Applied Data Analytics

Visualisations with plotly

Line graphs vs. bar charts

Hans-Martin von Gaudecker and Aapo Stenhammar

Picking the right graph

All kinds of data:

- Histograms: Bar charts, potentially stacked

Cardinal data:

- Levels of variable with fractional scale: Bar charts (including zeros!)
- Changes: Line charts

Never

- Pie charts

```
import plotly.express as px

df = px.data.gapminder()
df = df.query("continent=='Africa' & pop > 10_000_000")
df = df[["country", "year", "lifeExp"]]
life_exp = df.pivot(index='year', columns='country', values='lifeExp')
life_exp = life_exp.dropna(axis="columns", how="any")

life_exp.round(1)
```

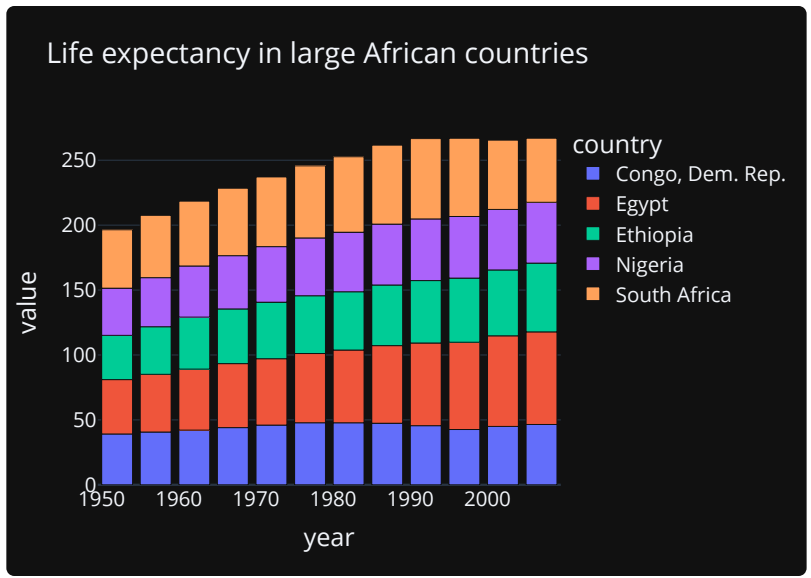
year	Congo, Dem. Rep.	Egypt	Ethiopia	Nigeria	South Africa
1952	39.1	41.9	34.1	36.3	45
1957	40.7	44.4	36.7	37.8	48
1962	42.1	47	40.1	39.4	50
1967	44.1	49.3	42.1	41	51.9
1972	46	51.1	43.5	42.8	53.7
1977	47.8	53.3	44.5	44.5	55.5
1982	47.8	56	44.9	45.8	58.2
1987	47.4	59.8	46.7	46.9	60.8
1992	45.5	63.7	48.1	47.5	61.9
1997	42.6	67.2	49.4	47.5	60.2
2002	45	69.8	50.7	46.6	53.4
2007	46.5	71.3	52.9	46.9	49.3

Features of the data

- Homogenous data
  - All entries are life expectancies
  - This fact is implicit, cannot tell from the table
  - Contrast with survey data on sex, age, income, etc.
- Cardinal data on a fractional scale
- No missing values

Out-of-the-box bar chart

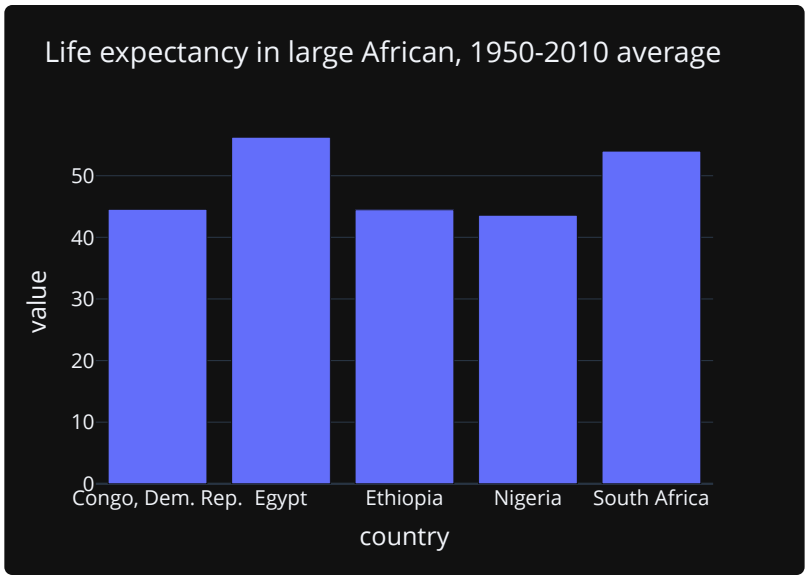
```
fig = life_exp.plot.bar(
  title='Life expectancy in large African
)
fig.show()
```



Plot levels as bars

```
fig = life_exp.mean().plot.bar(
    title='Life expectancy in large Africa
)
fig.update_layout(showlegend=False)
fig.show()
```

Note that bars have to start at zero!



Plot changes as lines

```
fig = life_exp.plot.line(
    title='Life expectancy in large Africa
)
fig.show()
```

Focus on changes, hence focus the scale of the y-axis on the relevant range