

Exploratory Data Analysis & Visualization

Multivariate Data

Ben Winjum

Today: Multivariate Analysis & Viz

Multivariate in the context of statistics

- Descriptive
 - Univariate: single variable
 - Bivariate: relationships between two variables
 - **Multivariate: relationships between multiple variables**
 - Dependency of a response on multiple variables; relationships between variables conditioned on the value of another variable; codependencies; ...
- Inferential
 - Hypothesis testing: testing whether your assumptions are true or not
 - Model fitting: (predictive analytics) use your insights to make decisions or predictions

Multivariate analysis may be necessary to explain most
real world problems

- Weather depends only on temperature

Multivariate analysis may be necessary to explain most real world problems

- Weather depends only on temperature
 - No! humidity, dewpoint, date, location, pollution, ...

Multivariate analysis may be necessary to explain most real world problems

- Weather depends only on temperature
 - No! humidity, dewpoint, date, location, pollution, ...
- Sales depend only on the number of Facebook likes

Multivariate analysis may be necessary to explain most real world problems

- Weather depends only on temperature
 - No! humidity, dewpoint, date, location, pollution, ...
- Sales depend only on the number of Facebook likes
 - No! media advertising, Yelp reviews, customer service, competitor values, product category...

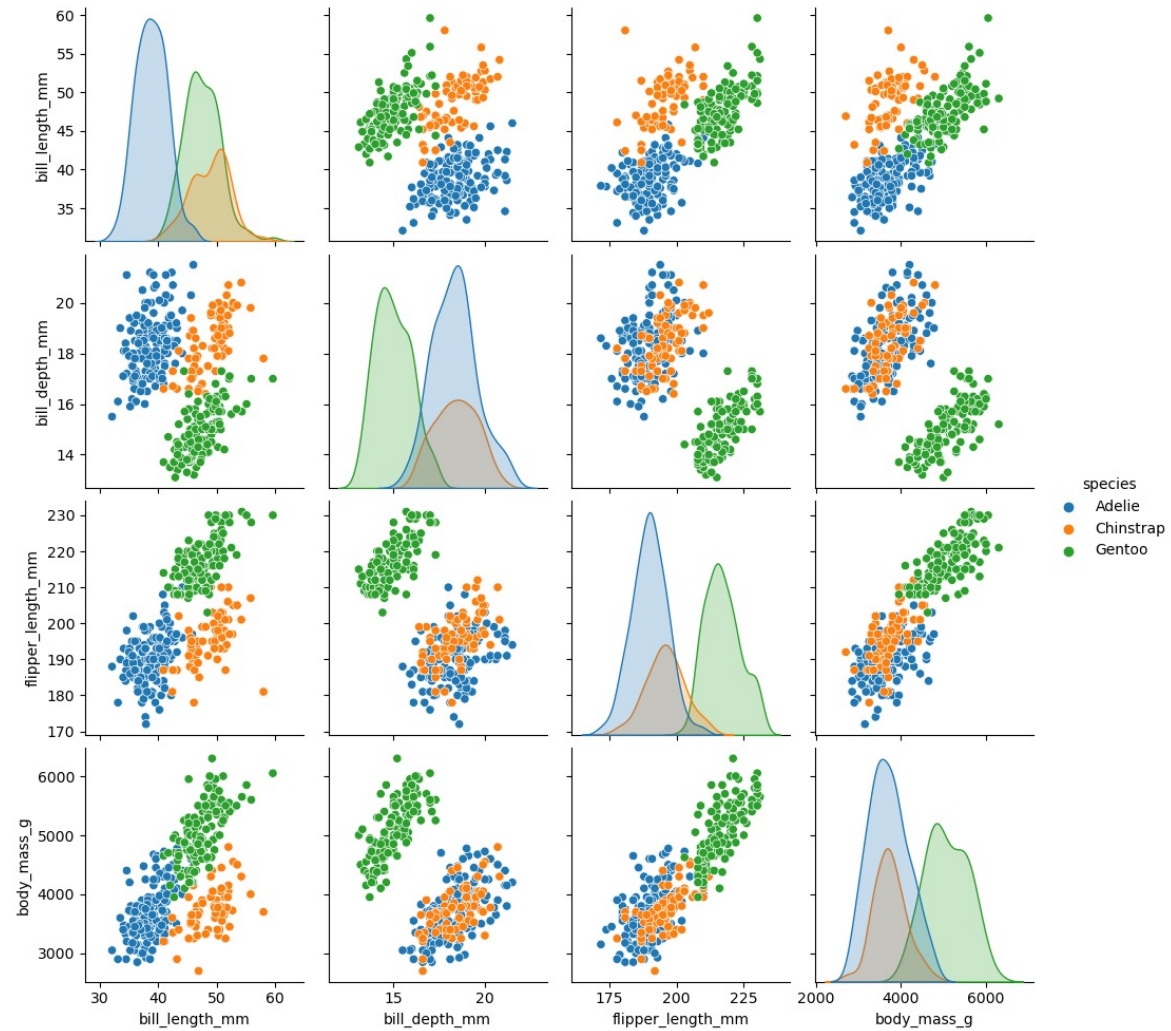
Multivariate analysis may be necessary to explain most real world problems

- Weather depends only on temperature
 - No! humidity, dewpoint, date, location, pollution, ...
- Sales depend only on the number of Facebook likes
 - No! media advertising, Yelp reviews, customer service, competitor values, product category...
- COVID-19 transmission rates depend only on President Biden's attitude

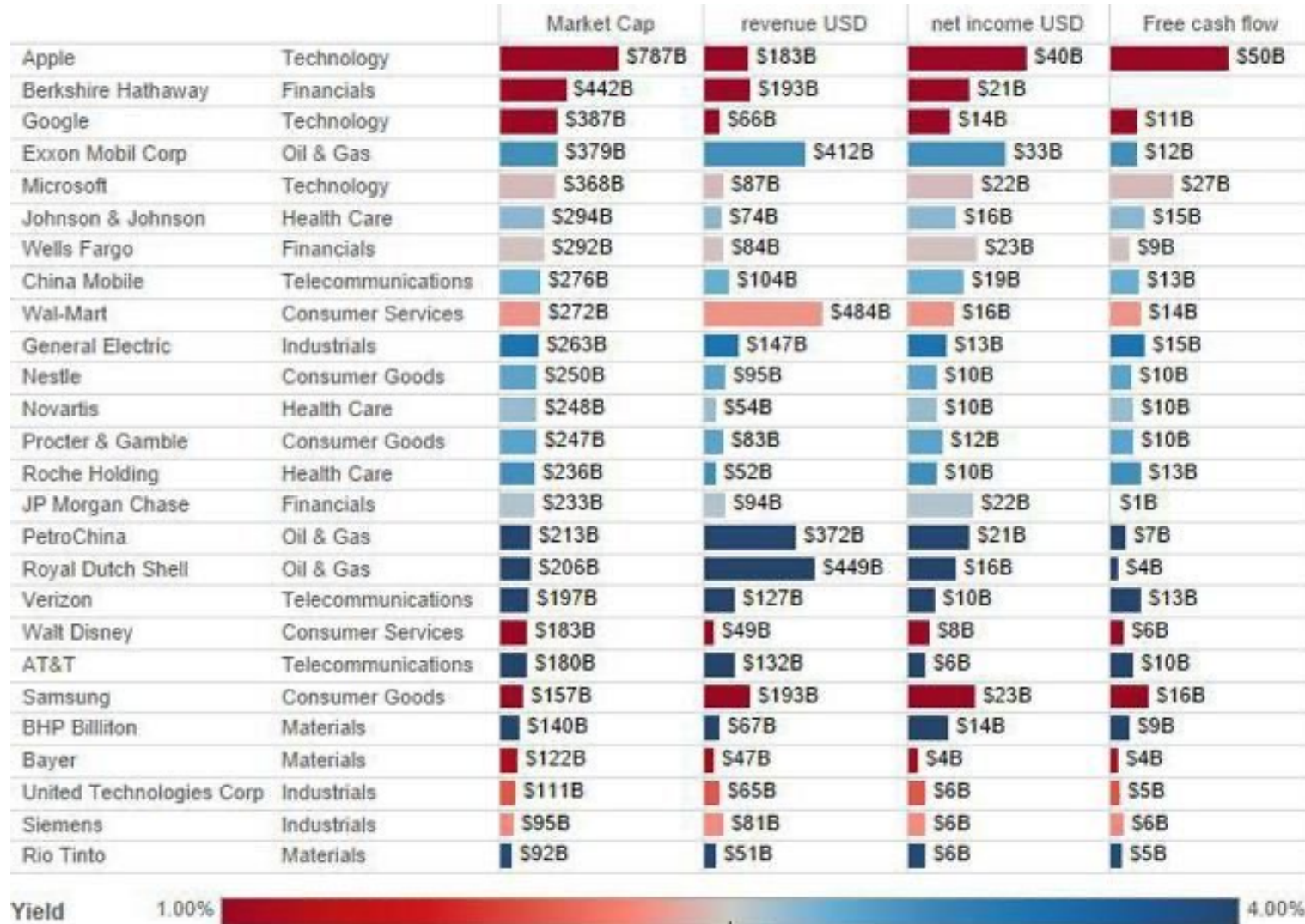
Multivariate analysis may be necessary to explain most real world problems

- Weather depends only on temperature
 - No! humidity, dewpoint, date, location, pollution, ...
- Sales depend only on the number of Facebook likes
 - No! media advertising, Yelp reviews, customer service, competitor values, product category...
- COVID-19 transmission rates depend only on President Biden's attitude
 - What? Public policy, healthcare services, population, public transportation methods, public behavior, ...

Multivariate visualization



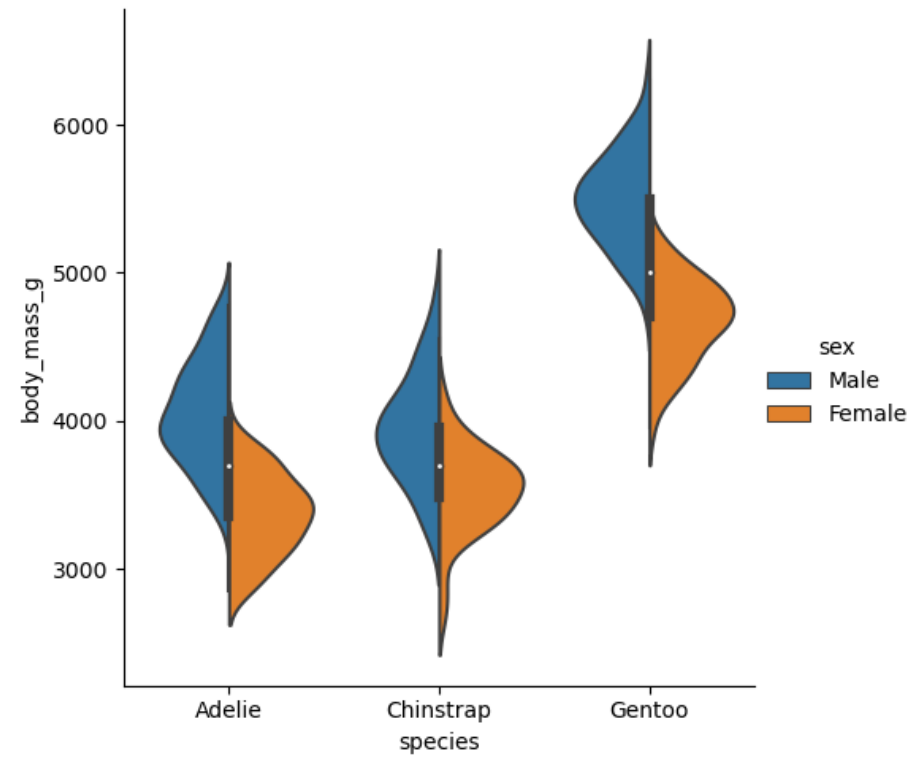
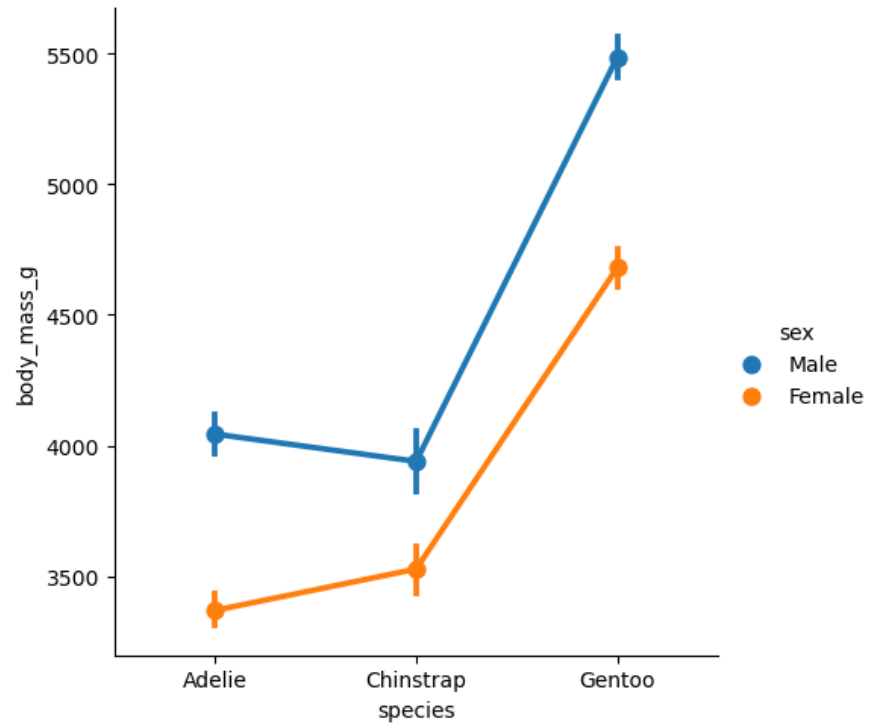
Multivariate visualization

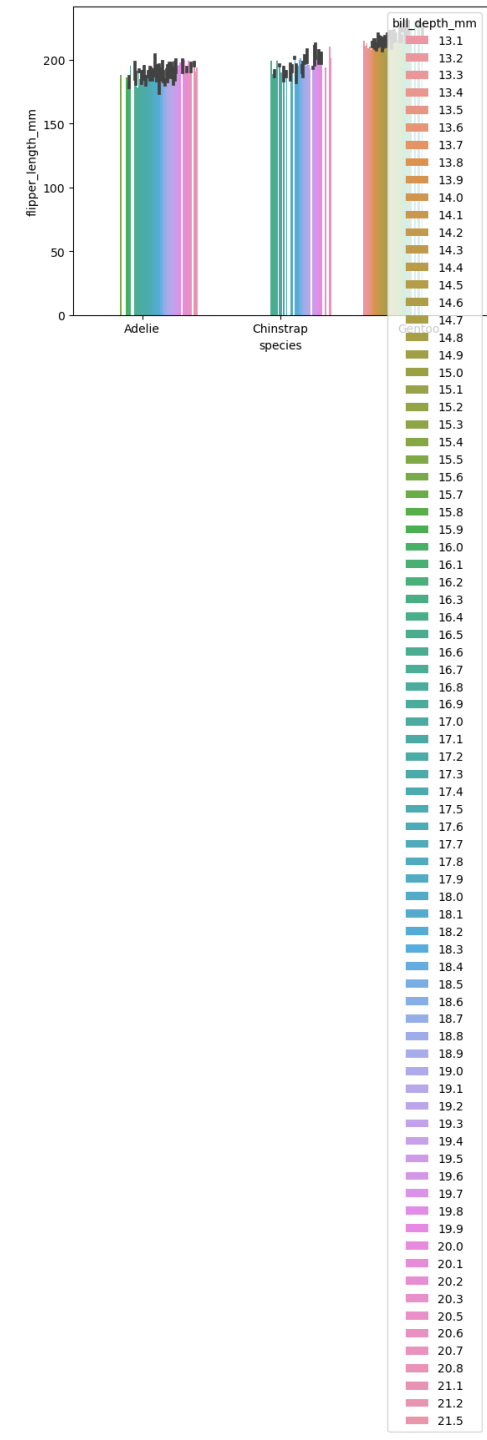


Multivariate visualization

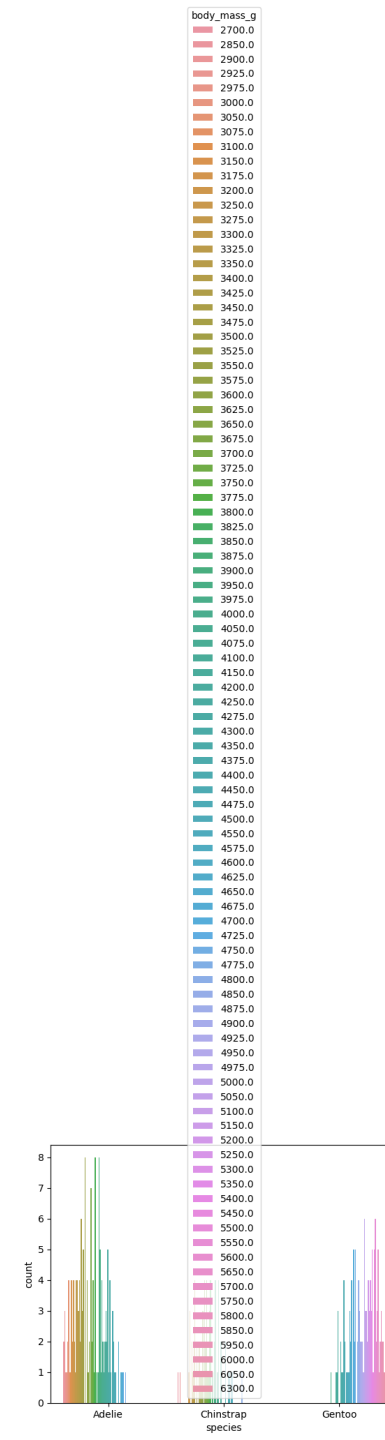
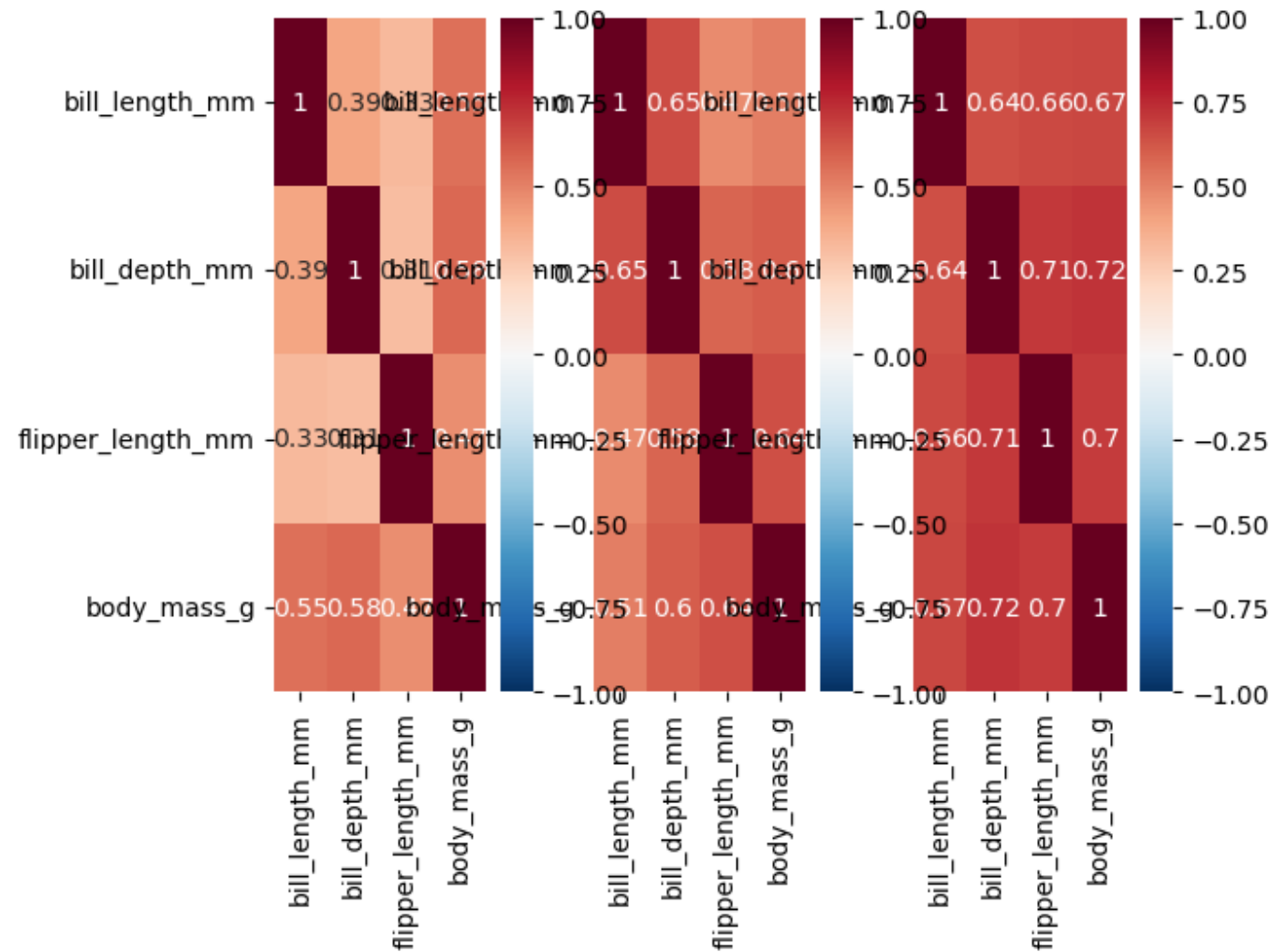
- Data visualizations can be used to understand and assess statistical models.
- Prematurely committing to a model, or foregoing adequate verification and validation, can lead you astray
- Tailor your viz to your questions, but ALSO, be open to surprises!
- Even simple data can merit a variety of graphical investigations

Follow your objective

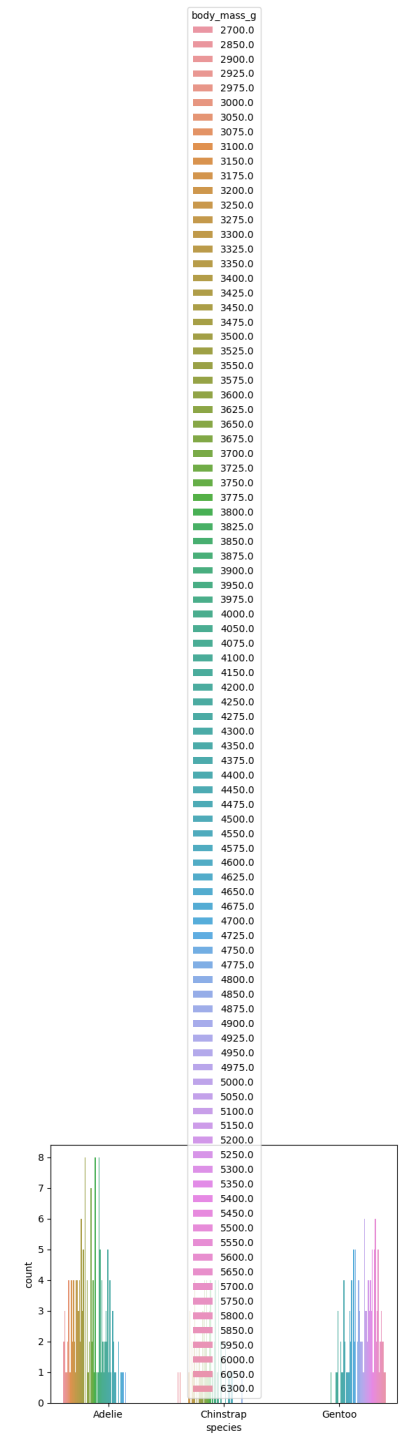
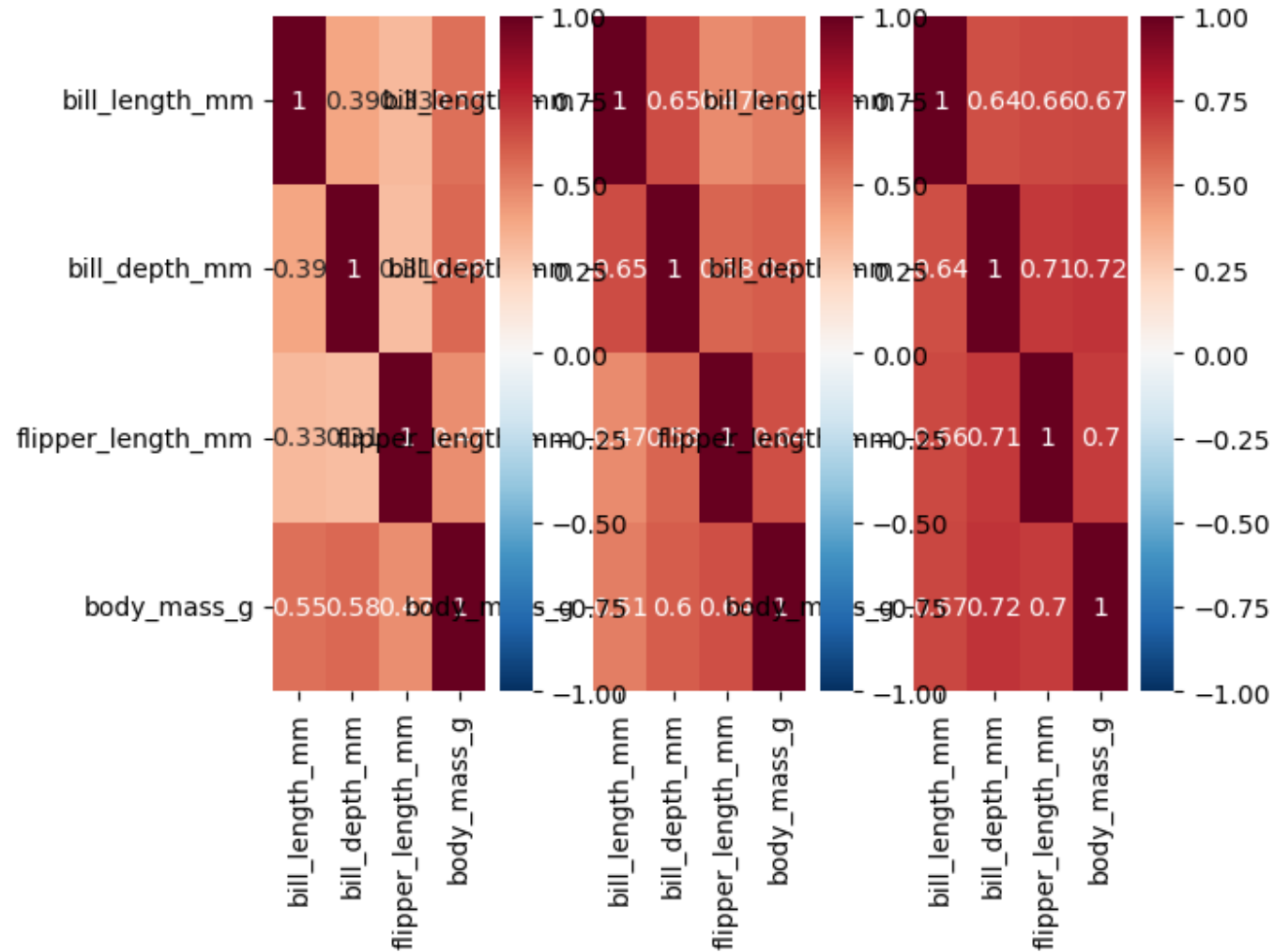
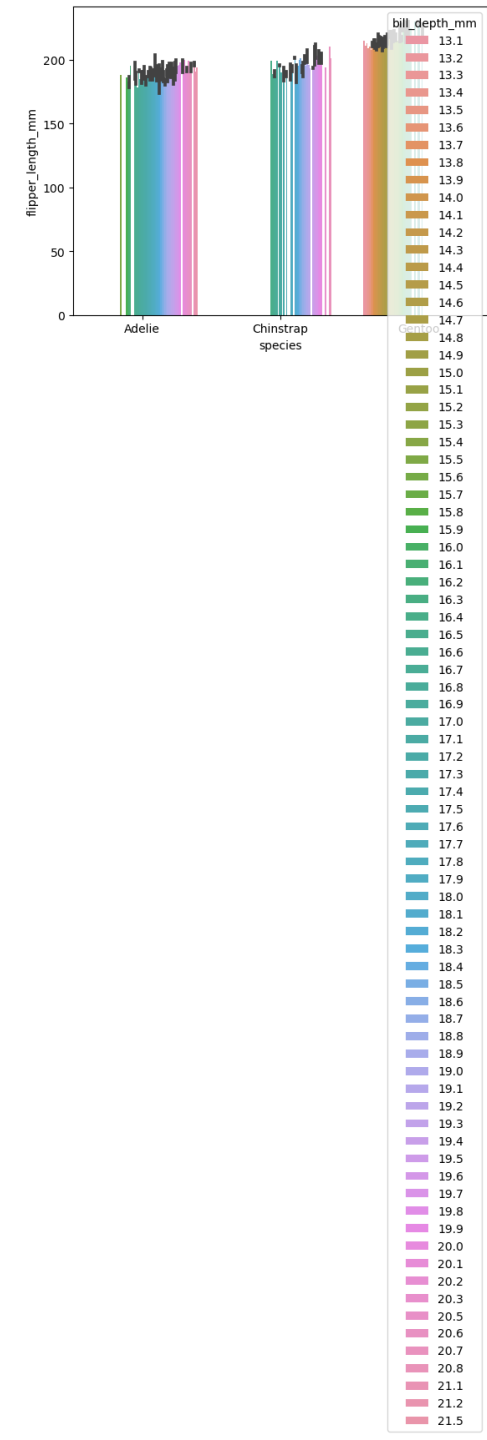




Do not let the visualization scare you



Data visualization is an exploratory journey



Multivariate analysis may be necessary to explain most real world problems

BUT

- It requires complex computations to arrive at satisfactory conclusions
- It can be a time-consuming process
- It can require a lot of visual experimentation
- It can be fun
 - There is rarely a single visualization that answers all questions. Instead, the ability to generate appropriate visualizations quickly is key
- Today -> more development of visualization skills
- Saved for later -> some useful analytical calculations to use in exploration