DSO 545: Statistical Computing and Data Visualization

Abbass Al Sharif

Fall 2019

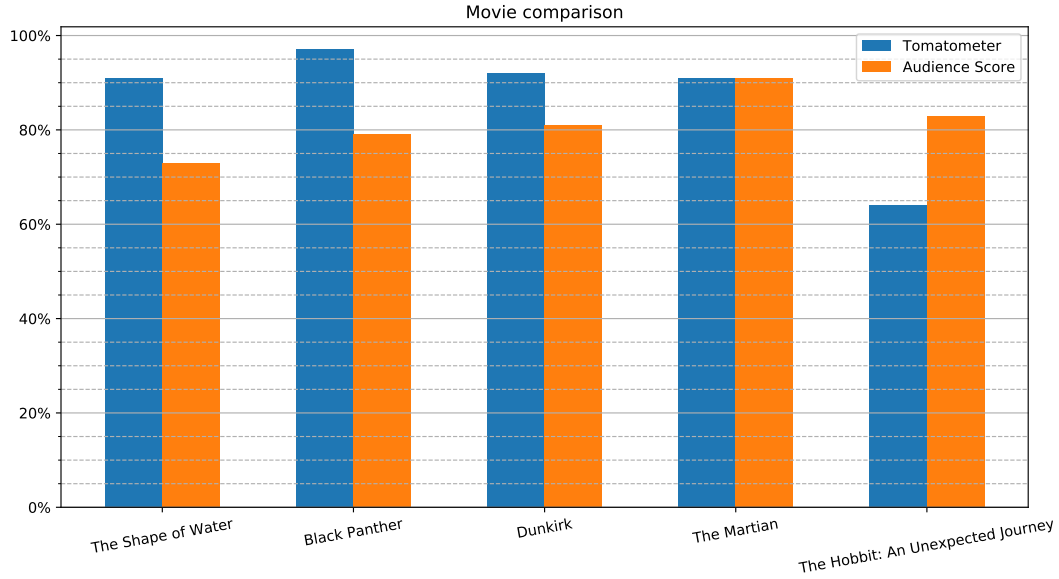
Lab 5: Data Visualization Using Matplotlib (part 2)

1. Read the dataset `movie_scores.csv` into Python, and clean the data if necessary.

```
##                          MovieTitle  Tomatometer  AudienceScore
## 0                The Shape of Water           91             73
## 1                     Black Panther           97             79
## 2                           Dunkirk           92             81
## 3                       The Martian           91             91
## 4  The Hobbit: An Unexpected Journey           64             83
```

2. Create the following bar plot to compare the movie scores from Tomatometer and Audience Score.

```
## <BarContainer object of 5 artists>
## <BarContainer object of 5 artists>
## ([<matplotlib.axis.XTick object at 0x118b53610>, <matplotlib.axis.XTick object at 0x11f072950>, <matp
## ([<matplotlib.axis.YTick object at 0x11f015350>, <matplotlib.axis.YTick object at 0x11f009a10>, <matp
## [Text(0, 0, 'The Shape of Water'), Text(0, 0, 'Black Panther'), Text(0, 0, 'Dunkirk'), Text(0, 0, 'T
## [Text(0, 0, '0%'), Text(0, 0, '20%'), Text(0, 0, '40%'), Text(0, 0, '60%'), Text(0, 0, '80%'), Text(
## [<matplotlib.axis.YTick object at 0x11f02a690>, <matplotlib.axis.YTick object at 0x11f0726d0>, <matpl
```



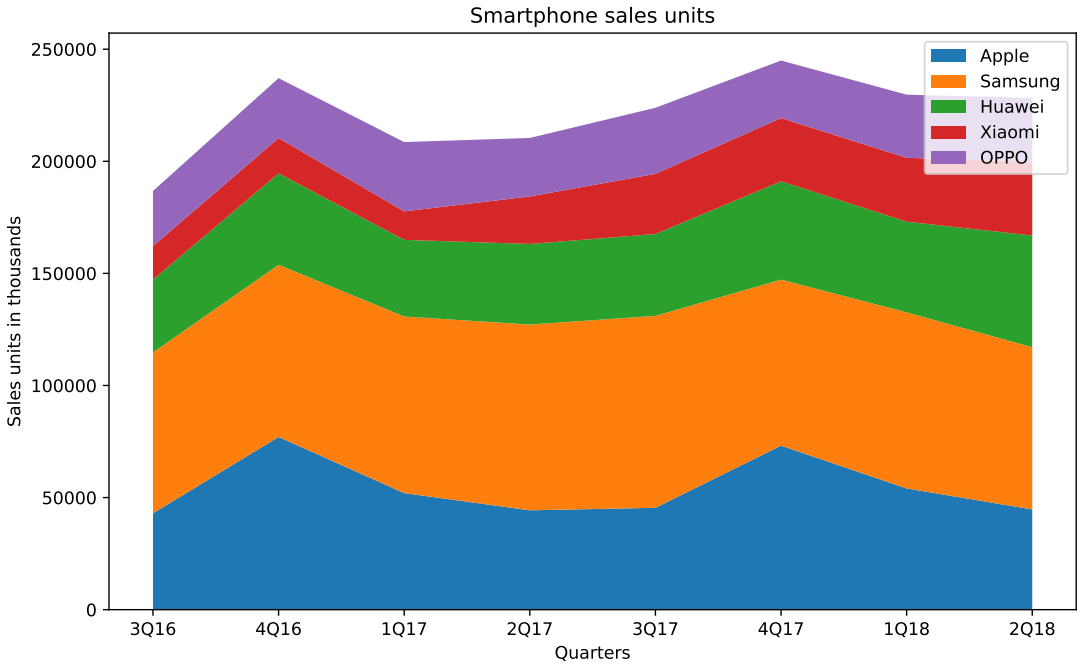
3. Read the dataset `smartphone_sales.csv` into Python, and clean the data if necessary.

```
##   Quarter  Apple  Samsung  Huawei  Xiaomi   OPPO
## 0    3Q16  43001    71734   32490   14926  24591
## 1    4Q16  77039    76783   40804   15751  26705
## 2    1Q17  51993    78776   34181   12707  30922
```

```
## 3    2Q17  44315    82855   35964   21179  26093
## 4    3Q17  45442    85605   36502   26853  29449
## 5    4Q17  73175    74027   43887   28188  25660
## 6    1Q18  54059    78565   40426   28498  28173
## 7    2Q18  44715    72336   49847   32826  28511
```

4. Create the following stacked area plot to compare the sales units of different smart phone manufacturer.

```
## [<matplotlib.collections.PolyCollection object at 0x121ea63d0>, <matplotlib.collections.PolyCollecti
```



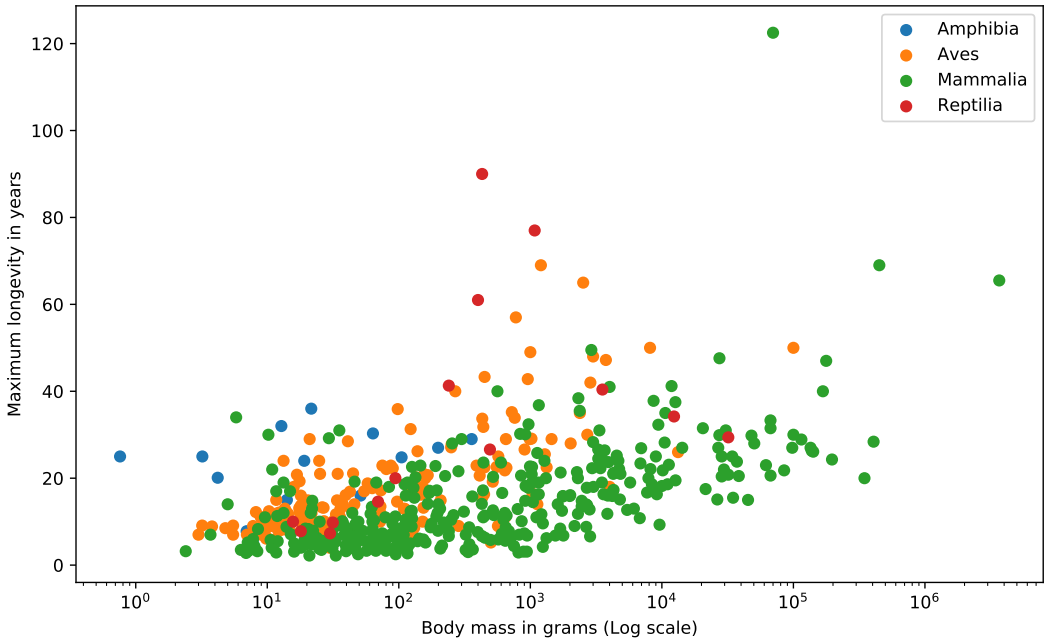
5. Read the dataset `anage.csv` into Python, and clean the data if necessary.

6. The dataset `anage.csv` is not complete. Filter the data so that you end up with samples containing a body mass and a maximum longevity.

7. Create 4 subsets of the given dataset: amphibia, aves, mammalia, and reptilia

```
## array(['Amphibia', 'Aves', 'Mammalia', 'Reptilia'], dtype=object)
```

8. Create a scatter plot that shows the corrlation between the body mass and the maximum longevity. Use log scale for the x-axis.


Amphibia
Aves
Mammalia
Reptilia
Maximum longevity in years
Body mass in grams (Log scale)
0
20
40
60
80
100
120
10^0
10^1
10^2
10^3
10^4
10^5
10^6