Data Visualisation

Scientific Programming with Python

Christian Elsasser
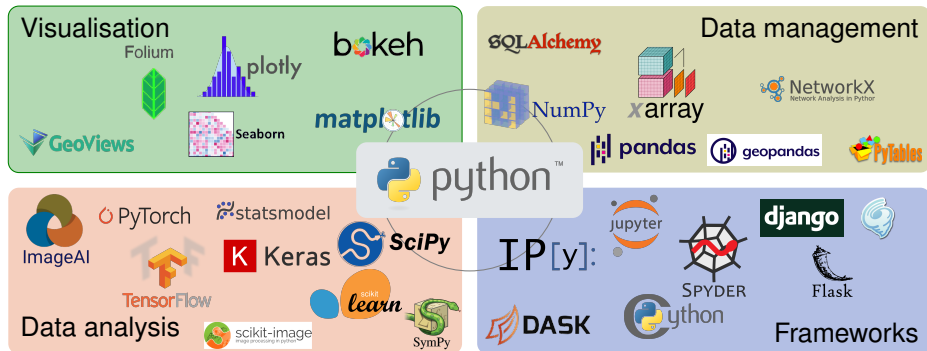
Python offers a large ecosystem for scientific analytics and beyond

Visualisation is too often treated as an afterthought

80-90% of the information is consumed visually by humans

The human brain processes visuals 1'000 to 10'000$\times$ faster than text

Almost everybody will read this last! (if at all)

You will read this first
because it is in a colored box and big!

You will read this probably second!

And then most of the people read this!

Three simple guidelines might help to achieve an optimised way of creating visualisations
based on Jean-luc Dummont "Trees, maps, and theorems"

1. **Adapt to the audience and purpose**
   Take into account the audience's pre-knowledge and how the visualisation is consumed.

2. **Maximise the signal-to-noise ratio**
   Limit distraction as much as possible on your visualisation.

3. **Use effective redundancy**
   Leverage multiple characteristics and channels to convey the insights and data.

Three simple guidelines might help to achieve an optimised way of creating visualisations

based on Jean-luc Dummont "Trees, maps, and theorems"

1. **Adapt to the audience and purpose**
   Take into account the audience's pre-knowledge and how the visualisation is consumed.

2. **Maximise the signal-to-noise ratio**
   Limit distraction as much as possible on your visualisation.

3. **Use effective redundancy**
   Leverage multiple characteristics and channels to convey the insights and data.

Three simple guidelines might help to achieve an optimised way of creating visualisations
based on Jean-luc Dummont "Trees, maps, and theorems"

1. **Adapt to the audience and purpose**
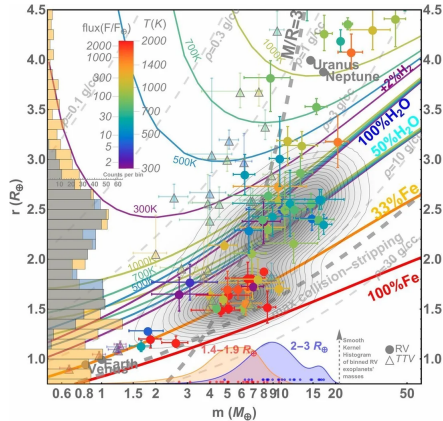   Take into account the audience's pre-knowledge and how the visualisation is consumed.

2. **Maximise the signal-to-noise ratio**
   Limit distraction as much as possible on your visualisation.

3. **Use effective redundancy**
   Leverage multiple characteristics and channels to convey the insights and data.

Three simple guidelines might help to achieve an optimised way of creating visualisations

based on Jean-luc Dummont "Trees, maps, and theorems"

1. **Adapt to the audience and purpose**
   Take into account the audience's pre-knowledge and how the visualisation is consumed.

2. **Maximise the signal-to-noise ratio**
   Limit distraction as much as possible on your visualisation.

3. **Use effective redundancy**
   Leverage multiple characteristics and channels to convey the insights and data.

Three simple guidelines might help to achieve an optimised way of creating visualisations

based on Jean-luc Dummont "Trees, maps, and theorems"

1. **Adapt to the audience and purpose**
   Take into account the audience's pre-knowledge and how the visualisation is consumed.

2. **Maximise the signal-to-noise ratio**
   Limit distraction as much as possible on your visualisation.

3. **Use effective redundancy**
   Leverage multiple characteristics and channels to convey the insights and data.

Example 1: Couldn't there be any more information in this chart?

Source: Zeng et al. *Growth model interpretation of planet size distribution*, PNAS **116** (2019), 20
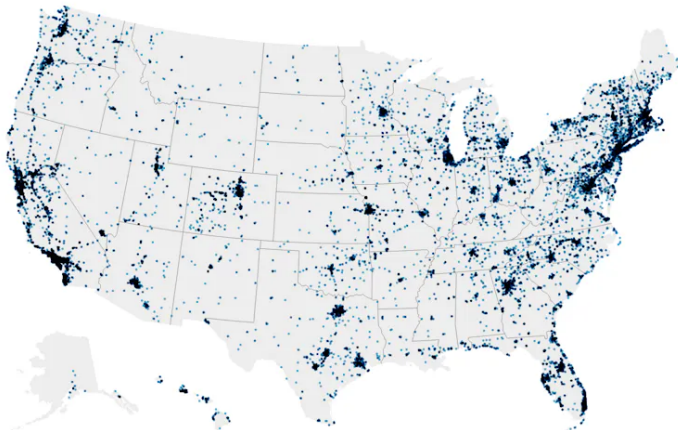
Example 2: What is the purpose?
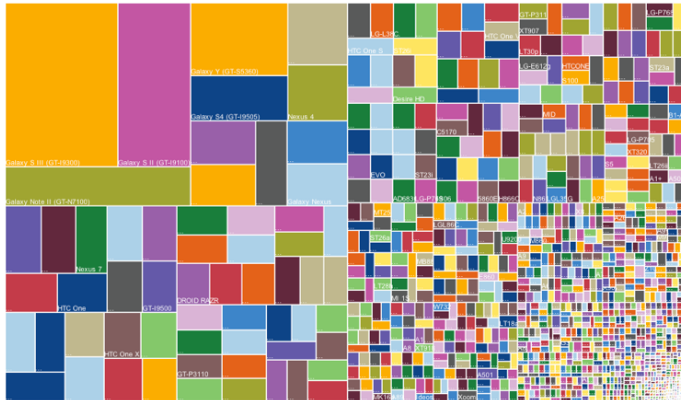
Source: Open Signal
Distribution of the Open Signal App among different smartphones
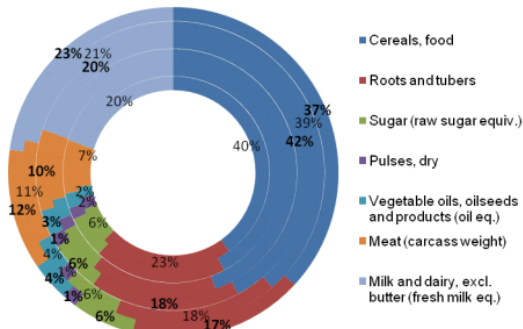
Example 3: Be mindful of the data representation

Source: US Department of Energy
Original caption: [EV] Charging stations as of August 2021. Darker areas are clusters of many stations, primarily in large cities.

Example 4: Be clear what message you want to convey

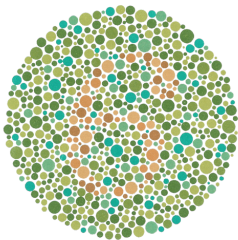World Dietary shares: (from inside to outside) 1970, 1980, 1990, 2000, 2030, 2050

- Cereals, food
- Roots and tubers
- Sugar (raw sugar equiv.)
- Pulses, dry
- Vegetable oils, oilseeds and products (oil eq.)
- Meat (carcass weight)
- Milk and dairy, excl. butter (fresh milk eq.)

Source: Northern Ireland Assembly

Note: figures for 1980 and 1990 shares are not shown for sake of clarity.

Example 5: Without words. . .
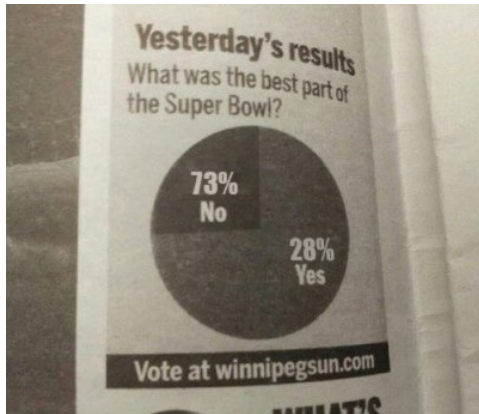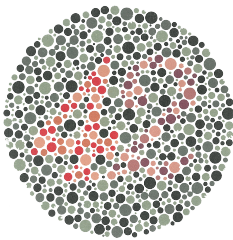
Source: Winnipeg Sun

Dedicated color schemes can be used to make colors better distinguishable for colorblind people and when read in black & white

6

42

- ▶ 8% of Caucasian, 5% of Asian, and 4% of African males are so-called "red-green" Colorblind.
- ▶ Many papers are still printed in black & white.

- ▶ Matplotlib has suitable color gradients like `viridis`, `cividis`, `magma`, `inferno`, `plasma`
- ▶ There are helpful tools like **monolens** to test the suitability of visualisation for such situations

Different visualisation dimensions allow us to create redundancy and emphasis
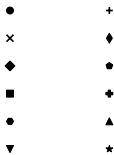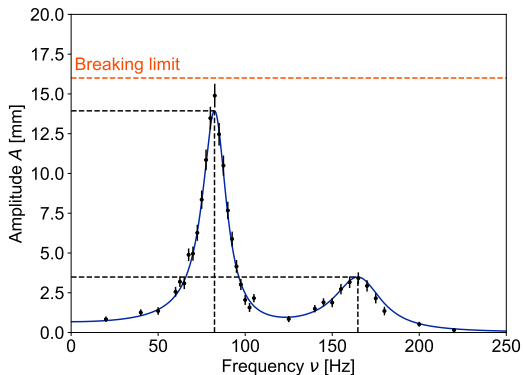
Size   **Line style**   **Fill style**   **Marker shape**

But ensure that you don't overload a visualisation as shown in the examples before!

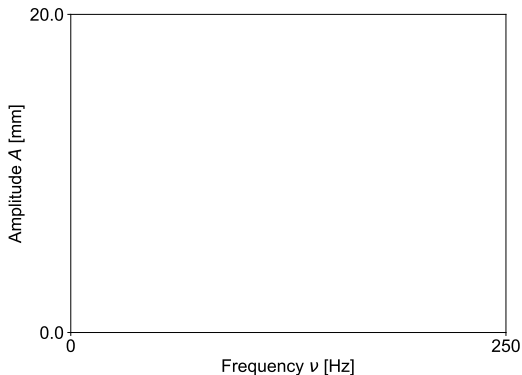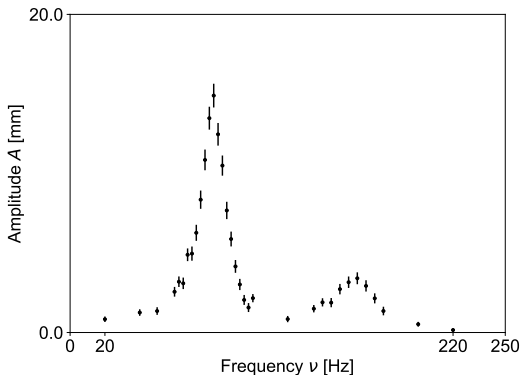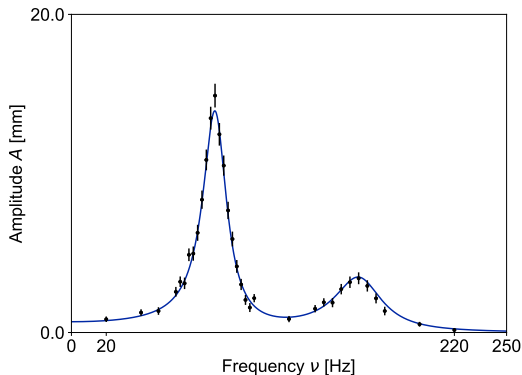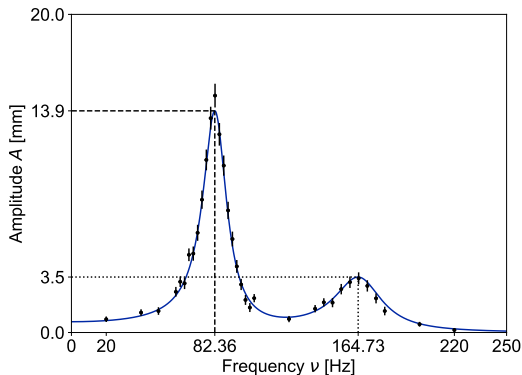# It does not take much to make better charts – and Python libraries can help us

It does not take much to make better charts – and Python libraries can help us
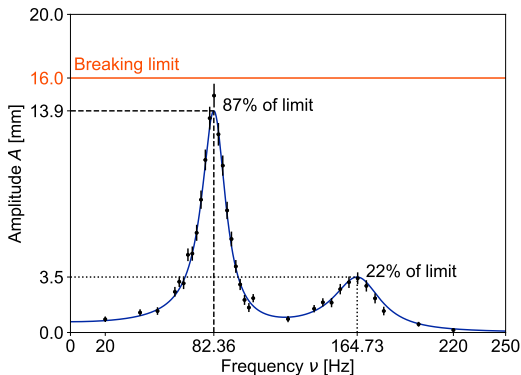
It does not take much to make better charts – and Python libraries can help us

It does not take much to make better charts – and Python libraries can help us

It does not take much to make better charts – and Python libraries can help us

It does not take much to make better charts – and Python libraries can help us

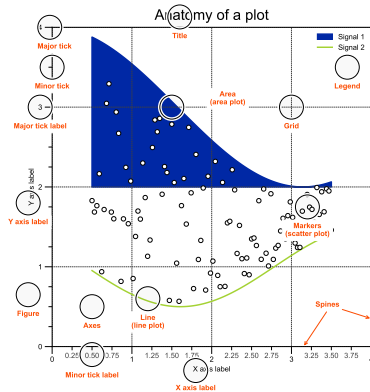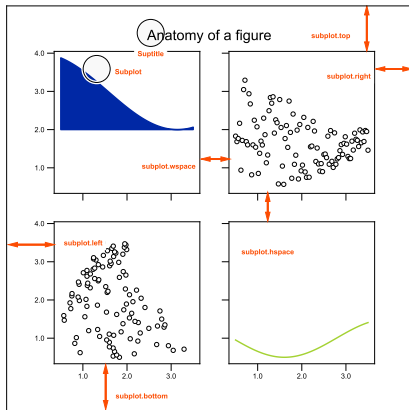# It does not take much to make better charts – and Python libraries can help us
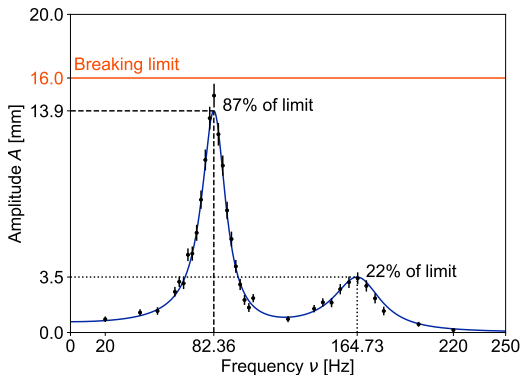
The structure of a figure and plot in matplotlib
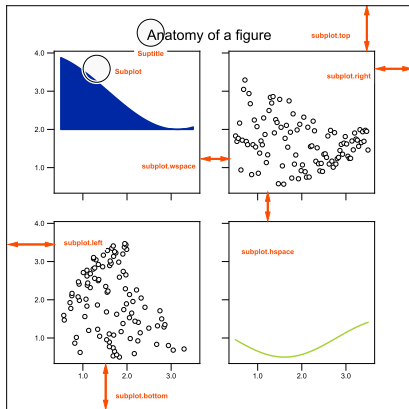
The structure of a figure and plot in matplotlib

References

- matplotlib cheatsheet
- Claus O. Wilke, Fundamentals of Data Visualization
- Datavizcatalogue
- Python Graph Gallery