Introduction to Programming

For Archaeologists

Part 4: Graphs & Plots

2021-2022

Universiteit Leiden
The Netherlands

Topics of this lecture series

1. Introduction: Python, variables, comments
2. Lists & Loops
3. Loading and manipulating data
4. **Graphs & Plots**
5. SQL & Databases
6. Advanced methods: Machine Learning, QGIS integration

Assignment

Assignment deadlines

- Assignment 1: 22 April
- Assignment 2: **6 May**
- Assignment 3: 20 May

Assignment 2 due tonight at 23.59

Topics of this lecture

- Why use Python for graphs?
- Data types
- Graph types
  - Bar chart
  - Pie chart
  - Box plot
  - Histogram
  - Scatter plot
  - Line plot
  - Heatmap
  - Geographical maps

After this lecture:

- You know the advantages and disadvantages of making graphs with Python
- You can explain the differences between nominal, ordinal, ratio & interval data
- You know which types of graphs can be used for which types of data
- You know the basic graph types and can create them with the Pandas and Seaborn packages

Graphs in Python

- Graphs are also called plots, charts, graphics, visualisations
- Excel can be easier due to point-and-click GUI
- Python graphs generally better:
  - More structured
  - Clear what you did
  - Repeatable (open science)
  - Customisable
  - Quick to re-use with new data
  - Easy export to many formats
  - (more beautiful graphs)

Excel vs Python

- Excel is best when doing **small and one-time analyses** or creating basic visualizations quickly. It is easy to become an intermediate user relatively without too much experience due to its GUI.
- Python is harder to learn because you have to download many packages and set the correct development environment on your computer. However, it provides a big leg up when working with big data and creating **repeatable, automatable analyses**, and in-depth visualizations.

Reminder on data types / level of measurement

Levels of Measurement

Nominal	Ordinal	Interval	Ratio
"Eye color"	"Level of satisfaction"	"Temperature"	"Height"
Named	Named	Named	Named
	Natural order	Natural order	Natural order
		Equal interval between variables	Equal interval between variables
			Has a "true zero" value, thus ratio between values can be calculated

In archaeology

- **Nominal**: artefact type (flint, pottery, bone)
  - Discrete, not ordered
- **Ordinal**: state of preservation (bad, good, perfect)
  - Discrete, ordered
- **Interval**: dates / years (1000 BCE, 0, 500 CE)
  - Continuous, no true zero (100 CE is not twice 50 CE)
- **Ratio**: weight / height (3 cm, 40 kg, but *not* -10 cm!)
  - Continuous, true zero (100g is twice 50g)

Level of measurement	Properties	Example	Descriptive statistics	Graph types
Nominal	Discrete Not ordered	Artefact type	Frequencies Percentages Mode	Bar Pie
Ordinal	Discrete Ordered	Preservation level	Frequencies Mode Median Percentiles	Bar Pie
Interval	Continuous No true zero	Time Temperature	Frequencies Mode Median Mean Std	Bar Pie Box plot Histogram Scatter plot Line plot
Ratio	Continuous True zero	Weight Height	Mean Std	Histogram Boxplot Scatter plot Line plot

Bar chart

Values are indicated by the length of bars, each of which corresponds with a measured group

Use with:
- Nominal
- Ordinal
- Interval
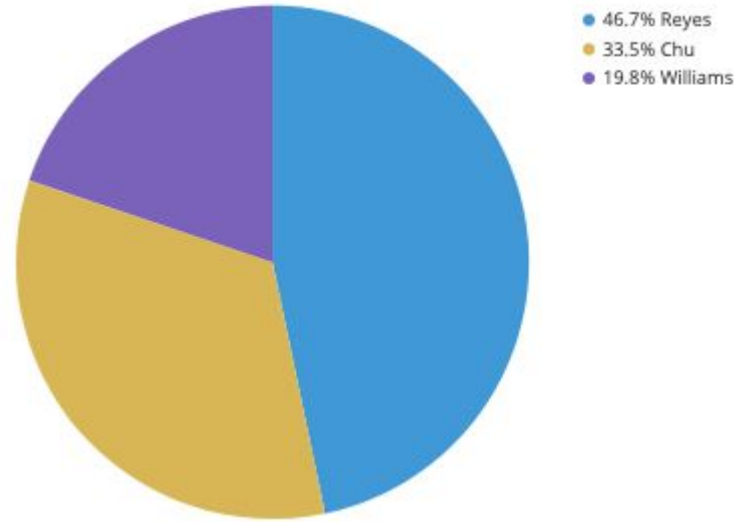

Purchases by User Type
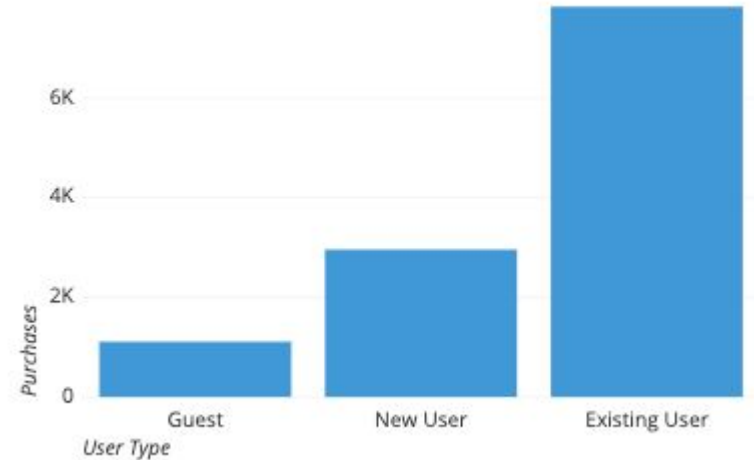
Pie chart

Values are indicated by size of slices, each of which corresponds with a measured group
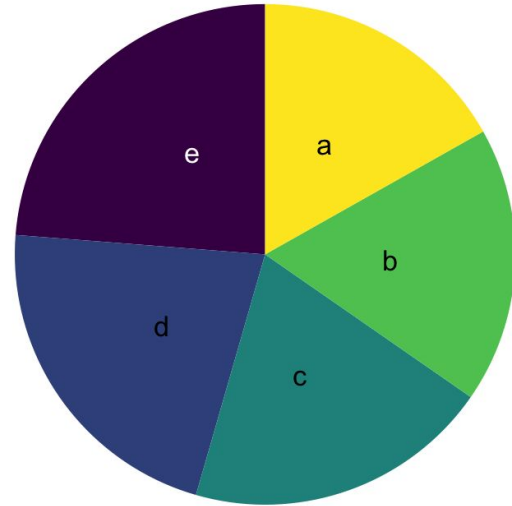
Good for %, not for absolute value

Use with:
- Nominal
- Ordinal

- 46.7% Reyes
- 33.5% Chu
- 19.8% Williams

Pie chart - warning

- Humans are bad at guessing angles
- Try to order the slices to the right from biggest to smallest
- Pretty difficult right?
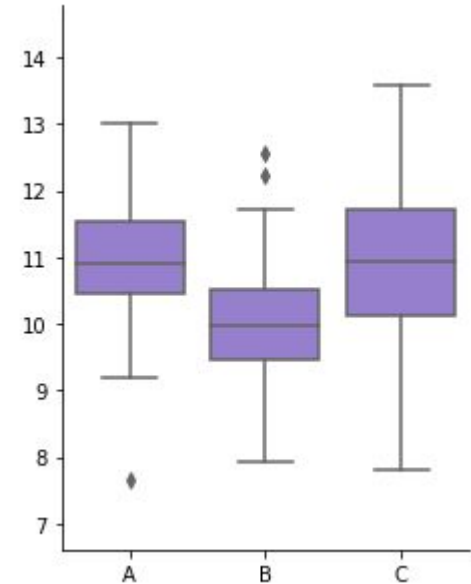- That's why pie charts are not great…
- Generally bar charts are fine!

Box plot

Uses boxes and whiskers to summarise the distribution of values within measured groups.

Positions of the box and whisker ends show the regions where the majority of the data lies
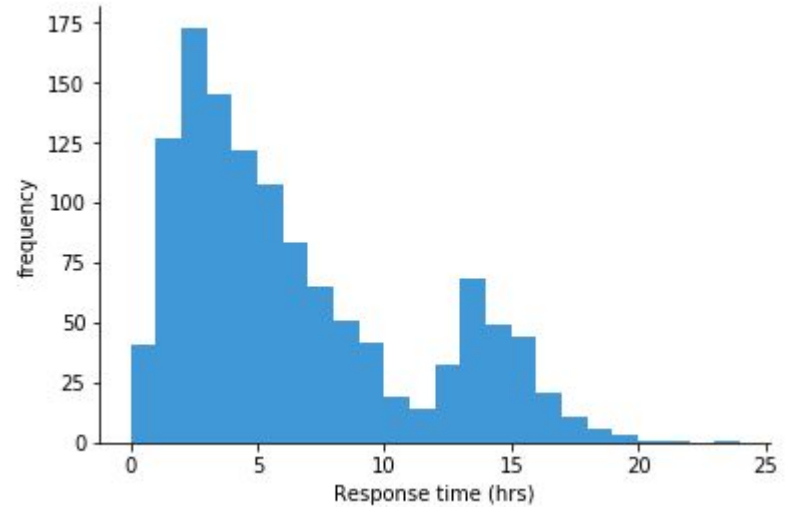
Use with:
- Interval
- Ratio

Histogram

Like bar chart, but for continuous variables

Uses 'bins' to group continuous variable, height of bar = count

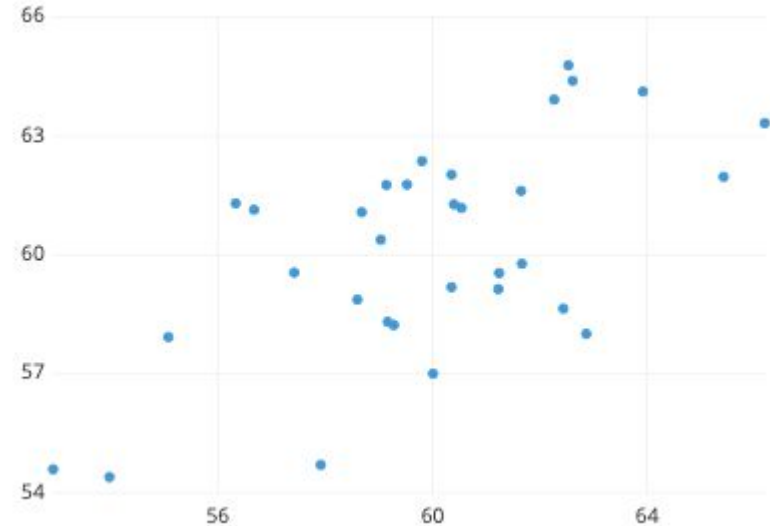Use with:
- Interval
- Ratio

Scatter plot

Displays values on two numeric variables using points positioned on two axes: one for each variable

Relationship between the plotted variables

Use with:
- Interval
- Ratio

Line chart

Show changes in value across continuous measurements, such as those made over time
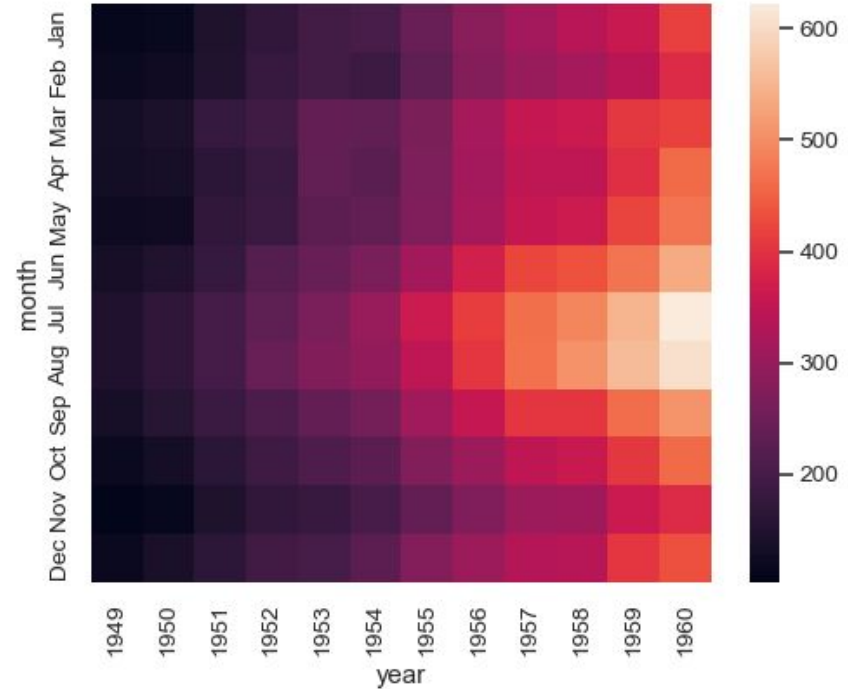
Use with:
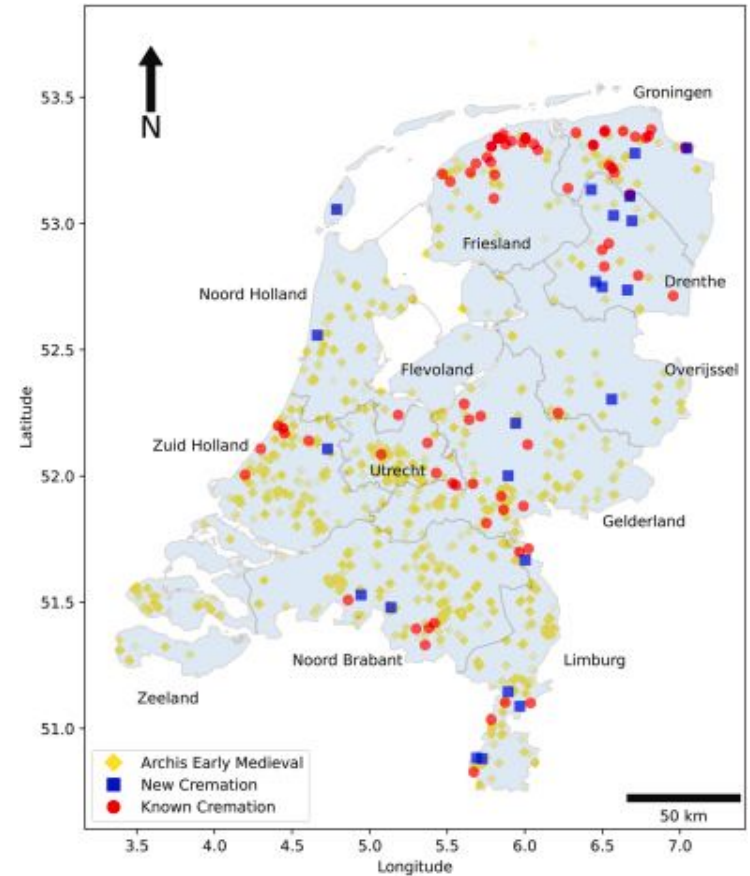- Interval
- Ratio

ZZD to QQY Exchange Rates

Heat map

Show continuous variable across 2 category variables

Geographical map

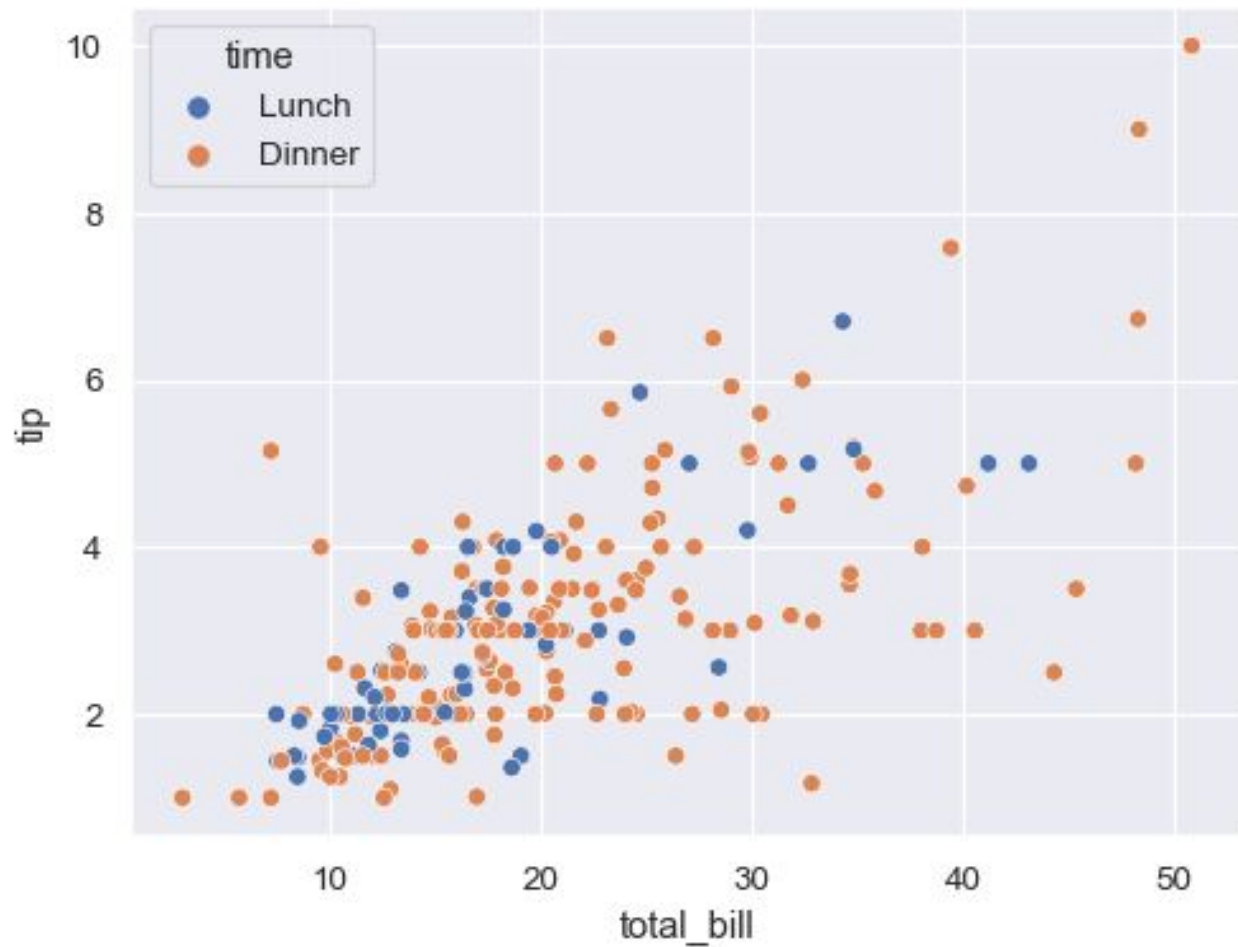Show variable across physical space, on a land map

Seaborn

- Another package you can import
- Allows easy ways to create plots
- Can use DataFrame as data source
- Looks like:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


sns.scatterplot(data=tips, x="total_bill",
y="tip", hue="time")
```

Seaborn

After this lecture:

- You know the advantages and disadvantages of making graphs with Python
- You can explain the differences between nominal, ordinal, ratio & interval data
- You know which types of graphs can be used for which types of data
- You know the basic graph types and can create them with the Pandas and Seaborn packages

Questions?

- **Any questions about any of the subjects?**

- Contact me at
    - a.brandsen@arch.leidenuniv.nl

Slides are available on Brightspace

Exercises

github.com/alexbrandsen/Introduction-to-Programming-for-Archaeologists

- Go to github
- Click on 'modules'
- Click click on the 4th module
- Right click 'raw', then select 'save link as' or 'download as'
- Save the file in the 'modules' folder within your own Scripts folder
- Start Anaconda
- Start Jupyter Notebook
- Navigate to the notebook file and run it