

# Homework 01

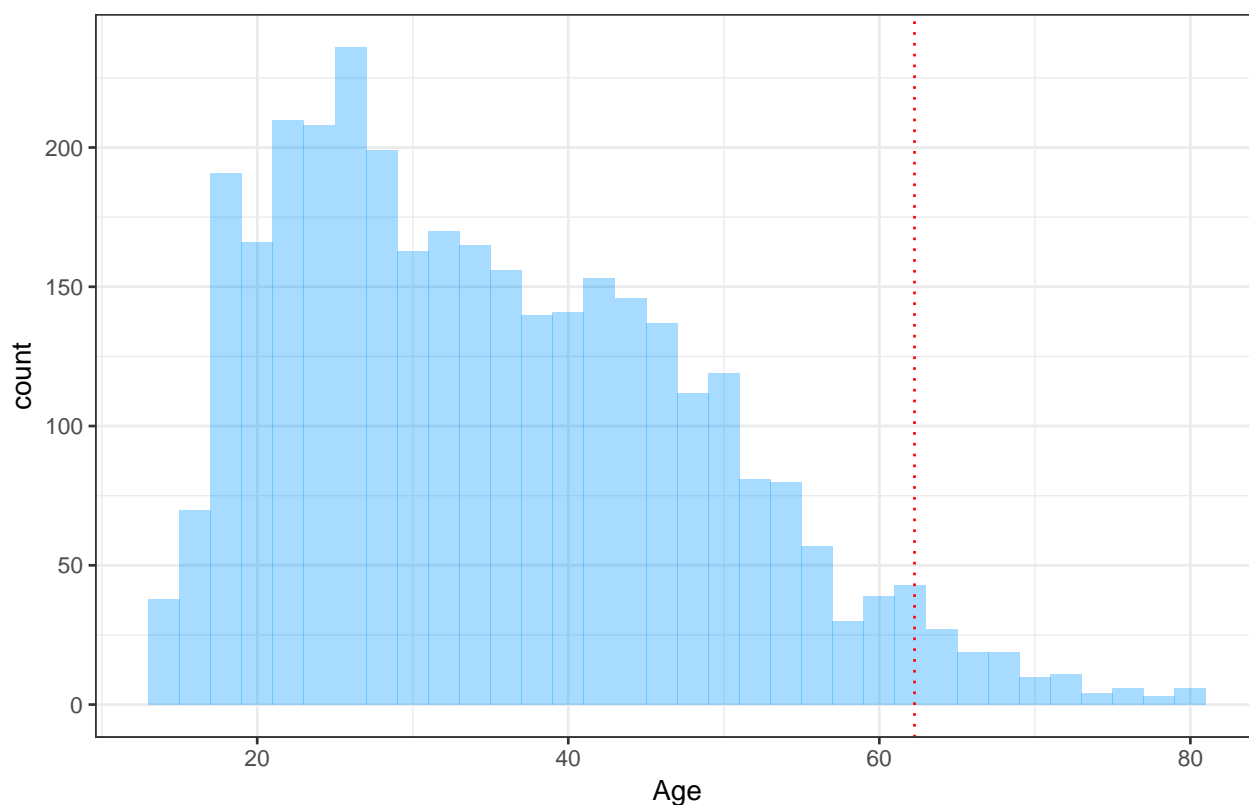
*Cody Frisby*

*8/21/2017*

**NOTE:** I have excluded all rows where age is less than 13. The range for age on this subset is 13 to 80.

## 1: How many statistical outliers are there in the Age Attribute?

The outliers, defined by  $\hat{\mu} \pm 2\hat{\sigma}$ , are any observations that are beyond 9.3084727, 62.2623172. As can be seen from the plot, it appears we are right skewed and there are no “outliers” to the left. This shouldn’t be too surprising since we are left censored at  $age = 13$ , meaning and age below this value is not possible. Here, I’ve added a vertical line at  $Age = 62.2623172$ . All observations beyond this line are counted as outliers.



The total number of observations that are outliers by the above definition is 128.

## 2: What is the most common type of Community Membership?

The most common type of community membership by the respondents appears to be **Hobbies** with 1554.

### 3: How many men and how many women responded to the survey?

The number of respondents who identified as women and men was **1236** and **2119** respectively. (*Recall that I've excluded all ages < 13*).

### 4: What is the primary reason people don't purchase on the internet?

**Security** appears to be the primary reason that people do not make internet purchases with **1391** people out of those surveyed.

Bad_experience	72
Bad_press	369
Cant_find	293
Company_policy	109
Easier_locally	729
Enough_info	547
Judge_quality	1321
Never_tried	384
No_credit	334
Not_applicable	428
Not_option	250
Other	356
Prefer_people	461
Privacy	987
Receipt	618
Security	1391
Too_complicated	192
Uncomfortable	492
Unfamiliar_vendor	608

### 5: Which country accounted for the fewest number of survey responses?

It appears from the data that more than one country had only one respondent. The country codes that had one respondent include **46, 60, 73, 78, 79, 84, 88, 90, 96, 97, 104, 107, 109, 118, 119**. Without more information regarding the coding of this variable no more reliable information can be given.

### 6: What is the median household income category for survey respondents?

Income catagories ranged from 0 to 8. Below is a table with the income catagories and the number of survey respondents for each.

0	1	2	3	4	5	6	7	8
507	225	232	392	424	387	621	288	279

The median catagory is **4**.

## 7: What was the most effective form of survey promotion?

The variables that appear to answer this question begin with “How\_You\_Heard\_About\_Survey\_”. The most frequent response was “**WWW\_Page**” with **1737** total respondents. The table below contains the number of respondents for each category.

Banner	270
Friend	201
Mailing_List	220
Others	476
Printed_Media	90
Remebered	96
Search_Engine	289
Usenet_News	242
WWW_Page	1737

## 8: Excluding outliers, what is the most common marital status?

Excluding all observations = 99, the most common marital status is **3**.

0	1	3	4	5	6
49	299	1340	48	1264	44

## 9: Assuming any rating above 4 indicates “a lot”, how many survey respondents claimed they falsified information “a lot”?

This question corresponds to the variable “**Falsification\_of\_Information**”. This is just  $145 + 262 = 407$ .

0	1	2	3	4	5	6
98	1789	754	189	118	145	262

## 10: What percent of survey respondents have their internet access paid for by their employer?

The percentages are displayed on the table below. Approximately **23.4%** of respondents have their internet paid for by their employer.

Dont_Know	Other	Parents	School	Self	Work
0.5	2.9	4.6	11.4	57.1	23.4