

Information Visualization 1

Abgabe über die NextCloud bis 23:59 Uhr des o.g. Datums.

Aufgabe 1 Pareto Chart

(5 Punkte)

In `task10_1.py` wird eine Tabelle mit Daten zu CO₂-Emissionswerten von 223 Ländern geladen. In der Spalte `Country` finden sich die Ländernamen. In der Spalte `Emissions` finden sich die kumulativen Emissionswerte aller Jahre von 1751 bis 2017.¹ In dieser Aufgabe soll mit einem *Pareto Chart* visualisiert werden, welche Länder den größten Anteil an den kumulativen Emissionen haben.

a) (2 Punkte)

Plotten Sie die Daten der 20 Länder mit den höchsten Emissionswerten in absteigender Reihenfolge als Balkendiagramm via `axis.bar(x,y)`. Die y-Achse soll dabei in Milliarden Tonnen skaliert sein (die Werte in der Tabelle sind in Tonnen, d.h. der Faktor ist 10^{-9}).

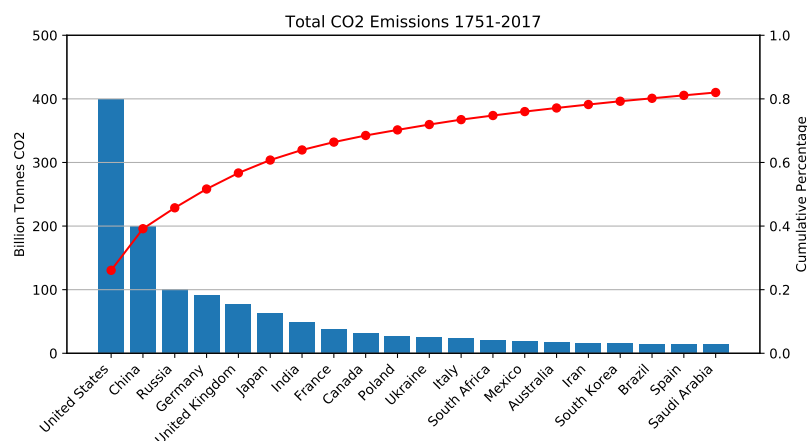
b) (3 Punkte)

Zeichnen Sie die Linie ein, die im Pareto Chart den kumulativen Anteil der absteigend sortierten Komponenten zeigt. Damit Balken und Linie erkennbar sind, soll eine zweite y-Achse verwendet werden, die zwischen 0 und 1 skaliert ist. Eine solche Achse wurde bereits mit `ax2 = ax1.twinx()` erstellt und anschließend korrekt skaliert. Wird beim Erzeugen der Linie der Parameter `'-ro'` übergeben, erhält man eine rote Linie mit Punkten an den Datenwerten. Dadurch wird die Darstellung übersichtlicher.

Hinweise:

- Aus einem *pandas dataframe* lässt sich eine Spalte mit ihrem Namen `dataframe['Name']` extrahieren.
- Außerdem nützlich ist der `.iloc[]`-Operator, sowie `.sort_values()` und `.sum()`:
 - <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.iloc.html>
 - https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.sort_values.html
 - <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.sum.html>
- Die Numpy-Funktion `cumsum()` darf verwendet werden (siehe Cheatsheet *Useful build-in functions*).

Das Ergebnis sieht so aus:



¹Es handelt sich um reale Daten von <https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions>.

Aufgabe 2 Theorie

(10 Punkte)

Geben Sie die Antworten auf die Theorieaufgaben in der Multiple-Choice-Datei `MC10.txt` an. Es ist immer genau eine Auswahlmöglichkeit richtig. Bitte keine anderen Anmerkungen in diese Datei schreiben und den Dateinamen nicht verändern.

a) (1 Punkt)

Welche Aussage bezüglich des Pareto Charts aus Aufgabe 1 ist wahr?

- (a) Alle Emissionen kombiniert entsprechen in etwa 500 Milliarden Tonnen.
- (b) Chinas Anteil an den globalen Emissionen entspricht 40%.
- (c) Die vier Länder mit den meisten Emissionen sind für über 50% der Gesamtemissionen verantwortlich.

b) (1 Punkt)

Als Attribute bezeichnet man die Eigenschaften eines Informationsobjektes.

- (a) Wahr.
- (b) Falsch.

c) (1 Punkt)

Was ist *keine* Interaktionsaufgabe?

- (a) Unwichtige Datenbereiche Herausfiltern.
- (b) Sich einen Überblick verschaffen.
- (c) Teile der Daten für Weiterverarbeitung extrahieren.
- (d) Die Beziehung zwischen zwei Datenobjekten betrachten.
- (e) Sicherstellen, dass die Framerate konstant bleibt.

d) (1 Punkt)

Die Bilder einer Überwachungskamera fallen in welche der folgenden Kategorien?

- (a) Univariate Daten.
- (b) Bivariate Daten.
- (c) Multidimensionale Daten.

e) (1 Punkt)

Eine Temperaturmessung einer Wetterstation fällt in welche der folgenden Kategorien?

- (a) Univariate Daten.
- (b) Multivariate Daten.
- (c) Multidimensionale Daten.

f) (1 Punkt)

Es ist generell sinnvoll die Balken eines Histogramms absteigend nach ihrer Größe zu sortieren, um einen besseren Überblick zu erhalten.

- (a) Wahr.
- (b) Falsch.

g) (1 Punkt)

Warum kann die Hauptdiagonale einer Scatterplot Matrix für etwas anderes als einen Scatterplot genutzt werden?

- (a) Ein Attribut würde hier mit sich selbst verglichen werden.
- (b) Der Scatterplot lässt sich an dieser Stelle nicht berechnen.
- (c) Die an diesen Stellen dargestellten Beziehungen kann man auch an den anderen Scatterplots ablesen.

h) (1 Punkt)

Der Nachteil von parallelen Koordinaten ist, dass sie nur für maximal 5 Dateneinträge genutzt werden können.

- (a) Wahr.
- (b) Falsch.

i) (1 Punkt)

Die Achsen von parallelen Koordinaten können in beliebiger Reihenfolge angeordnet werden.

- (a) Wahr.
- (b) Falsch.

j) (1 Punkt)

Welche der folgenden Techniken eignet sich *nicht* um *visual clutter* bei parallelen Koordinaten zu reduzieren?

- (a) Brushing.
- (b) Clustering.
- (c) Normalization.