NLP2Chart

Introduction

In this evaluation, we want to investigate the use of NLP for the user interface of information visualisation software. For this purpose, we created two versions of a prototype of a web application. We will give some tasks to be solved with both versions. These tests will be recorded using Screencapture.

You can use the software https://atomisystems.com/download/ for this purpose, for example. Or record a private session with Zoom.

Answer a few questions to get started

https://docs.google.com/forms/d/1vlX_5N6wsAWBEKCOZUOTS86iCyQFo42Ccb5FaLFEqQM

Watch the Tutorial Videos

Familiarise yourself with both versions of the application by watching the following two videos.

Prototype:: https://youtu.be/UiCSczhslAs
Prototype V: https://youtu.be/cuR1U0Y1vJQ

Preparation

The prototype is at https://share.streamlit.io/astoeckl/nlp2chart2/main.py
The prototype V is at https://share.streamlit.io/astoeckl/nlp2chart2---vegalite/main.py

Data:

We use data from Gapminder, a foundation set up by Hans Rosling to provide data and software tools that promote an evidence-based worldview, for the evaluation of our approach.

Download the Data from:
https://github.com/astoeckl/NLP2Chart-Data

The data is available as a CSV file. The first line contains the name of the feature, each further line a data record. Each data record consists of:

- Country - String
- Continent - String
- Year- Date / Int
- Life expectancy - lifeExp - Float
- Population size - pop - Int
- GDP per capita - gdpPercap - Float

	A	B	C	D	E	F
1	country	continent	year	lifeExp	pop	gdpPercap
2	Afghanistan	Asia	1952	28.801	8425333	779.4453145
3	Afghanistan	Asia	1957	30.332	9240934	820.8530296
4	Afghanistan	Asia	1962	31.997	10267083	853.10071
5	Afghanistan	Asia	1967	34.02	11537966	836.1971382
6	Afghanistan	Asia	1972	36.088	13079460	739.9811058
7	Afghanistan	Asia	1977	38.438	14880372	786.11336
8	Afghanistan	Asia	1982	39.854	12881816	978.0114388
9	Afghanistan	Asia	1987	40.822	13867957	852.3959448
10	Afghanistan	Asia	1992	41.674	16317921	649.3413952
11	Afghanistan	Asia	1997	41.763	22227415	635.341351
12	Afghanistan	Asia	2002	42.129	25268405	726.7340548
13	Afghanistan	Asia	2007	43.828	31889923	974.5803384
14	Albania	Europe	1952	55.23	1282697	1601.056136
15	Albania	Europe	1957	59.28	1476505	1942.284244
16	Albania	Europe	1962	64.82	1728137	2312.888958
17	Albania	Europe	1967	66.22	1984060	2760.196931

Starting the video recording

Open each of the two web applications in a tab of your web browser and import the gapminder-data.csv file.

Exercises:

Get an overview of the distributions of the numerical characteristics with the following steps.

Exercise 1)

Create histograms for the numerical characteristics.

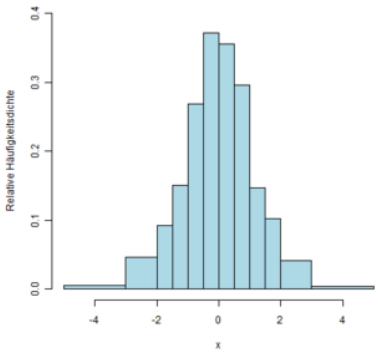
These are in this task:

- lifeExp
- pop
- gdpPercap

Format each of the 3 graphs by:

- Assigning the title "Histogram".
- Take 50 bars for the subdivision ("Bins").
- Colouring the bars with "Green".
- Labelling the y-axis with "Count".
- Labelling the x-axis with "Feature".

A histogram is a graphical representation of the following form:



(Source: Wikipedia)

Exercise 2)

Plot the numerical characteristics per continent using bar charts.

These are in this task:

- lifeExp
- pop
- gdpPercap

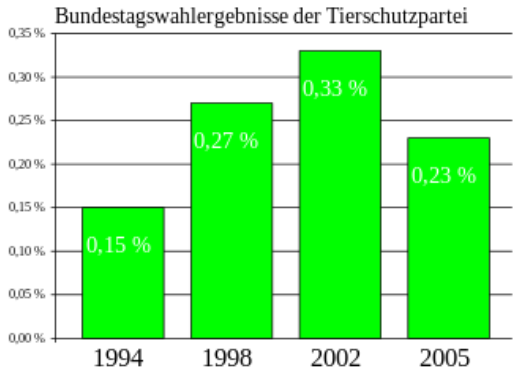
Grouped according to:

- continent

Format each of the 3 graphs by:

- Titling them "Life Expectancy", "Population" and "GDP".
- Colour the bars with "Green".

A bar chart is a graphical representation of the following form:



(Source: Wikipedia)

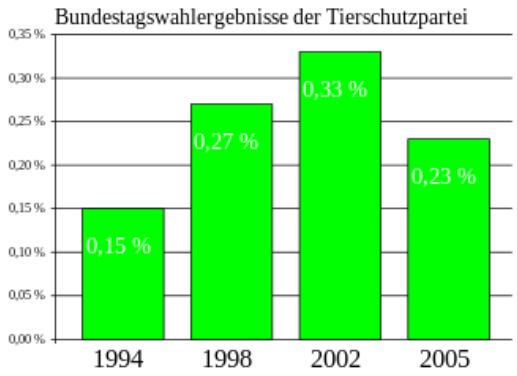
Exercise 3)

Display the GDP per country with the help of a bar chart.
The top 15 countries should be displayed in descending order.

Format the graph by:

- Give "Top GDP Countries" as the title.
- Colour the bars with "Green".

A bar chart is a graphical representation of the following form:

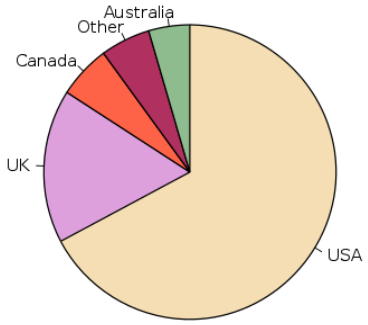


(Source: Wikipedia)

Exercise 4)

Create a pie chart that shows the distribution of the world's population by continent.

A pie-chart is a graphical representation of the following form:



(Source: Wikipedia)

Let us now look at the temporal course of the numerical characteristics.

Exercise 5)

Create a graph that shows the time course of mean life expectancy averaged over all countries. Use a line chart for this.

Format the graph by:

- Assigning the title "Life Expectancy per Year".
- Colouring the line with "Green".
- Using a dashed line.
- Label the y-axis with "LifeExp".
- The lower limit of the y-axis starts at 0.
- The upper limit of the y-axis ends at 80.
- Label the x-axis with "Year".

A line chart is a graphical representation of the following form:

(Source: Wikipedia)

Find correlations in the data.

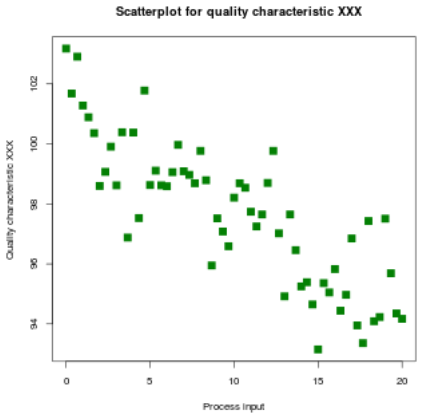
Exercise 6)

Show the relationship between GDP per inhabitant of a country and life expectancy with the help of a scatter diagram. In addition, use a colour code that reflects the population size.

Format the graph by:

- Assigning the title "LifeExp versus GDP".

A scatterplot chart is a graphical representation of the following form:



(Source Wikipedia)

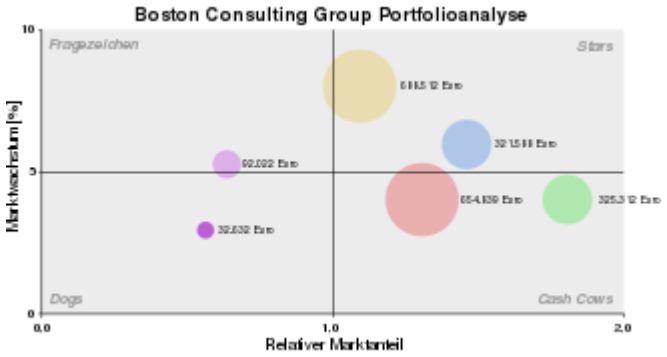
Exercise 7)

Show the relationship between GDP per inhabitant of a country and life expectancy and population size using a bubble chart.

Format the graph by:

- Assign the title "LifeExp versus GDP versus Population".

A bubble chart is a graphical representation of the following form:



(Source Wikipedia)

Post Questions

Complete the following questionnaire separately for both versions:

Normal Version:

https://docs.google.com/forms/d/1q9qUtpKjdDeoEqLwz9gA0ESGJhtJHokhMXAu4jkTUmM

V Version:

https://docs.google.com/forms/d/1lXiRJEKmodNsGD5Sr-OM45o7blWw3uEHoggscWX3lr4