Chart Families

Ashish Anand

1 Objective

I am making this rolling[1] tutorial discussing different chart types. In visualization task, we face two primary questions:

- what is the primary objective of the visualization: whether we want to explore data distribution; whether we want to show hierarchical relations or temporal pattern etc.

- what are the different possible charts one can use for the identified objective and which chart type among all possible options will be the most appropriate choice for conveying the message in a comprehensive manner.

Aim of the tutorial is to help you in answering the two questions based on the organization of chart families as discussed in [1].

2 Chart Families

Different charts offer distinct ways to represent and visualize data. In [1], Kirk has categorized the charts into **five** main families based on their primary objective. The five families are -

1. **Categorical:** For comparison of categories and distributions of quantitative values

2. **Hierarchical:** Revealing *part-to-whole* relationships and hierarchies

3. **Relational:** Exploring correlations and connections

4. **Temporal:** Plotting trends and intervals over time

5. **Spatial:** Mapping spatial patterns through overlays and distortions

Kirk has given a five letter mnemonic **CHRTS** representing these chart families [1].

Now, we will discuss representative charts from each of the families. It is noteworthy to mention that -

---

[1]I will keep updating this tutorial

- some charts are likely to fall into more than one families, however, it was kept in one which suits the primary objective of the chart most. It does not mean that the particular chart cannot or should not be used in other scenario. We will highlight this aspect while discussing such charts.

- you can add additional *presentation* aspects such as "making it interactive", "adding texts and legends" etc.

Since, during the course of assignments, we already have explored several chart types belonging to **categorical** family, I will start the tutorial with discussion of **temporal** family.

2.1 Temporal Family

Charts from this family are primarily used to show trends, events/actions or activities at a set of time-intervals. Let us look at some of the examples.

Line Chart

A line chart displays how quantitative values have changed over time for different categorical items. A line chart is drawn around a **continuous temporal x-axis** and **quantitative y-axis** with values plotted at relevant coordinates. Connecting lines join up adjacent and related categorical items to form slopes which are then extended along the full timescale.

Fig 1 shows the the cumulative runs scored in Test matches by English Test batsmen between 1947 and 2018.
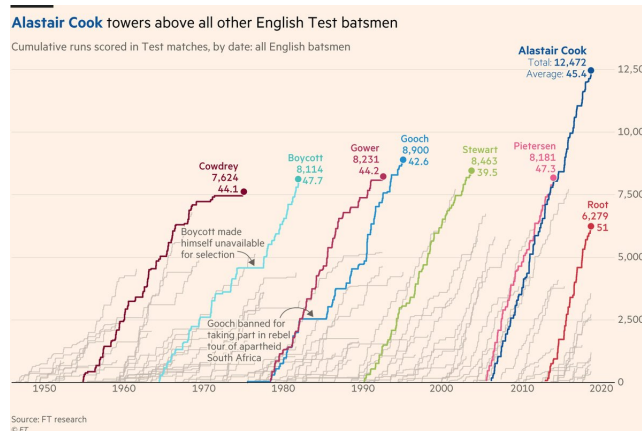

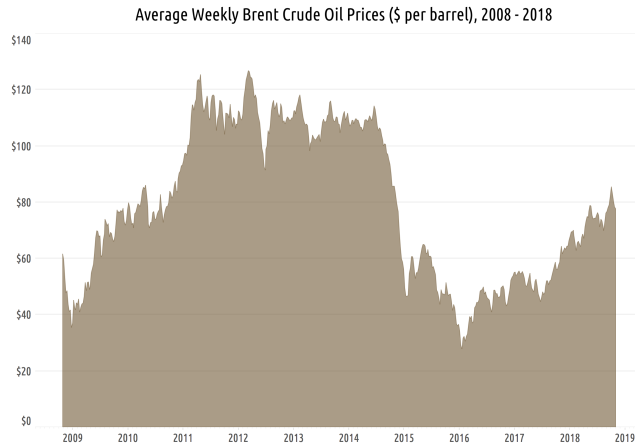Figure 1: Runs scored by English Test batsmen between 1947 and 2018. Source: Andy Kirk, Chapter 6 [1]

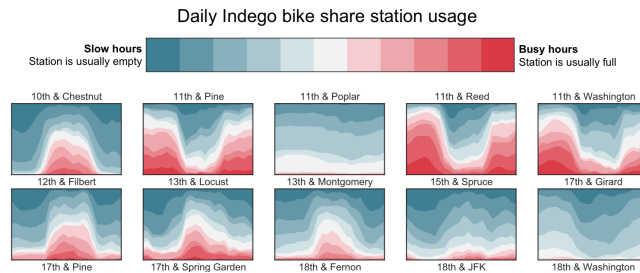Figure 2: Crude oil price for Brent (US$ per barrel) 2008-2018. Source: Andy Kirk, Chapter 6 [1]

Figure 3: Source: Andy Kirk, Chapter 6 [1]

Area Chart

Area chart is an extension of line chart, with the area between the x-axis (representing time) and the line (quantitative measure).

Fig. 2 shows changes in the average weekly price (US$ per barrel) of Brent crude oil between 2008 and 2018.

Stacked Area Chart

Stacked area chart is an area chart displaying how quantitative values have changed over time for multiple categorical items. The idea is very similar to stacked bar chart and allows visualization of evolution of both total and contribution of component sub-categories. However, if there are too many sub-categories, then one can think of using **Area Chart** with **faceting**.

These are the 15 most important political problems in Germany

The chart shows which topics Germans are the most active in this general election and what significance they had in previous elections.
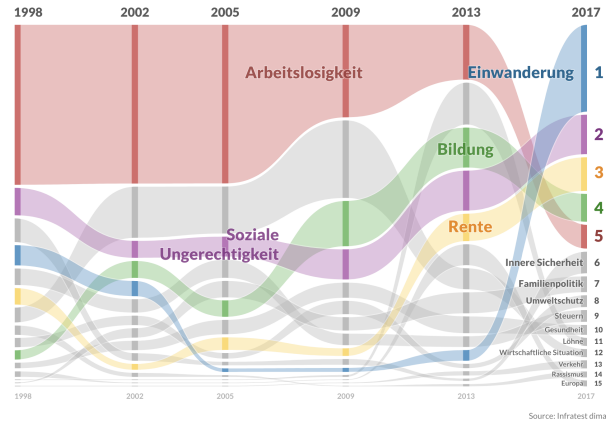
Figure 4: Changes in rank of the most politically important topics for Germans between 1998 and 2017. Source: Andy Kirk, Chapter 6 [1]

Bump Chart

Bump chart is an alternative to Line chart, when we want to show how quantitative values, in terms of *ranking measurement*, have changed over time for different categorical items. In this case, we keep rank on the Y-axis and X-axis represents continuous time scale.

Bump chart usually uses variation in the size (width) of each line to represent a quantitative measure, usually the absolute value informing the ranking measurement (Fig. 4).

Slope Graph

Slope Graph is used to show trend of change of quantitative values during the two time points for different category items. The graph consists of two parallel quantitative axes with a common value range. A line is plotted connecting the two axes together with the vertical position on axes representing the respective quantitative values. These connecting lines form slopes that indicate the upward, downward, or stable trend between the two temporal axes.

Fig. 5 shows an example of slope graph comparing the changes in the share of power sources across US states between 2004 and 2014.
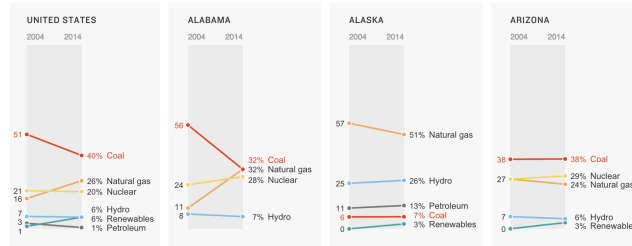
4

How Each State Generates Electric Power (2004-2014)

Figure 5: How US states generates power from different sources: Coal, gas, nuclear, hydro? Source: Andy Kirk, Chapter 6 [1]

Stream Graph

Stream Graphs display the changes in data over time of different category items. It uses flowing, organic shapes that somewhat resemble a river-like stream. The axis that a Stream Graph flows parallel to, is used for the timescale.

In a Stream Graph, the size of each individual stream shape is proportional to the values in each category. Often more than one attributes of the category item are shown. The area occupied is filled with an attribute of color to represent another categorical attribute. For example, in the Fig. **??** changes in the total domestic gross takings (in US$) and the longevity of all movies released between 1986 and 2008 are shown. Height of the stream indicates weekly box office revenues and width shows longevity at the box office. The area (and associated color) indicates total domestic gross. Instead of plotting values against a fixed,

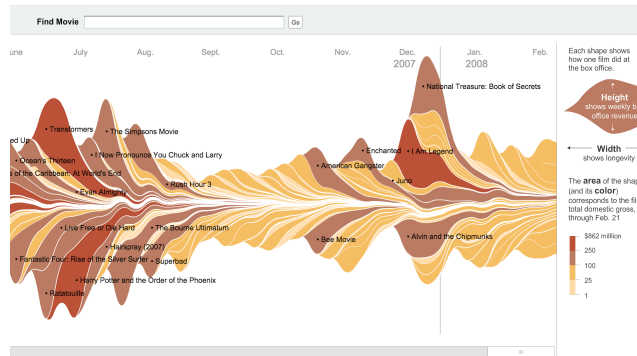The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Figure 6: Changes in the total domestic gross and the longevity of all movies released between 1986 and 2008. Source: Andy Kirk, Chapter 6 [1]

straight axis, a Stream Graph has values displaced around a varying central baseline.

Figure 7: Changes in the daily price and availability of Super Bowl tickets on the secondary market. Source: Andy Kirk, Chapter 6 [1]

Category items may be plotted up and down the implied y-axis, in order to optimise the layout and not to reflect any notion of positive or negative values.

Connected Scatter Plot

Connected scatter plot is an extension of joined version of *line chart* and *scatter plot*. It has two variations. In one of the variation, *line chart* and *scatter plot* are combined together to show the relation between two quantitative variables.

The second variation is used to compare two quantitative variables over a time series. Thus, as an input this variation requires three variables. The chart is formed of two quantitative x- and y-axes, and with the values represented by point marks at the respective coordinates, one for each measurement over time. The individual points are then connected using lines joining each consecutive time points to form a sequence of change.

Fig. 7 displays changes in the daily price (variable 1) and availability of Super Bowl tickets (variable 2) on the secondary market four weeks prior to the event across five Super Bowl finals.

Gantt Chart

Gantt chart is used to visualize sub-tasks/sub-events of an event or a project over time. Typically, x-axis is used to represent the continuous time intervals and sub-tasks/sub-events on the y-axis. Corresponding to each item on the y-axis, size (length) of a horizontal bar indicates the duration of the activity. Horizontal bars corresponding to the different activities may have different start

and end time points. It can be further extended to visualize dependencies among several sub-tasks.
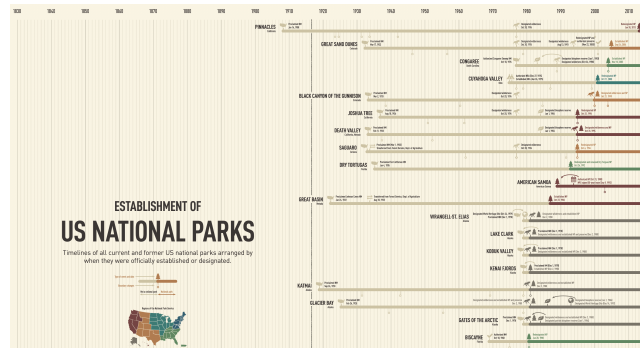

Figure 8: Establishment of the U.S. National Parks. Source: Andy Kirk, Chapter 6 [1]

References

[1] Andy Kirk. *Data Visualisation: A Handbook for Data Driven Design.* Sage, London, 2019.