Name: Solutions

SA433 · Data Wrangling and Visualization Fall 2024 · Uhan

Exam 1 – Part 1 – 10/2/2024

Problem	Max	Score
1	10	
2	10	
Total		/ 20

Instructions

- This part is worth 20 points total. The exam (all three parts) is worth 100 points total.
- You have 50 minutes to complete Parts 1 and 2 of the exam.
- For Parts 1 and 2 of the exam, you may not use any outside assistance. These parts of the exam are closed book, closed notes, and closed internet.
- **No collaboration allowed.** All work must be your own.
- You must turn in Part 1 before beginning Part 2.
- **Do not discuss the contents of this exam with any midshipmen until it is returned to you.**

Background

The problems in this part are based on a dataset on Pokemon from `kaggle.com`. The columns are as follows:

Column	Description
`name`	Name of Pokemon
`type`	Type of Pokemon
`hp`	Hit points
`attack`	Attack points
`defense`	Defense points
`special_attack`	Special attack points
`special_defense`	Special defense points
`speed`	Speed points
`generation`	Generation when Pokemon was introduced

The first five rows of the DataFrame look like this:

	name	type	hp	attack	defense	special_attack	special_defense	speed	generation
0	Bulbasaur	Grass	45	49	49	65	65	45	1
1	Ivysaur	Grass	60	62	63	80	80	60	1
2	Venusaur	Grass	80	82	83	100	100	80	1
3	VenusaurMega Venusaur	Grass	80	100	123	122	120	80	1
4	Charmander	Fire	39	52	43	60	50	65	1

Assume the following code has already been run to import Pandas and Altair, and read the data into a DataFrame `df`:

```
import pandas as pd
import altair as alt

df = pd.read_csv('data/Pokemon.csv')
```

Problem 1. Write code that will generate a scatterplot of *attack points* versus *defense points*.

```
alt.Chart(df).mark_point().encode(
    alt.X('attack:Q'),
    alt.Y('defense:Q')
)
```

Problem 2. Write code that will generate a histogram of the *hit points* of the Pokemon.

```
alt.Chart(df).mark_bar().encode(
    alt.X('hp:Q').bin(),
    alt.Y('count():Q')
)
```