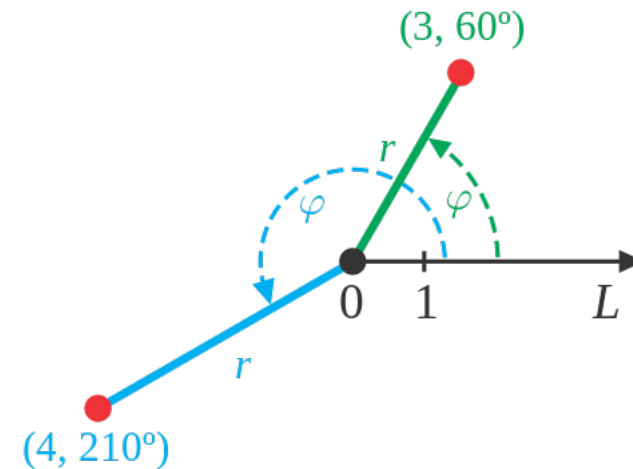
Some graphs are only different by the coordinate system

- Polar area graph types (such as pie chart, Nightingale graph, and spider graph) are some other graph types in the Cartesian coordinate system!
  - What are they?

- Is there any graph type that is not consistent with these coordinate systems?
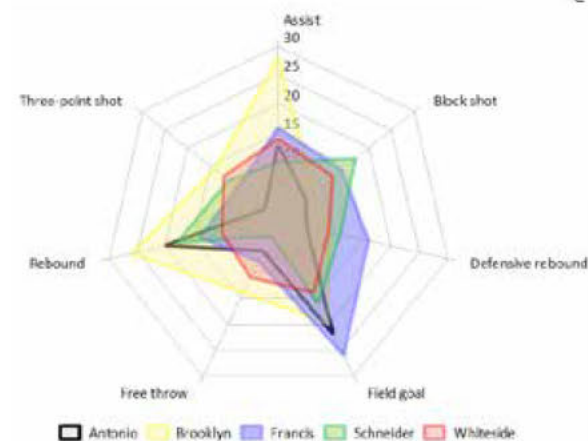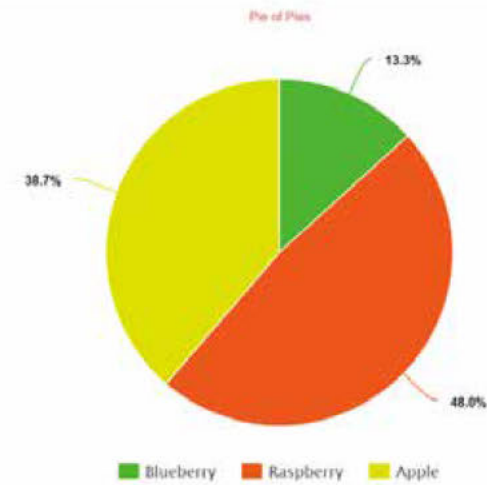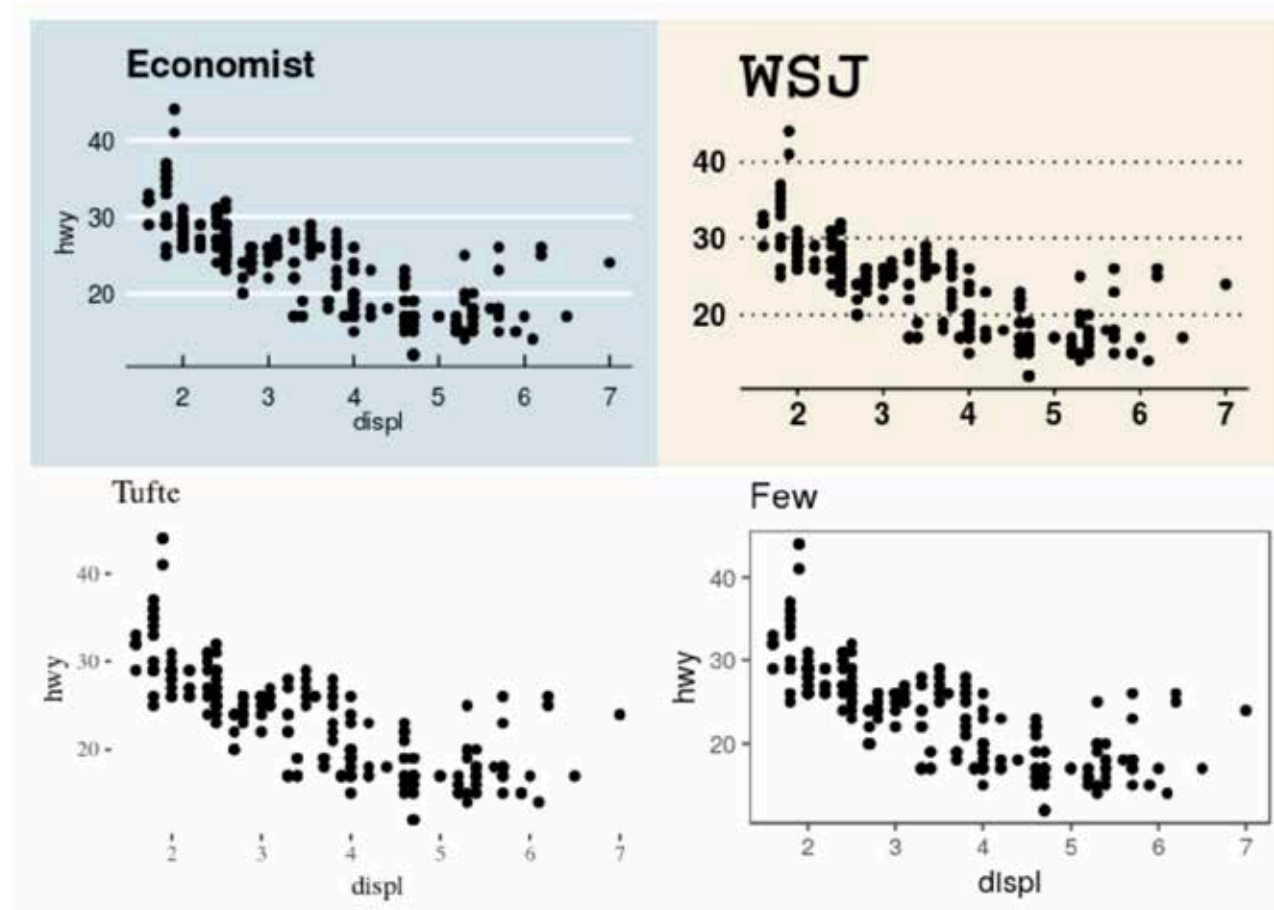
Why do we want to flip the two axes?

- In most cases, we use a flipped Cartesian coordinate system in a bar chart, where there are many objects to be compared with each other.
  - It is always easier to print your graph longer vertically than horizontally.
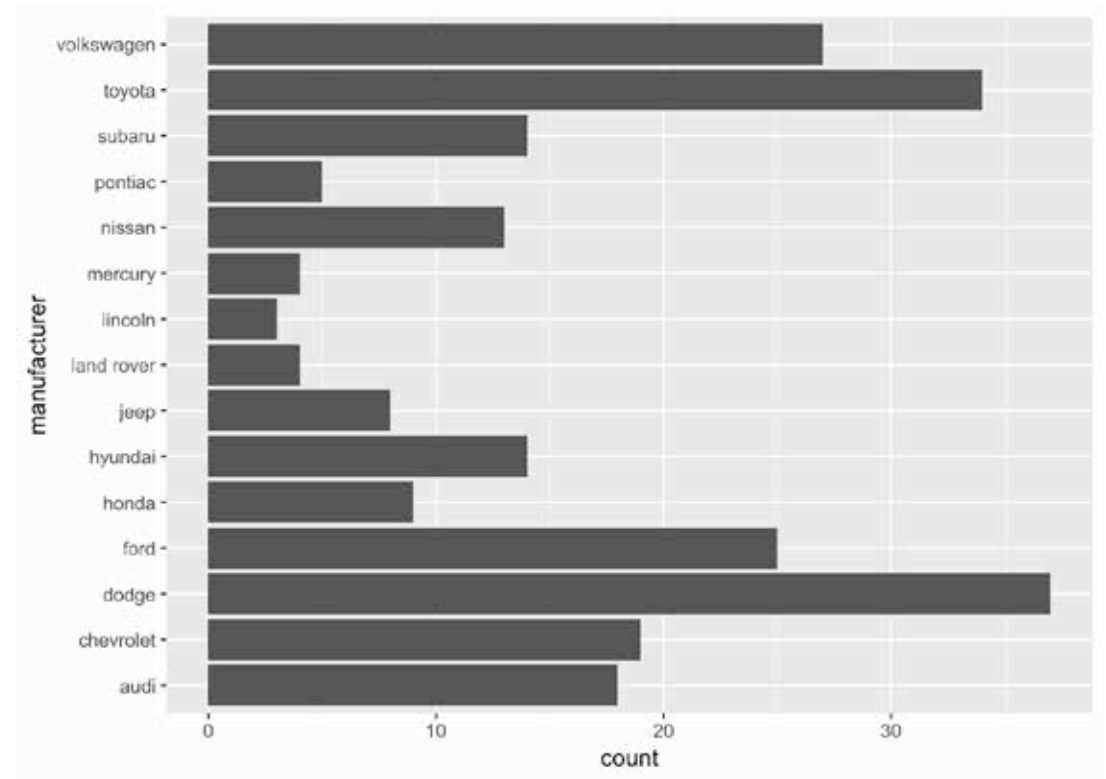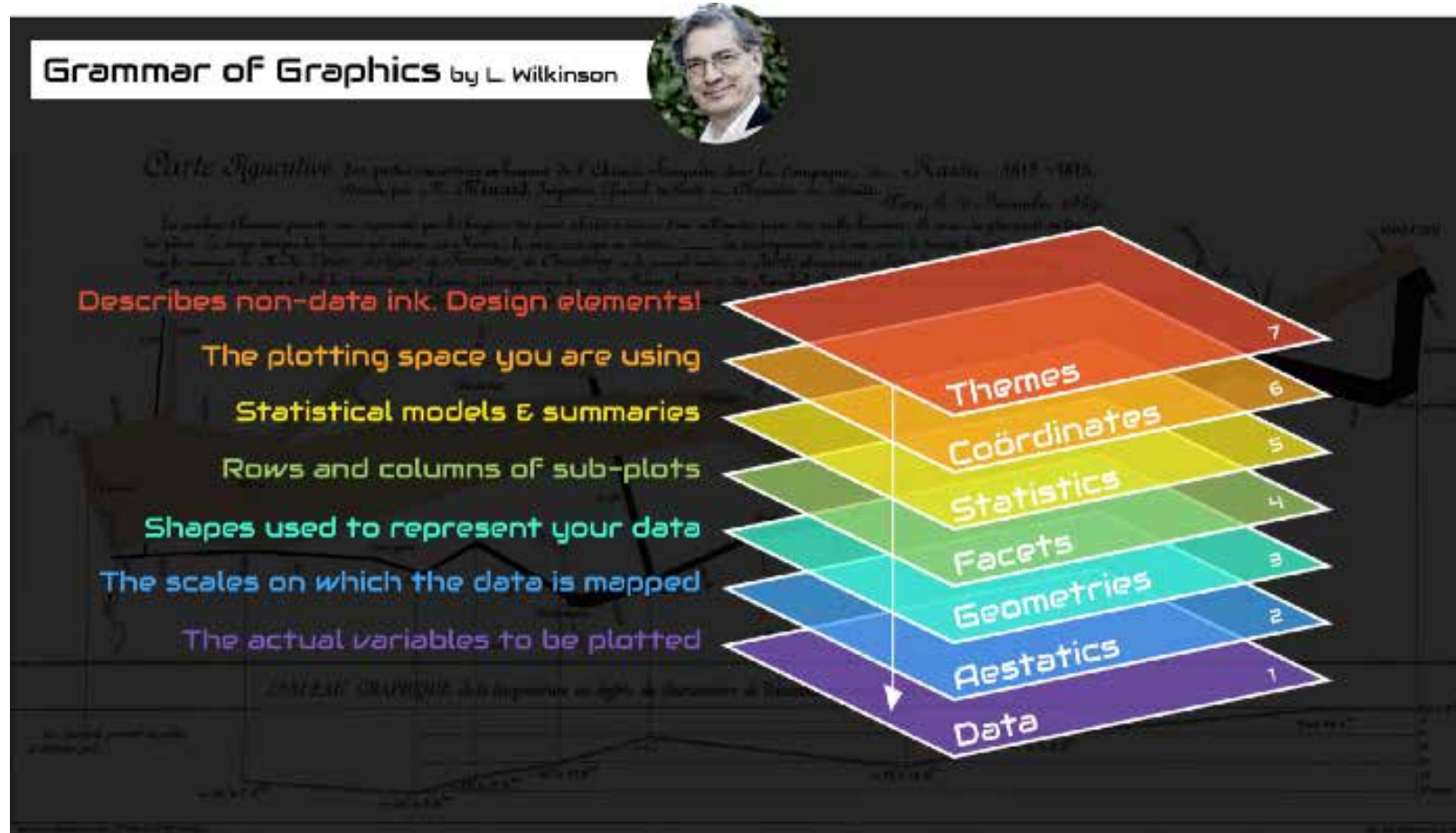
Theme (I)

- Theme is about how the non-data aspects of the graph should be like, such as:
  - Formats of titles and captions
  - Formats of axis and legend titles and labels
  - Background color
  - Displacement of components (such as guideline or labels)
- Theme is almost totally detached to the visual representations of data.

Theme (II)

Let's take a look at this picture again

Good resources

- https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368
- https://venngage.com/blog/how-to-choose-the-best-charts-for-your-infographic/
- https://www.data-to-viz.com/caveats.html