

21BDS0340 - Abhinav Dinesh Srivatsa

Data Mining Lab

Digital Assignment 1

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

```
In [ ]: temperature_data = pd.read_csv("./temperatures.csv")
temperature_data
```

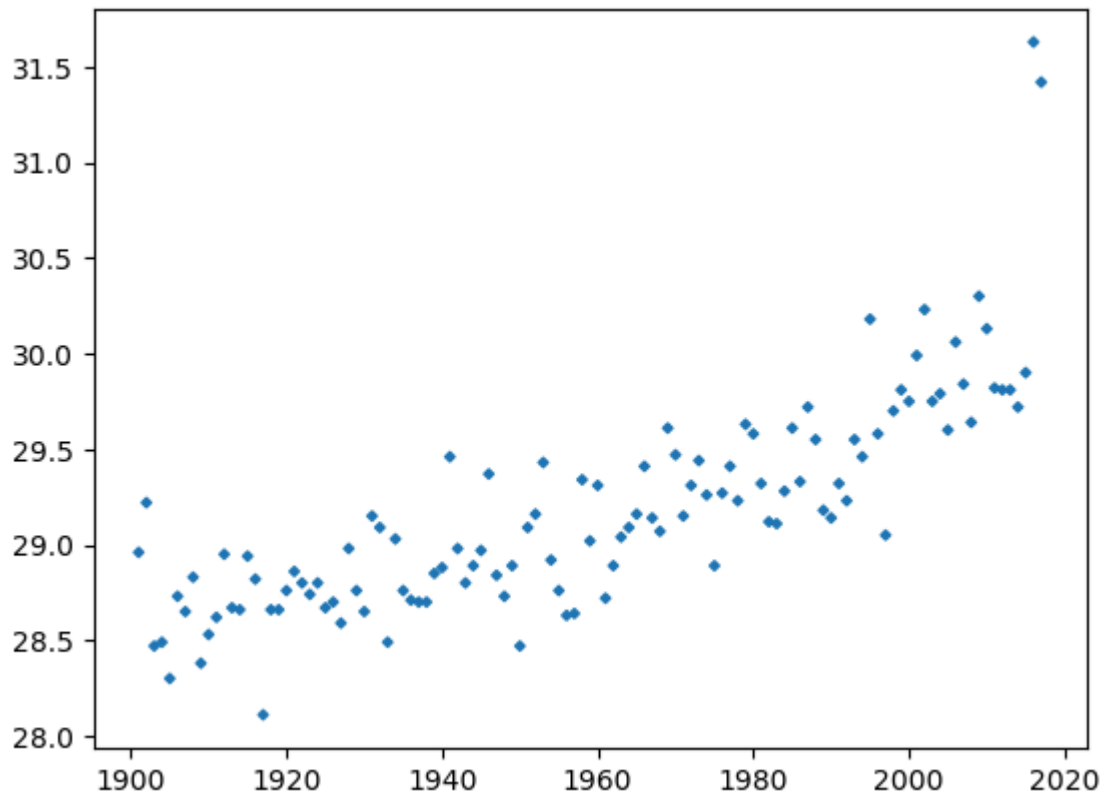
```
Out [ ]:
```

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
0	1901	22.40	24.14	29.07	31.91	33.41	33.18	31.21	30.39	30.47	29.97
1	1902	24.93	26.58	29.77	31.78	33.73	32.91	30.92	30.73	29.80	29.12
2	1903	23.44	25.03	27.83	31.39	32.91	33.00	31.34	29.98	29.85	29.04
3	1904	22.50	24.73	28.21	32.02	32.64	32.07	30.36	30.09	30.04	29.20
4	1905	22.00	22.83	26.68	30.01	33.32	33.25	31.44	30.68	30.12	30.67
...
112	2013	24.56	26.59	30.62	32.66	34.46	32.44	31.07	30.76	31.04	30.27
113	2014	23.83	25.97	28.95	32.74	33.77	34.15	31.85	31.32	30.68	30.29
114	2015	24.58	26.89	29.07	31.87	34.09	32.48	31.88	31.52	31.55	31.04
115	2016	26.94	29.72	32.62	35.38	35.72	34.03	31.64	31.79	31.66	31.98
116	2017	26.45	29.46	31.60	34.95	35.84	33.82	31.88	31.72	32.22	32.29

117 rows × 18 columns

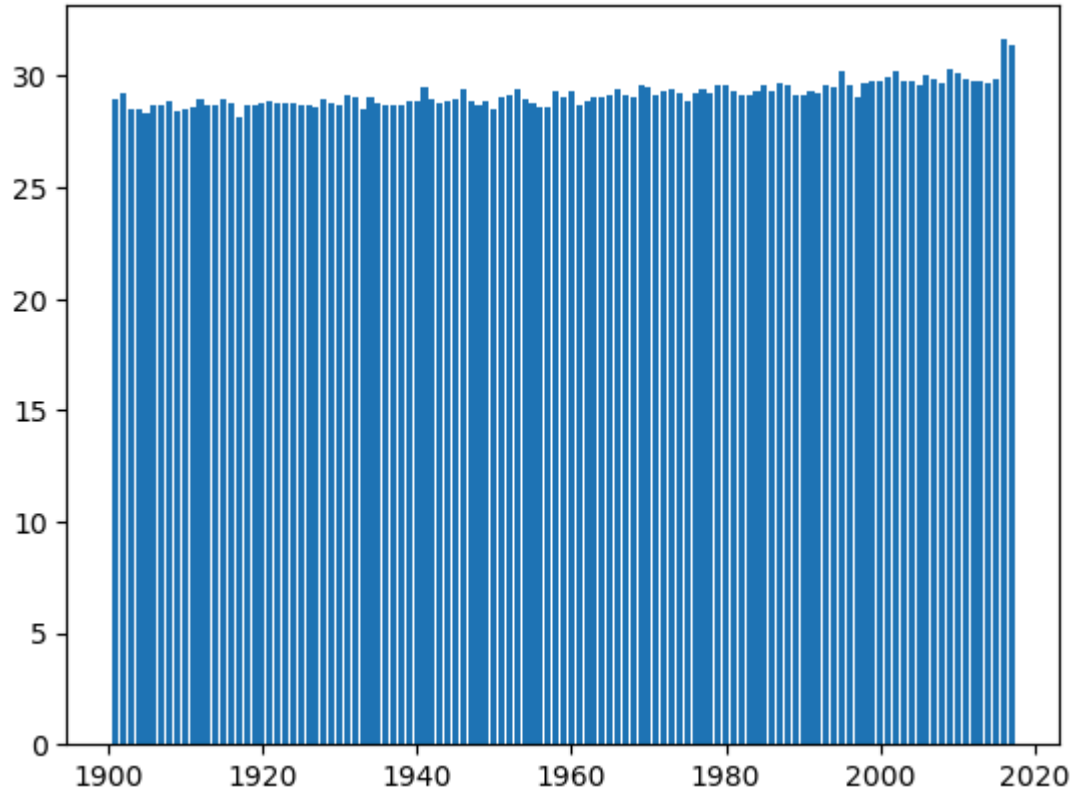
```
In [ ]: # question a)
sizes = np.sqrt(temperature_data["ANNUAL"])
plt.scatter(temperature_data["YEAR"], temperature_data["ANNUAL"], marker
```

```
Out [ ]: <matplotlib.collections.PathCollection at 0x1368ffd60>
```



```
In [ ]: # question b)
plt.bar(temperature_data["YEAR"], temperature_data["ANNUAL"])
```

Out[]: <BarContainer object of 117 artists>



```
In [ ]: # question c)
months = list(temperature_data.columns.values[1:-5])
ranges = {
    month: temperature_data[month].max() - temperature_data[month].min()
```

```
}
ranges
```

```
Out[ ]: {'JAN': 4.940000000000001,
        'FEB': 6.890000000000001,
        'MAR': 5.939999999999998,
        'APR': 5.370000000000001,
        'MAY': 3.9100000000000037,
        'JUN': 3.3799999999999955,
        'JUL': 2.9999999999999964,
        'AUG': 2.530000000000001,
        'SEP': 3.1499999999999986,
        'OCT': 4.390000000000001,
        'NOV': 4.41,
        'DEC': 4.990000000000002}
```

```
In [ ]: # question d)
months = list(temperature_data.columns.values[1:-5])
std_devs = {}
variances = {}
for month in months:
    std_devs[month] = temperature_data[month].std()
    variances[month] = temperature_data[month].var()
std_devs, variances
```

```
Out[ ]: ({'JAN': 0.834587606035784,
        'FEB': 1.1507567443147626,
        'MAR': 1.0684512315440673,
        'APR': 0.8894779942580806,
        'MAY': 0.7249049553033843,
        'JUN': 0.6331316748149578,
        'JUL': 0.468817576914948,
        'AUG': 0.4763119533699285,
        'SEP': 0.5442945910325506,
        'OCT': 0.705492200882088,
        'NOV': 0.7145175557391683,
        'DEC': 0.7826439391666521},
        {'JAN': 0.696536472148541,
        'FEB': 1.3242410845859118,
        'MAR': 1.1415880341880342,
        'APR': 0.7911711022693781,
        'MAY': 0.5254871942234016,
        'JUN': 0.40085571765399347,
        'JUL': 0.2197899204244032,
        'AUG': 0.22687307692307696,
        'SEP': 0.2962566018272916,
        'OCT': 0.4977192455054524,
        'NOV': 0.5105353374594754,
        'DEC': 0.6125315355142942})
```

```
In [ ]: # question e)
filtered = temperature_data[(temperature_data["YEAR"] >= 2010) & (tempera
median = filtered["ANNUAL"].median()
median
```

```
Out[ ]: 29.86
```

```
In [ ]: # question f)
temperature_data[["YEAR", "JAN-FEB", "MAR-MAY", "JUN-SEP", "OCT-DEC"]]
```

Out []:

	YEAR	JAN-FEB	MAR-MAY	JUN-SEP	OCT-DEC
0	1901	23.27	31.46	31.27	27.25
1	1902	25.75	31.76	31.09	26.49
2	1903	24.24	30.71	30.92	26.26
3	1904	23.62	30.95	30.66	26.40
4	1905	22.25	30.00	31.33	26.57
...
112	2013	25.58	32.58	31.33	27.83
113	2014	24.90	31.82	32.00	27.81
114	2015	25.74	31.68	31.87	28.27
115	2016	28.33	34.57	32.28	30.03
116	2017	27.95	34.13	32.41	29.69

117 rows × 5 columns

```
In [ ]: # question g)
temperature_data[temperature_data["ANNUAL"] > 25]["ANNUAL"].count()
```

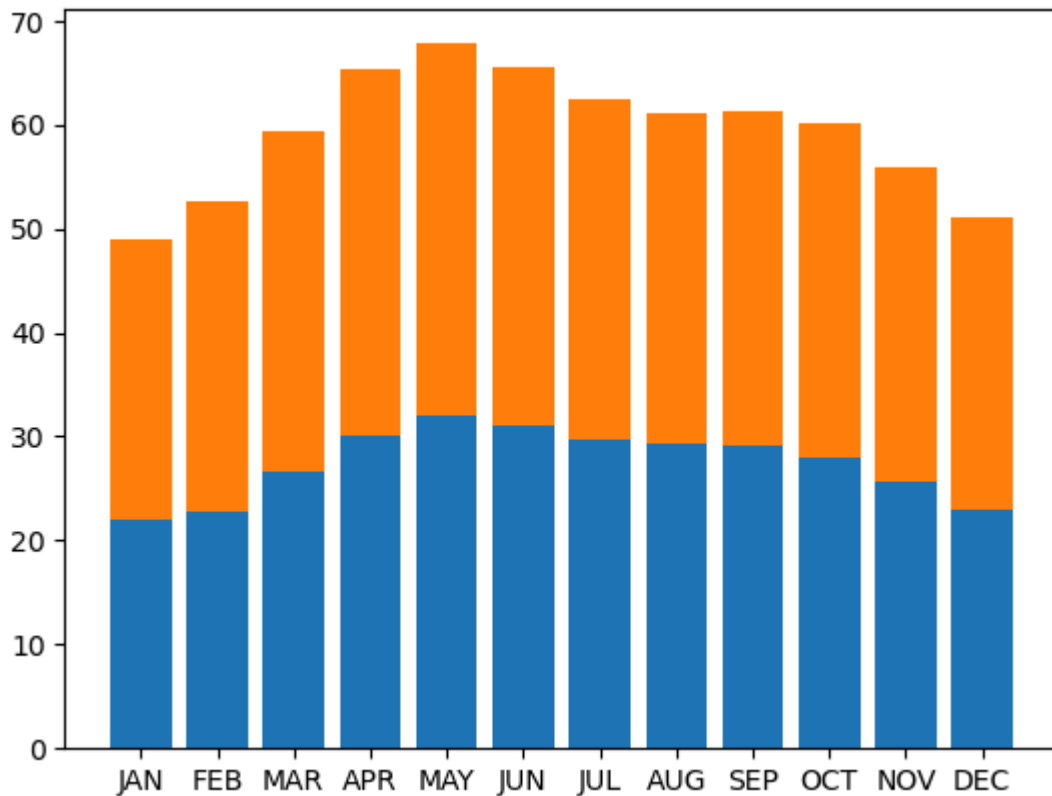
Out []: 117

```
In [ ]: # question h)
temperature_data["JUL"].mode()
```

Out []: 0 30.9
Name: JUL, dtype: float64

```
In [ ]: # question i)
months = list(temperature_data.columns.values[1:-5])
min_temps = []
max_temps = []
for month in months:
    min_temps.append(temperature_data[month].min())
    max_temps.append(temperature_data[month].max())
min_max_temps = pd.DataFrame({
    "MONTH": months,
    "MIN_TEMP": min_temps,
    "MAX_TEMP": max_temps
})
min_max_temps.set_index('MONTH', inplace=True)
plt.bar(months, min_temps)
plt.bar(months, max_temps, bottom = min_temps)
```

Out []: <BarContainer object of 12 artists>



```
In [ ]: # question j)
sns.heatmap(min_max_temps, annot=True, cmap='coolwarm', fmt='.2f', cbar_k
```

```
Out[ ]: <Axes: ylabel='MONTH'>
```

