Data Analytics II: Project Part A

57
Gabriela Siren, Didrik Gentili, Anna Gao, Ludwig Fredriksson
25601, 25550, 25811, XXXXX

2023-01-29

A.1 - Explorative Data Analysis (EDA)

```
## [1] 0.525
```

This table tells me...

[your comments here]

Table 1: Summary table for all variables

	Age	Income	Kids
Mean	40.74	51901.34	1.32
Stan. Deviaton	12.84	20160.88	1.38
Minimum	17.00	11165.00	0.00
Q1 .25%	33.00	39906.75	0.00
Median	39.00	52574.00	1.00
Q3 .75%	48.25	64865.50	2.00
Maximum	77.00	138959.00	7.00

	Gender	OwnsHome	Subscribes
Female/False	0.52330	0.51667	0.86667
Male/True	0.47667	0.48333	0.13333

Table 2: Summary table for age

Segment	Mean	StDev	Minimum	Maximum	Q1	Median	Q3
Moving up	36.01	4.24	27	47	33	35	38.00
Suburb mix	39.29	5.14	27	53	36	39	42.25
Travelers	57.71	7.75	43	77	52	58	62.25
Urban hip	23.08	2.21	17	28	22	23	24.00

Table 3: Summary table for income

Segment	Mean	StDev	Minimum	Maximum	Q1	Median	Q3
Moving up	50613.60	8805.88	34975	73554	45391.00	50746.5	55835.75
Suburb mix	55444.01	12391.52	26521	82911	47162.75	55095.0	65273.00
Travelers	67832.45	19695.61	29699	138959	53802.50	66614.5	80348.00
Urban hip	21129.08	5175.90	11165	34106	18201.25	20850.0	24536.75

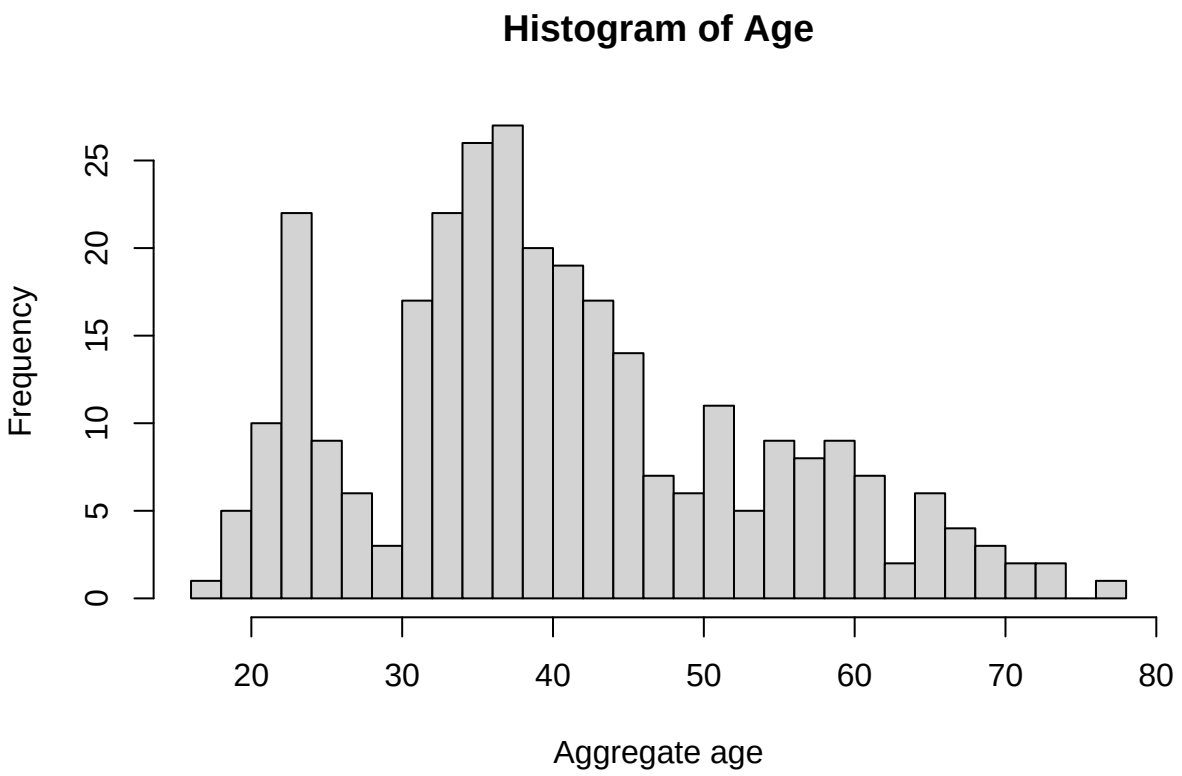
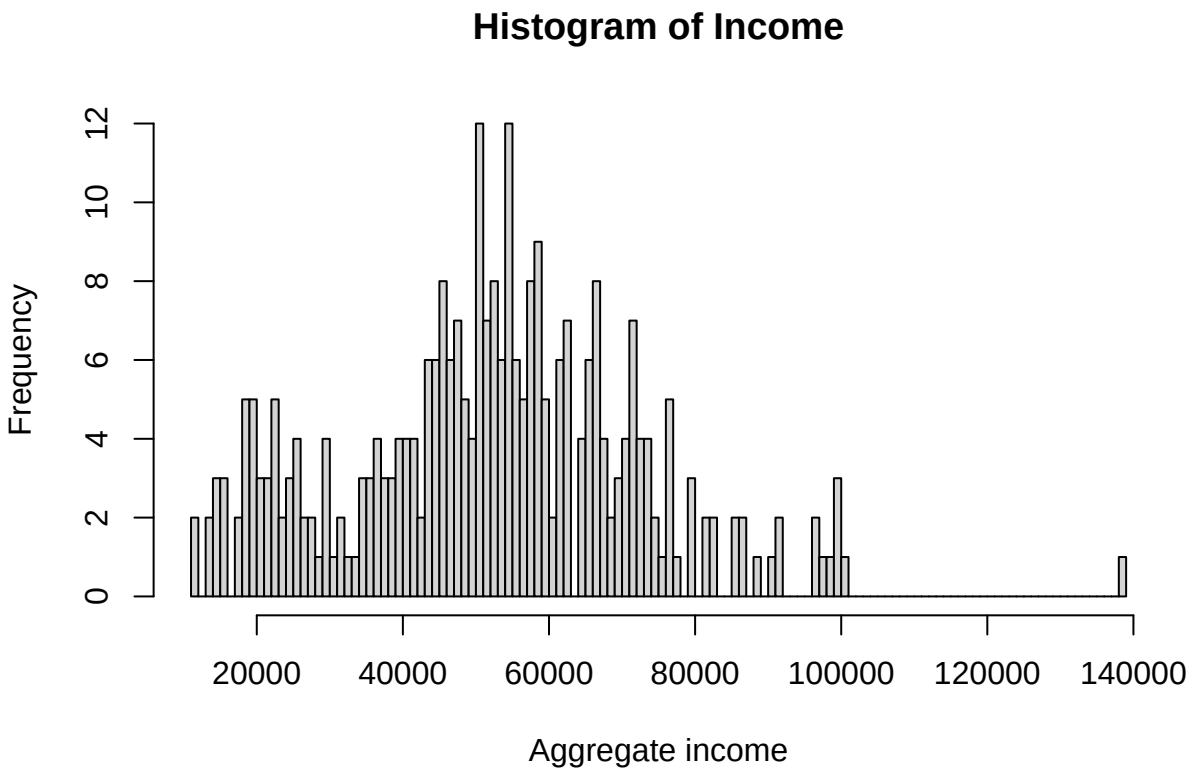
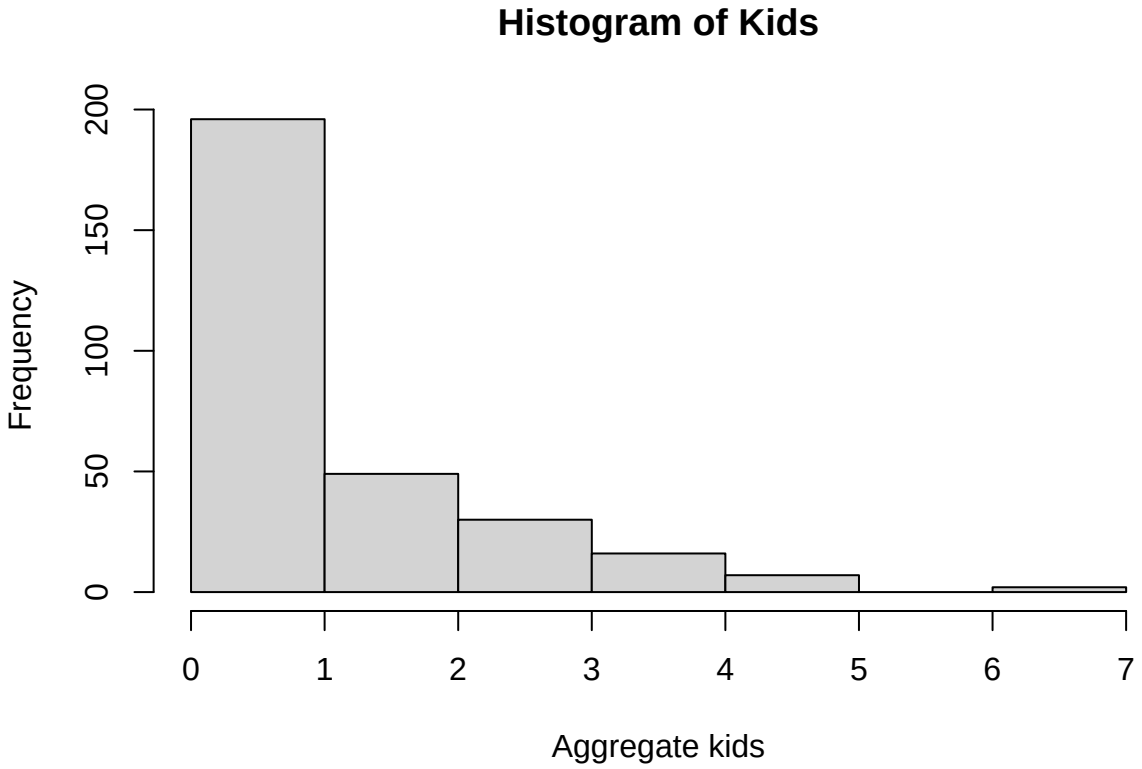
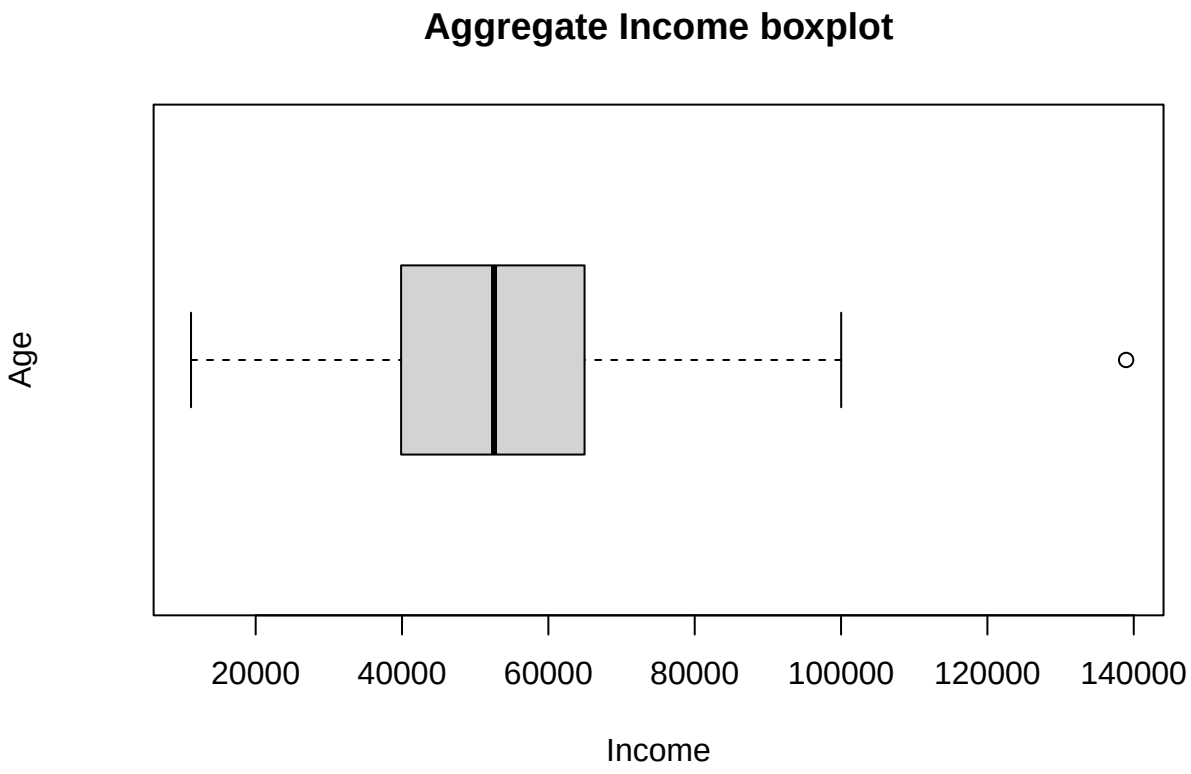


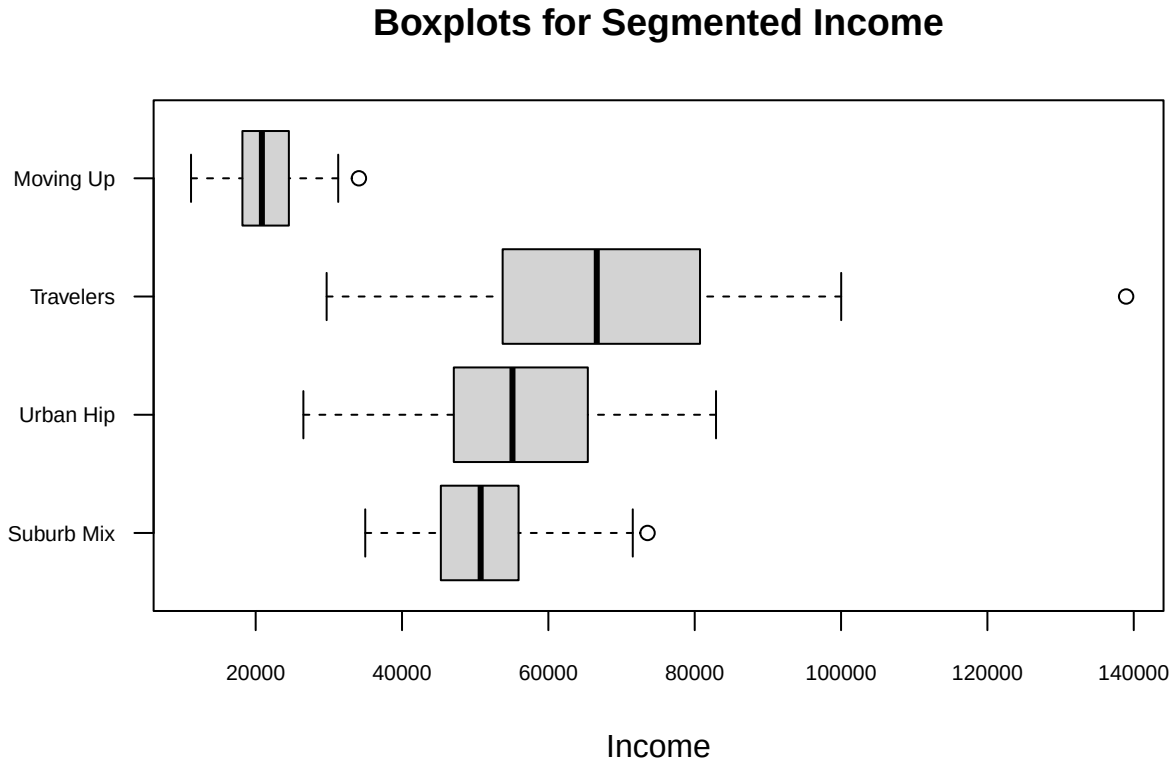
Table 4: Summary table for children

Segment	Mean	StDev	Minimum	Maximum	Q1	Median	Q3
Moving up	1.69	1.26	0	5	1	1	2
Suburb mix	1.90	1.49	0	7	1	2	3
Travelers	0.27	0.45	0	1	0	0	1
Urban hip	1.30	1.39	0	5	0	1	2


Histogram of Income
Frequency
0
2
4
6
8
10
12
20000
40000
60000
80000
100000
120000
140000
Aggregate income


Histogram of Kids
Frequency
0
50
100
150
200
0
1
2
3
4
5
6
7
Aggregate kids





[your comments here]

Do the variables Income and Age appear to be (approximately) normally distributed?

Can you think of a suitable distribution for the variable children?

A.2 - Confidence Intervals

[your comments here]

[your comments here]

Are the underlying assumptions for calculating this interval met in this data?

A.3 - Confidence Intervals and the Sample Size

```
## [1] 1690.965
```

```
## [1] 6763.859
```

[your comments here]

```
## [1] 6763.859
```

How should we adjust the confidence level as the size of the sample increases? Why?

A.4 - Confidence Intervals: Comparison of Population Means

```
## [1] "The income differnce is:"
```

```
## [1] 46703.37
```

```
## [1] "the 90 % confidence interval is:"
```

```
## [1] 42849.1
```

```
## [1] "to"
```

```
## [1] 50557.64
```

[your comments here]

Are underlying assumptions used to calculate this interval met in this data?