In [3]:

```
import pandas as pd
data = pd.read_excel("rainfall2018.xlsx")
```

In [4]:

```
data["rainy"] = data["rainfall"] > 0.0
```

In [5]:

```
rainy_days_counter = data.groupby("rainy")["rainy"].groups[True].size
rainy_days_counter
```

Out[5]:

141

Here, it shows that the number of rainy day in 2018 is 141.

In [6]:

```
# similarly, we compute the spring days in 2018
spring_days_counter = data.groupby("season")["season"].groups["spring"].size
spring_days_counter
```

Out[6]:

92

We have the number of spring days in 2018 is 92

In [7]:

```
rain_by_season = data.groupby("season")["rainy"].value_counts()
rain_by_season
```

Out[7]:

```
season  rainy
fall    False    53
        True     38
spring  False    73
        True     19
summer  True     66
        False    26
winter  False    72
        True     18
Name: rainy, dtype: int64
```

In [8]:

```
spring_rainy_days_counter = rain_by_season["spring"][True]
spring_rainy_days_counter
```

Out[8]:

19

Here, the number of rainy days in the spring season in 2018 is 19

In [9]:

```
days_counter = data.shape[0]
days_counter # show the number of days in 2018, though it seems obvious
```

Out[9]:

365

In [10]:

```
P_rainy = rainy_days_counter / days_counter
P_spring = spring_days_counter / days_counter
P_spring_rainy = spring_rainy_days_counter / days_counter
P_rainy, P_spring, P_spring_rainy
```

Out[10]:

(0.3863013698630137, 0.25205479452054796, 0.052054794520547946)

Here, the answers to (d), (e), (f) are shown respectively

In [11]:

```
# using counters calculated above, we can derive
P_spring_given_rainy = spring_rainy_days_counter / rainy_days_counter
P_rainy_given_spring = spring_rainy_days_counter / spring_days_counter
P_spring_given_rainy, P_rainy_given_spring
```

Out[11]:

(0.1347517730496454, 0.20652173913043478)

These are answers to (g), (h)

In [22]:

```
# a general estimator following the above procedure
def estimator(data, day):
    season = data.loc[day].season
    season_days_counter = data.groupby("season")["season"].groups[season].size
    rain_by_season = data.groupby("season")["rainy"].value_counts()
    season_rainy_days_counter = rain_by_season[season][True]
    print("The probability of raining on this day is "
          + str(season_rainy_days_counter / season_days_counter)
          + ".")

# an example, we see the output is consistant with the fact that
# in 2018, it rained on 18 days out of 90 winter days.
estimator(data, 1)
```

```
The probability of raining on this day is 0.2.
```

In [ ]:

```

```