Visualization Design

Design Principles and Main Visual Forms

S. Santoni[1,2]

[1]Bayes Business School

[2]Soundcloud

MSc in Business Analytics, 2022/23

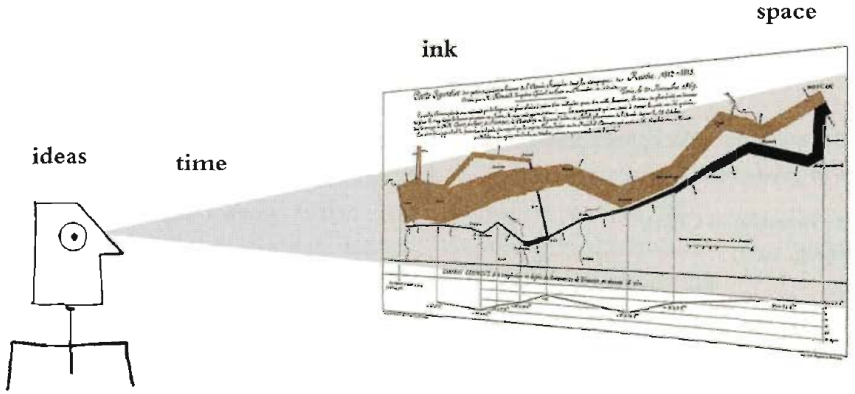
Outline

Outline

Graphical Excellence according to Tufte

Per Tufte's work [**tufte2001**], excellence in statistical graphs consists of complex *"ideas communicated with clarity, precision, and efficiency."*

Graphical displays pursuing clarity, precision, and efficiency *"give to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space."*

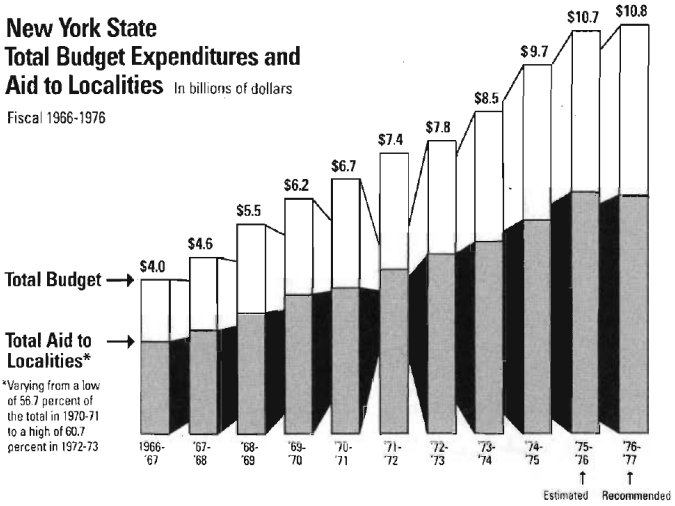


How to Reach Clarity, Efficiency, and Precision?

Tufte points out graphical displays should

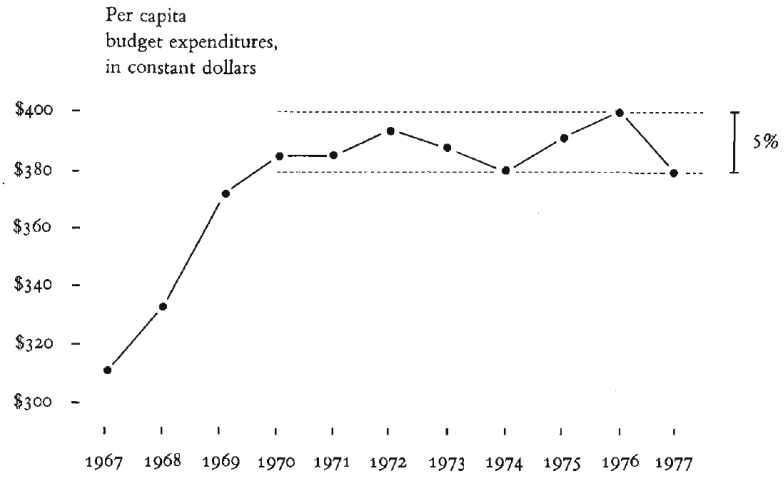
- Show the data
- Induce the viewer to think about the substance rather than about the methodology, graphical design, the technology of graphic production, or something else
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large datasets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- Be closely intergrated with the statistical and verbal description of a data set

Yet another Time Series that Lies



Source: New York Times, February 1, 1976, page IV-6.

Let Us Redesign the 'NYS Total Budget Expenditures' Graph



Your version of Tufte's chart

Outline

The Data Visualization Process acccording to Cairo

Data, information, knowledge, wisdom

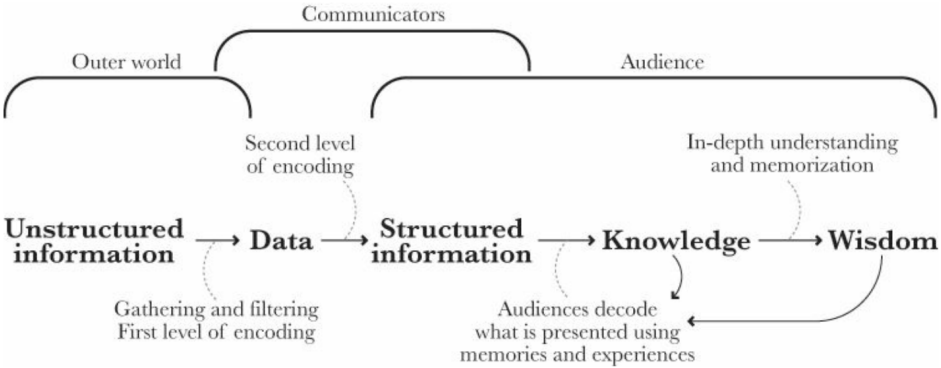


Figure 1.8. From reality to people's brains.

Source is [**cairo2012**]

How to Navigate the Data Visualization Process?

Tufte [**tufte2001**] suggests to adhere to a basic design principle:

> *The principle is the basis for a theory for a theory of data graphics. Show the data.*

> *The principle is the basis for a theory for a theory of data graphics.*

Fine — So What?!

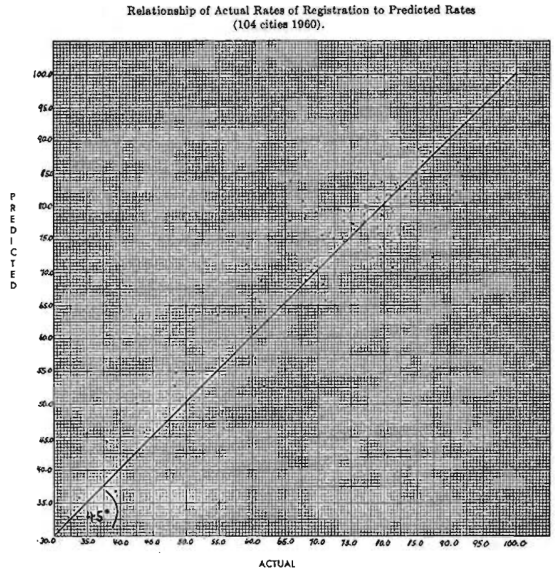
Maximize the Data-Ink Ratio!!

$$\text{Data-Ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic design}}$$

Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented.

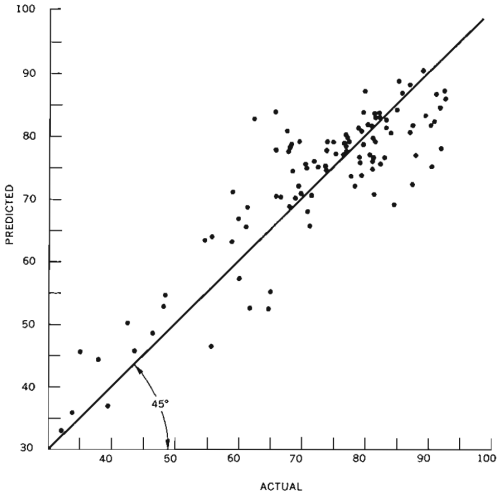
Theee Versions of the Same Scatter Diagram

Low data-ink ratio



Theee Versions of the Same Scatter Diagram

Substantial data-ink ratio

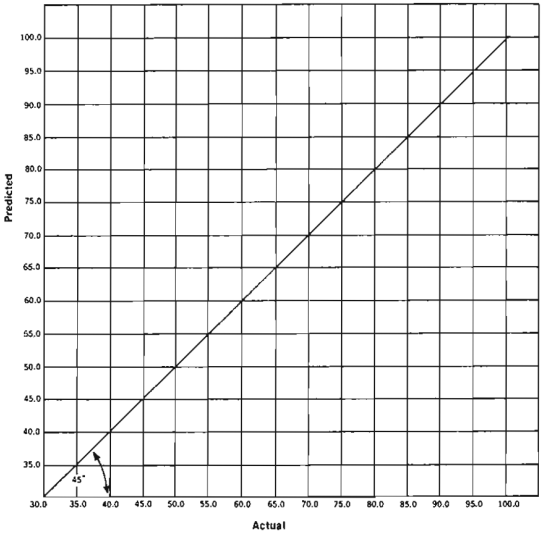


Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

Theee Versions of the Same Scatter Diagram

Null data-ink ratio

Figure 19.1 Relationship of Actual Rates of Registration to Predicted Rates (104 cities, 1960)



How to Improve the Data-Ink Ratio?

Pre-chart execution:
careful scoping

How to Improve the Data-Ink Ratio?

Pre-chart execution: careful scoping

!! Maximize the data-ink ratio !!

Every bit of ink on a graphic requires a reason — and the reason should be that the ink presents new information.

How to Improve the Data-Ink Ratio?

Pre-chart execution: careful scoping

!! Maximize the data-ink ratio !!

Every bit of ink on a graphic requires a reason — and the reason should be that the ink presents new information.

Post-chart execution: critical assessment

How to Improve the Data-Ink Ratio?

Pre-chart execution: careful scoping

!! Maximize the data-ink ratio !!

Every bit of ink on a graphic requires a reason — and the reason should be that the ink presents new information.

Post-chart execution: critical assessment

!! Erase the non data-ink !!

Non data-ink data clutters up the data, as is the case of a thick mesh of grid lines, or gratuitious decorations.

How to Improve the Data-Ink Ratio?

Pre-chart execution: careful scoping

!! Maximize the data-ink ratio !!

Every bit of ink on a graphic requires a reason — and the reason should be that the ink presents new information.

Post-chart execution: critical assessment

!! Erase the non data-ink !!

Non data-ink data clutters up the data, as is the case of a thick mesh of grid lines, or gratuitious decorations.

!! Erase redundant data-ink !!

Bilateral symmetry of data meausures also creates redundancy as in the box plot and the open bar.

Outline

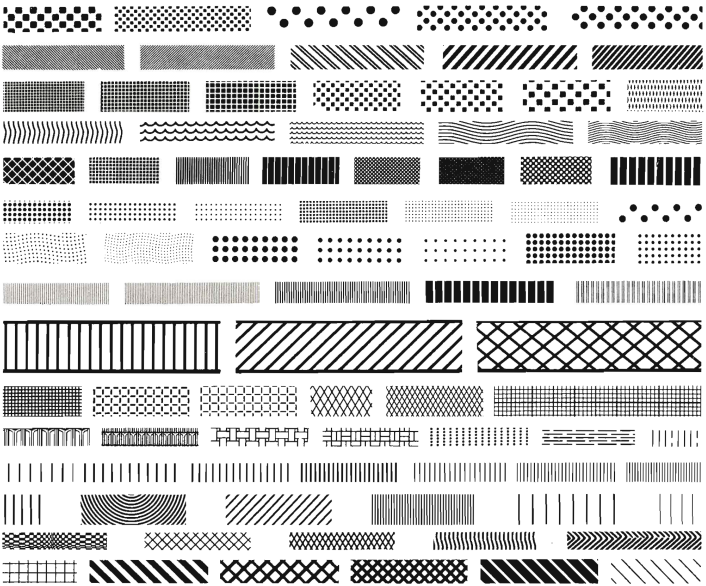
Chartjunk? Give me a Break!!

''The interior decoration of graphics generates a lot of ink that not thell the viewer anything new. The purpose of decoration varies — to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistics skills. Regardless of its cause, it is all non data-data-ink or redundant data-ink, and it is often chartjunk."

— Source is [**tufte2001**]

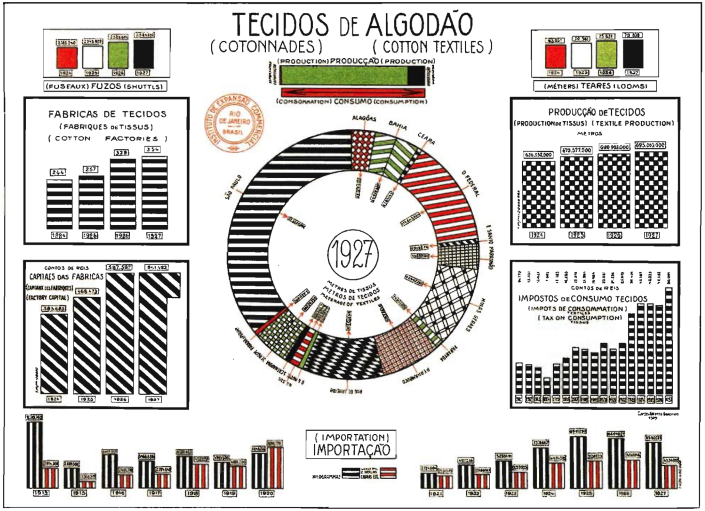
Moiré Effects

...geometric shapes creating a sense of movement



Get Rid of Moiré Effects

Example 1 — Source is "Instituto de Expansao Commercial (1929)"



Get Rid of Moiré Effects

Example 2 — Source is"Kouchoukos et al. (1994), New England Journal of Medicine"

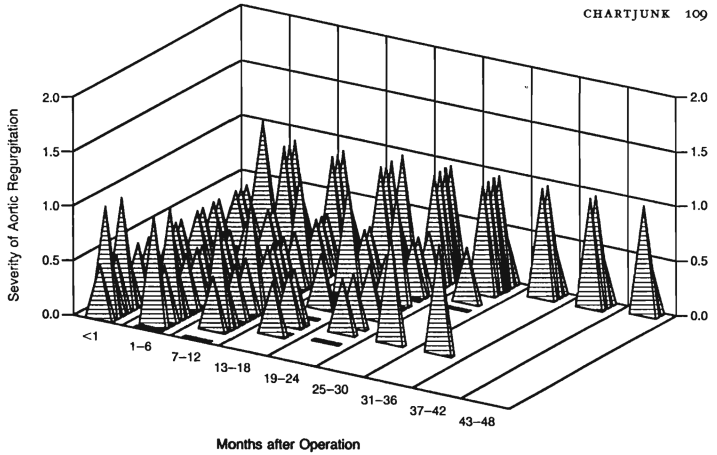
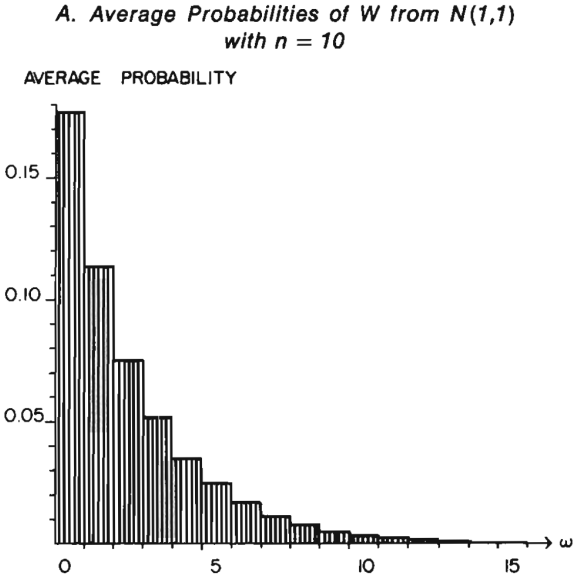


Figure 2. Serial Echocardiographic Assessments of the Severity of Regurgitation in the Pulmonary Autograft in 31 Patients. The numerical grades were assigned according to the severity of regurgitation, as follows: 0, none; 0.5, trivial; 1.0 to 1.5, mild; 2.0, moderate; and 3.0, severe.

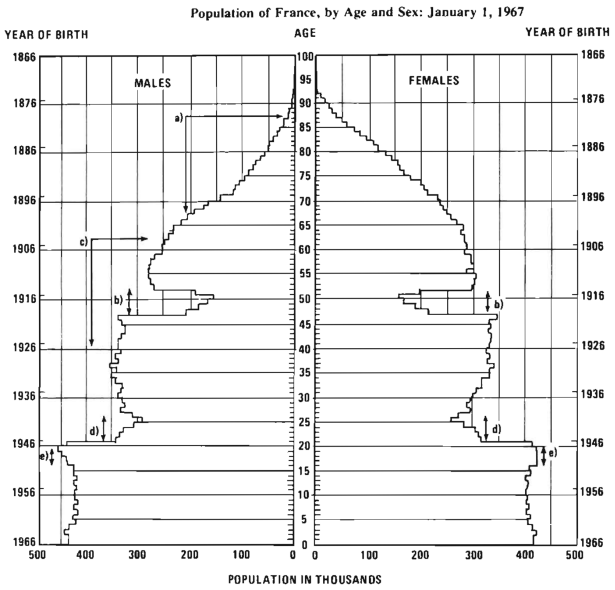
Get Rid of Moiré Effects

Example 3 — Source is "JASA style sheet (1976), Journal of the American Statistical Association"



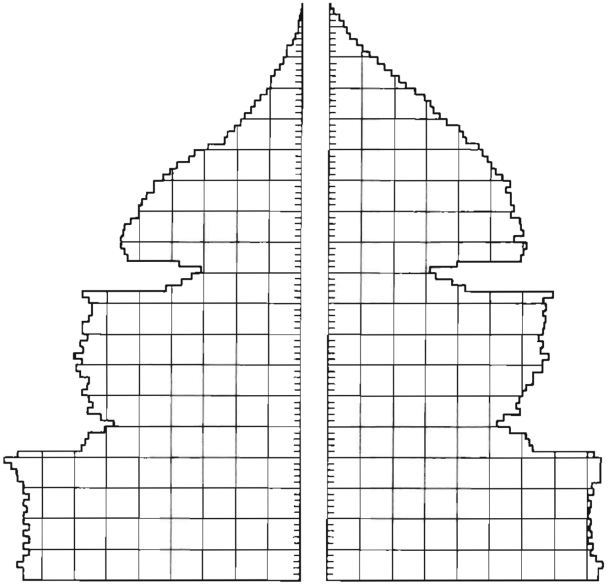
Use Grids Conscientiously

Example — Source is "Institut National de la Statistique et des Études Économiques"



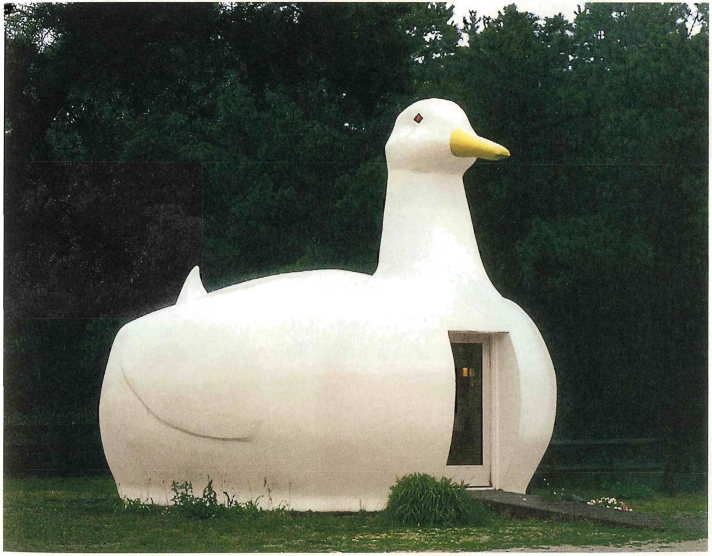
Use Grids Conscientiously

Redesigning the previous example



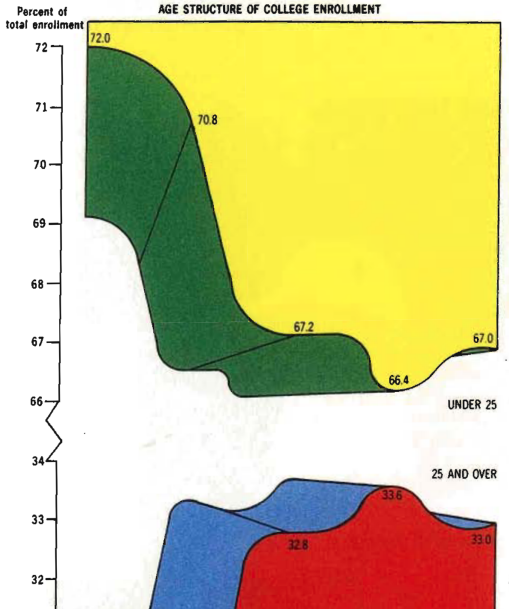
Have You Creted a 'Duck Chart'?

Big Duck, Flanders, New York, photograph by Edward Tufte, July 2000.



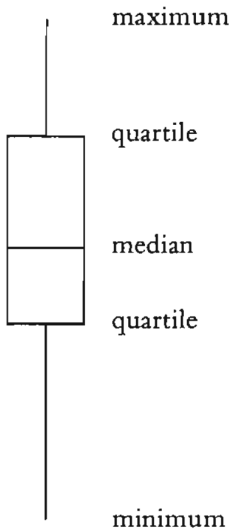
An Example of 'Duck Chart'

Source is [**tufte2001**]

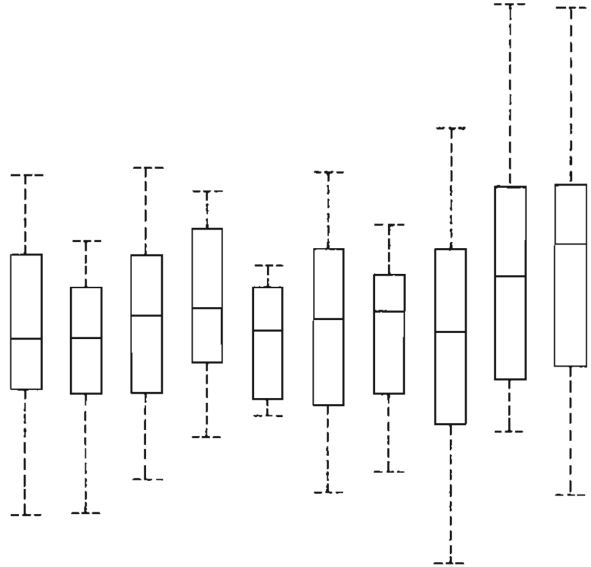


Outline

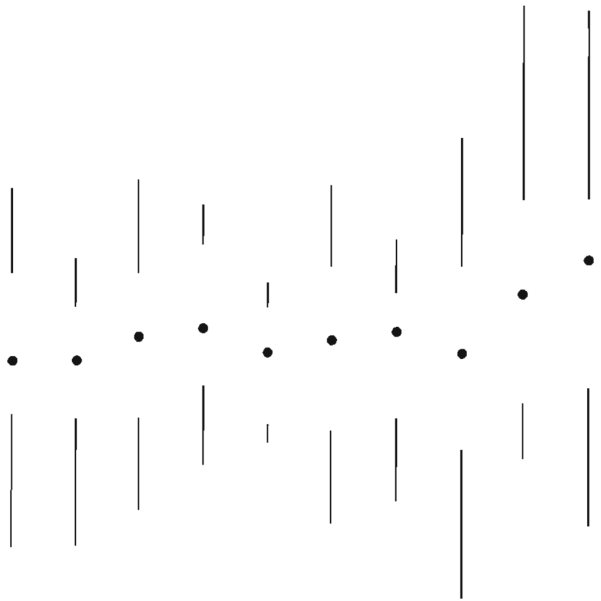
John Tukey's Box Plot



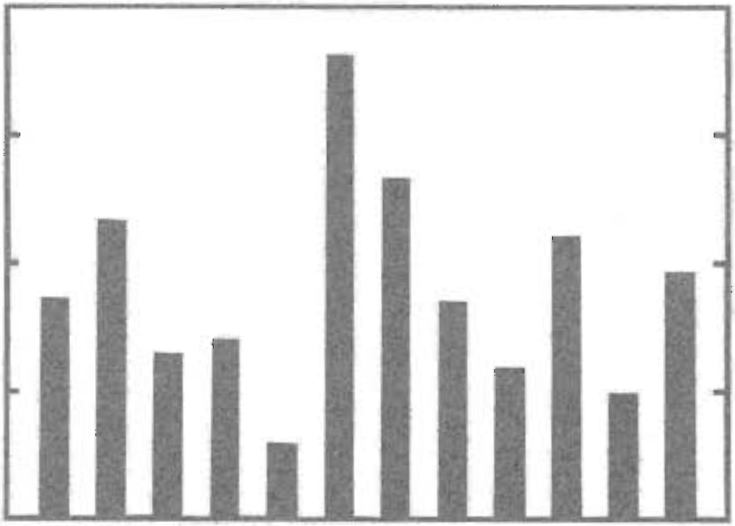
A Box Plot with a Limited Data-Ink Ratio



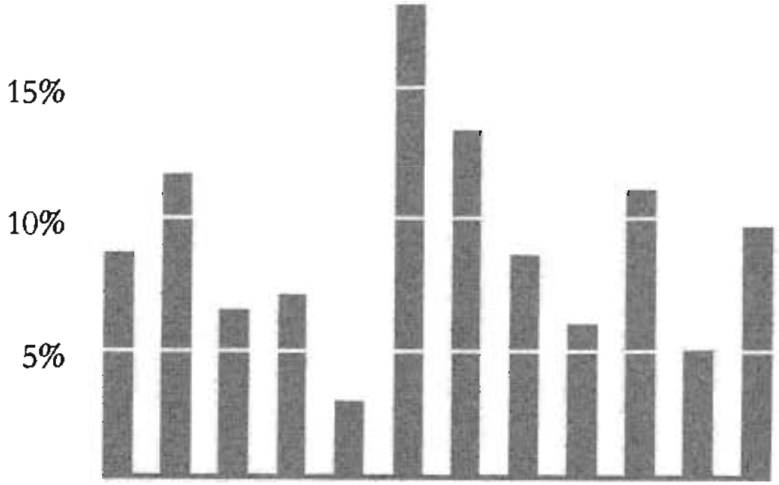
Tufte-Alike Box Plot



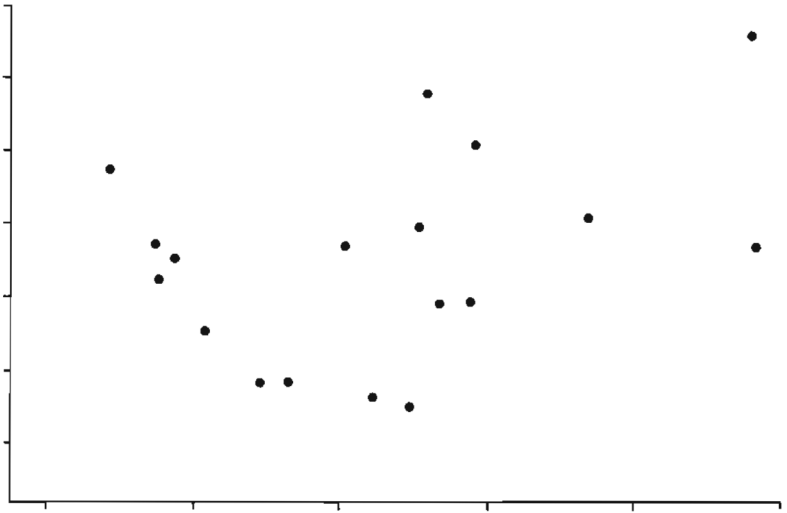
A Bar Chart with a Limited Data-Ink Ratio



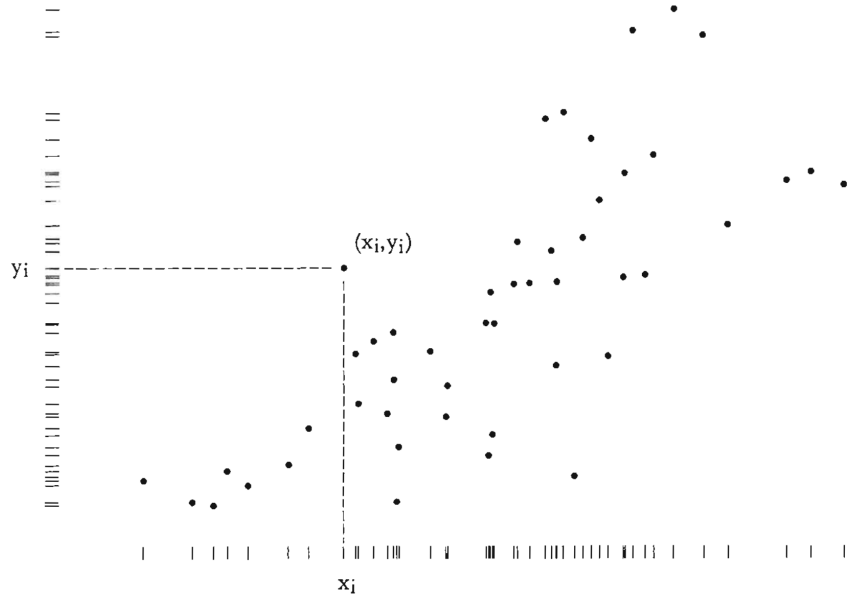
Tufte-Alike Bar Chart



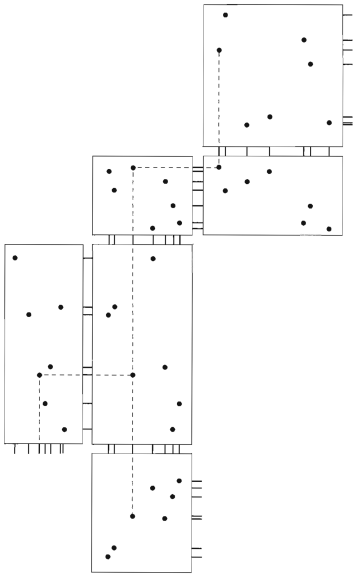
A Bare-Bone Scatter Diagram



Tufte-Alike Scatter Diagram



Tufte-Alike Scatter Diagram



Outline

Data Types Map onto Chart Types!

In SMM635, we focus on the following data types:

- Nominal
- Ordinal
- Quantitative
- Temporal

Nominal Data

Nominal data — also called categorical data — consist of category names.

With nominal data we can compare the equality of values: is value A the same or different than value B? (A = B), supporting statements like "A is equal to B" or "A is not equal to B". In the dataset above, the country field is nominal.

When visualizing nominal data we should readily perceive if values are the same or different: position, color hue (blue, red, green, etc.), and shape are all reasonable options. In contrast, using a size channel to encode nominal data might mislead us, suggesting rank-order or magnitude differences among values that do not exist!

Ordinale Data

Ordinal data consist of values that have a specific ordering.

With ordinal data we can compare the rank-ordering of values: does value A come before or after value B? (A ¡ B), supporting statements like “A is less than B” or “A is greater than B”. In the dataset above, we can treat the year field as ordinal.

When visualizing ordinal data, we should perceive a sense of rank-order. Position, size, or color value (brightness) might be appropriate, whereas color hue (which is not perceptually ordered) would be less appropriate.

Quantitative Data

With quantitative data we can measure numerical differences among values. There are multiple sub-types of quantitative data:

For interval data we can measure the distance (interval) between points: what is the distance to value A from value B? (A - B), supporting statements such as "A is 12 units away from B".

For ratio data the zero point is meaningful, so we can also measure proportions or scale factors: value A is what proportion of value B? (A / B), supporting statements such as "A is 10

In the dataset above, year is a quantitative interval field (the value of year "zero" is subjective), whereas fertility and life$_e$*xpectarequantitativeratiofields*(*zeroismeaningfulforcalculatingproportions*).*Vega* − *Literepresentsquantitativedata*, *butdoesnotmakeadistinctionbetweenintervalandratiotypes*.

Quantitative values can be visualized using position, size, or color value, among other channels. An axis with a zero baseline is essential for proportional comparisons of ratio values, but can be safely omitted for interval comparisons.

Temporal Data

Temporal values measure time points or intervals. This type is a special case of quantitative values (timestamps) with rich semantics and conventions (i.e., the Gregorian calendar).

Example temporal values include date strings such as “2019-01-04” and “Jan 04 2019”, as well as standardized date-times such as the ISO date-time format: “2019-01-04T17:50:35.643Z”. There are no temporal values in our global development dataset above, as the year field is encoded as an integer.

The temporal type in Vega-Lite supports reasoning about time units (year, month, day, hour, etc.), and provides methods for requesting specific time intervals. For more details about temporal data in Vega-Lite, see the TimeUnit documentation.

Outline

Time to wrap up!

References