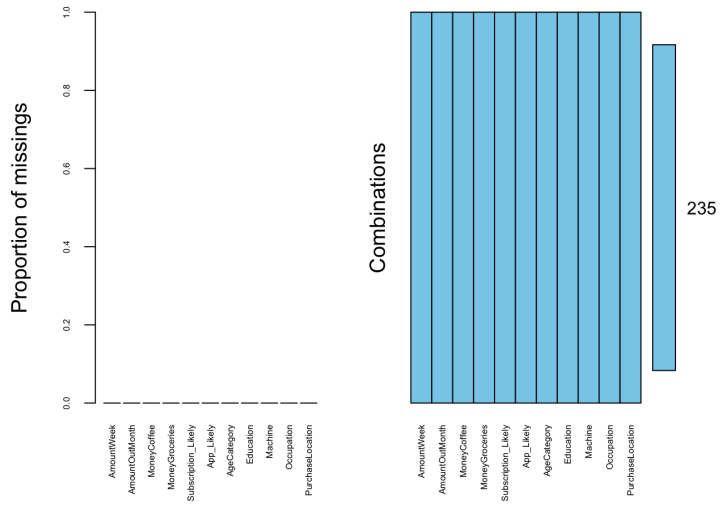
Thesis Data Analysis

15 September, 2021

Steps data analysis

- Univariate descriptions - categorical variables
    - Data table
    - Graphs
- Univariate descriptions - numerical variables
    - Summary
    - Graphs
- Boxplots - numerical
- Parametric testing
- Relationships & correlations
- Regressions

Introduction



Data set

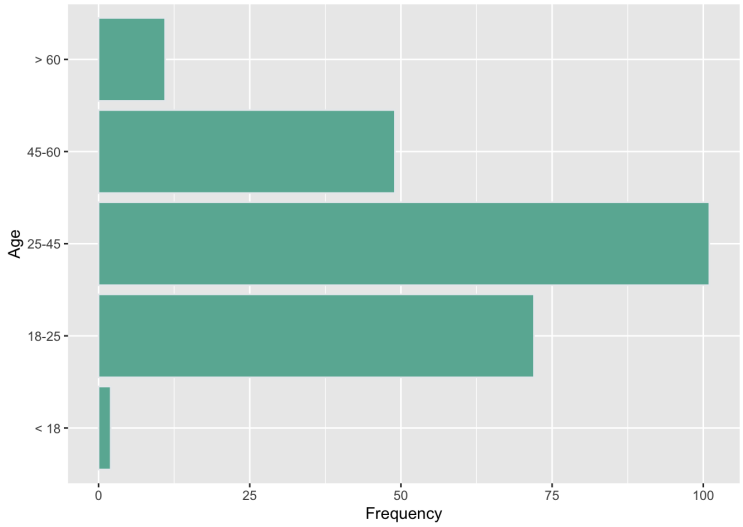
The variables included in the data set are:

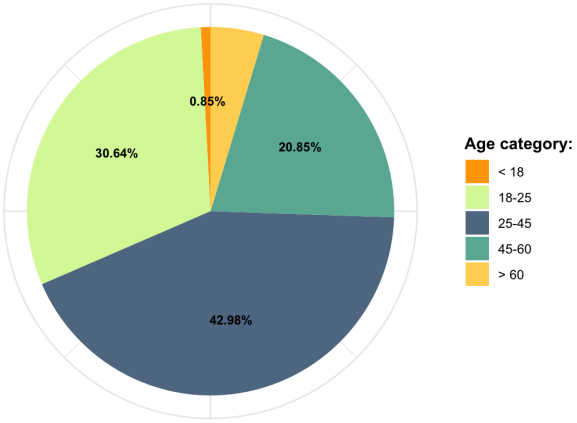
Field	Description
AmountWeek	How many cups of coffee do you typically consume weekly?
AmountOutMonth	How frequently do you drink out-of-home per month on average?
MoneyCoffee	How much money on average do you estimate you spend on coffee per month?
MoneyGroceries	How much on average do you spend on general groceries per month?
Machine	How do you brew your coffee at home?
Brand change	How often do you switch between coffee brands?
Purchase location	Where do you usually purchase your coffee?
Supermarket_Positive_Reasons	When you purchase coffee from the supermarket what are your main reasons for doing so?
Supermarket_Negative_Reasons	What would be reasons why you would not purchase coffee from the supermarket?
Criteria_Type_Coffee	What are your main criteria's or evaluation points for choosing the type of coffee?
KnowledgeCoffee	How would you describe your knowledge level regarding coffee in general?
Purchase_Price	I believe that the ____ is important to my decision on which coffee to purchase.
Purchase_Sustainability	I believe that the ____ is important to my decision on which coffee to purchase.
Purchase_Sustainability	I believe that the ____ is important to my decision on which coffee to purchase.
Purchase_Fairtrade	I believe that the ____ is important to my decision on which coffee to purchase.
Purchase_Packaging	I believe that the ____ is important to my decision on which coffee to purchase.
Frequency_Specialty	How often do you drink specialty coffee?
Subscription_Likely	How likely are you to have an online subscription for (specialty) coffee?
Subscription_Not_Likely	What is the number one reasons why you would be hesitant?
App_Likely	How likely are you to value and use an app for your online subscription?
Gender	What is your gender?
AgeCategory	What is your age category?
Occupation	What is your occupational status?
Education	What level of education have you completed?
Home	How would you describe the place you currently live in?

Univariate descriptions - Categorical variables

Age category

Age Category	Absolute	Relative
< 18	2	0.85%
18-25	72	30.64%
25-45	101	42.98%
45-60	49	20.85%
> 60	11	4.68%

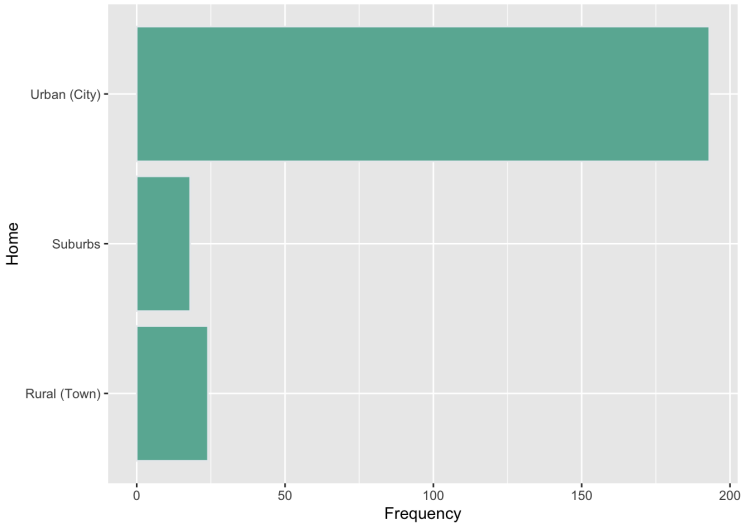


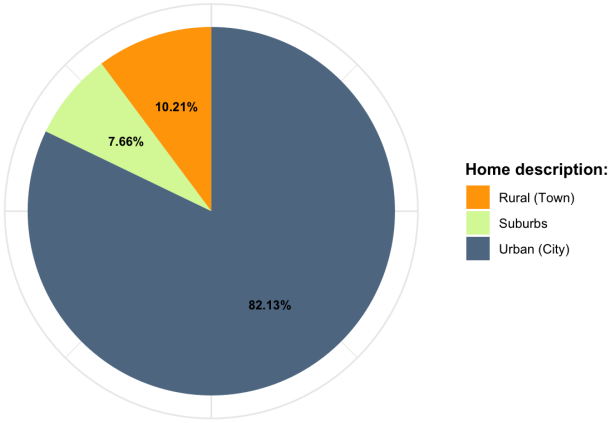


Home

Home	Absolute	Relative
Rural (Town)	24	10.21%

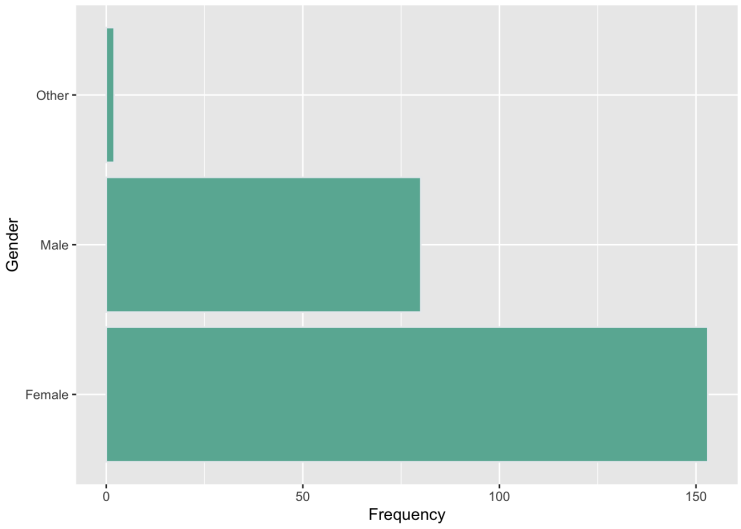
Suburbs	18	7.66%
Urban (City)	193	82.13%

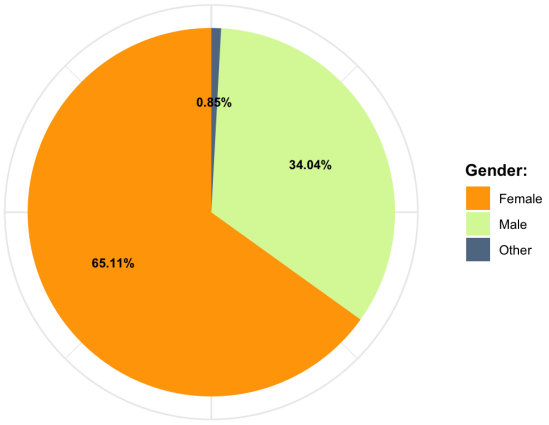




Gender

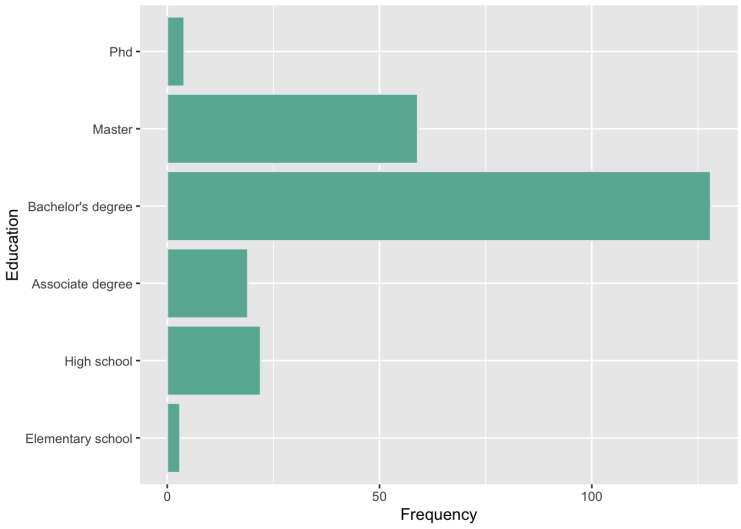
Gender	Absolute	Relative
Female	153	65.11%
Male	80	34.04%
Other	2	0.85%

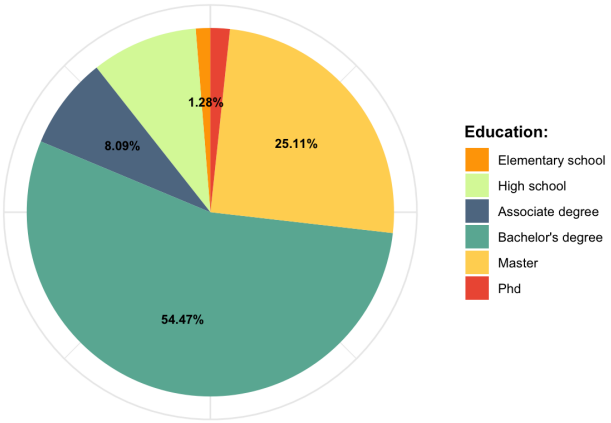




Education

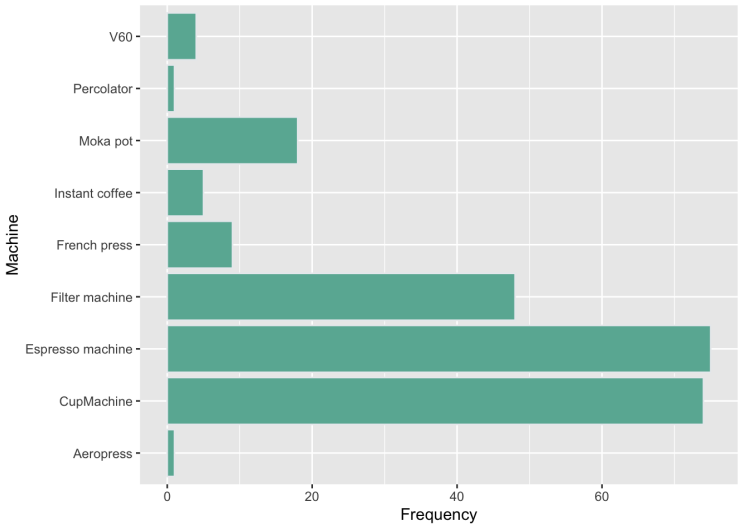
Education	Absolute	Relative
Elementary school	3	1.28%
High school	22	9.36%
Associate degree	19	8.09%
Bachelor’s degree	128	54.47%
Master	59	25.11%
Phd	4	1.70%

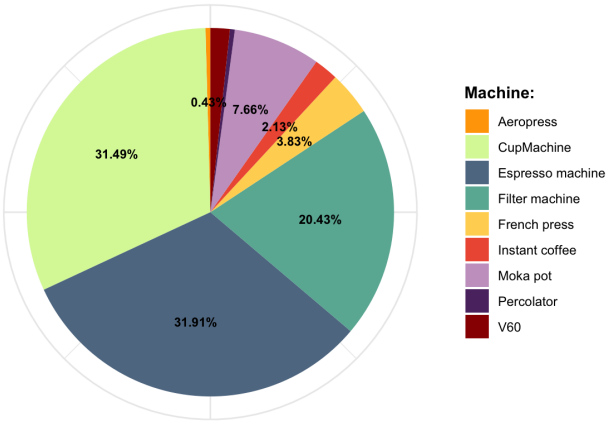




Machine

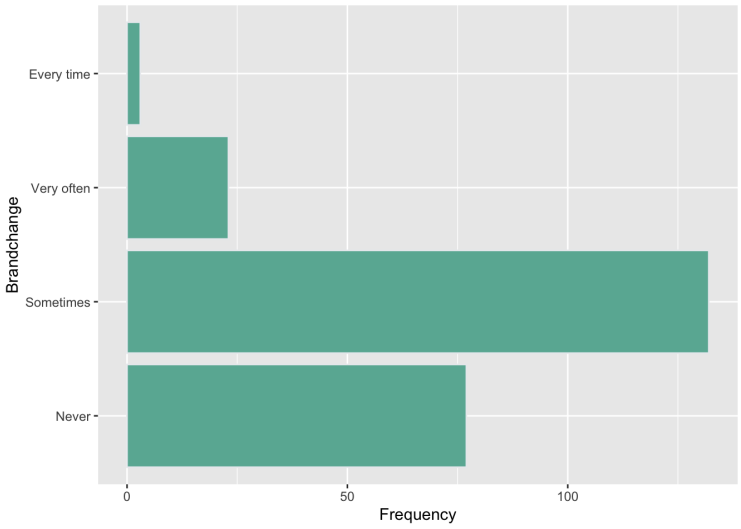
Machine	Absolute	Relative
Aeropress	1	0.43%
CupMachine	74	31.49%
Espresso machine	75	31.91%
Filter machine	48	20.43%
French press	9	3.83%
Instant coffee	5	2.13%
Moka pot	18	7.66%
Percolator	1	0.43%
V60	4	1.70%

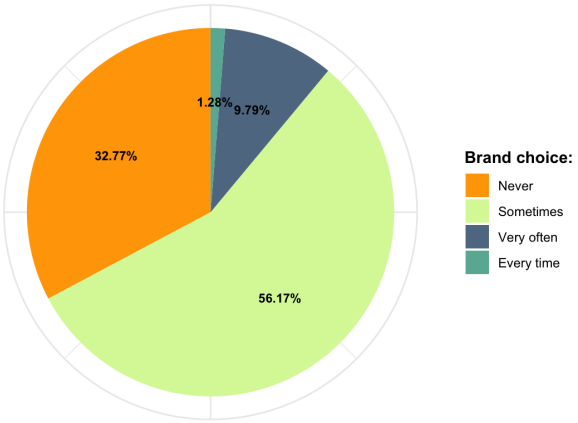




Brand choose

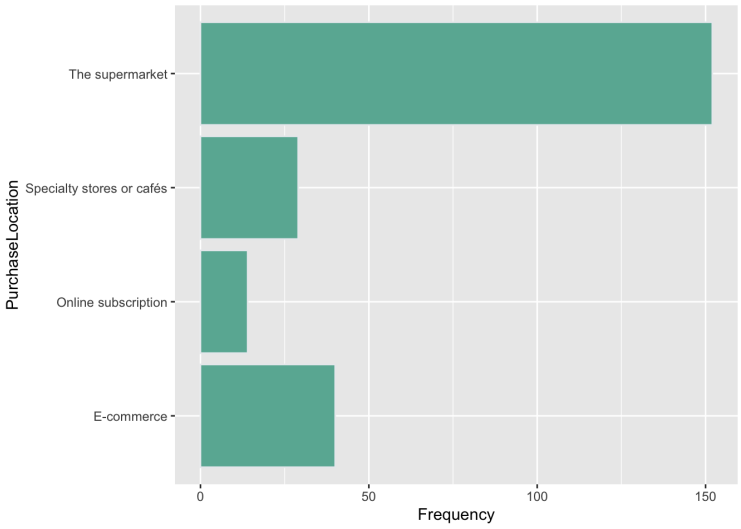
Brand choice	Absolute	Relative
Never	77	32.77%
Sometimes	132	56.17%
Very often	23	9.79%
Every time	3	1.28%

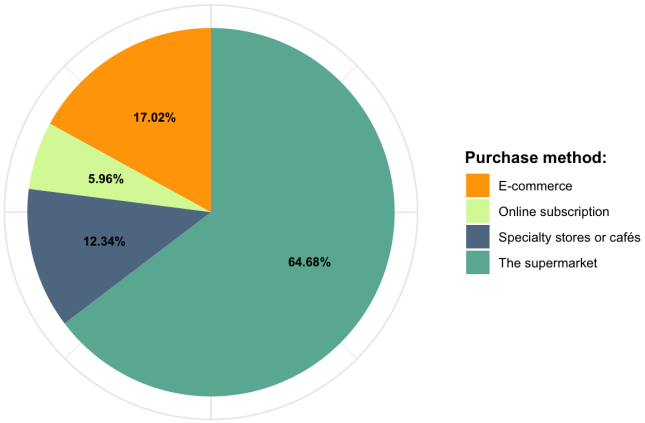




Purchase Method

Purchase Method	Absolute	Relative
E-commerce	40	17.02%
Online subscription	14	5.96%
Specialty stores or cafés	29	12.34%
The supermarket	152	64.68%

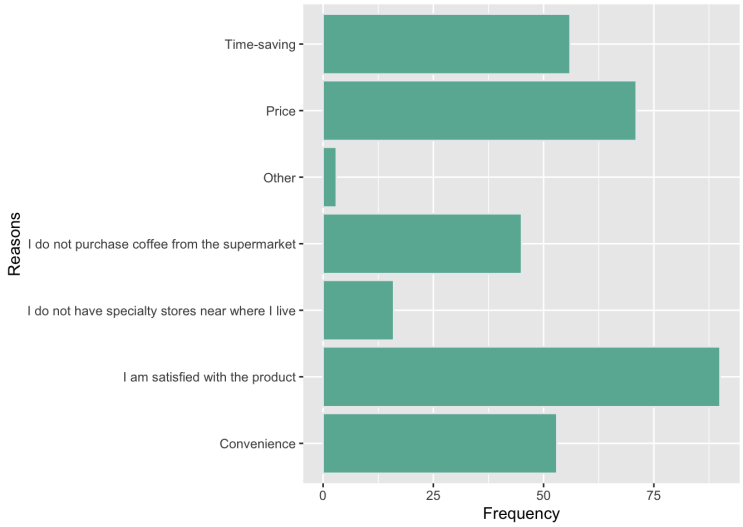

The supermarket
Specialty stores or cafés
Online subscription
E-commerce
PurchaseLocation
0
50
100
150
Frequency


17.02%
5.96%
12.34%
64.68%
Purchase method:
E-commerce
Online subscription
Specialty stores or cafés
The supermarket

Multiple option answers:

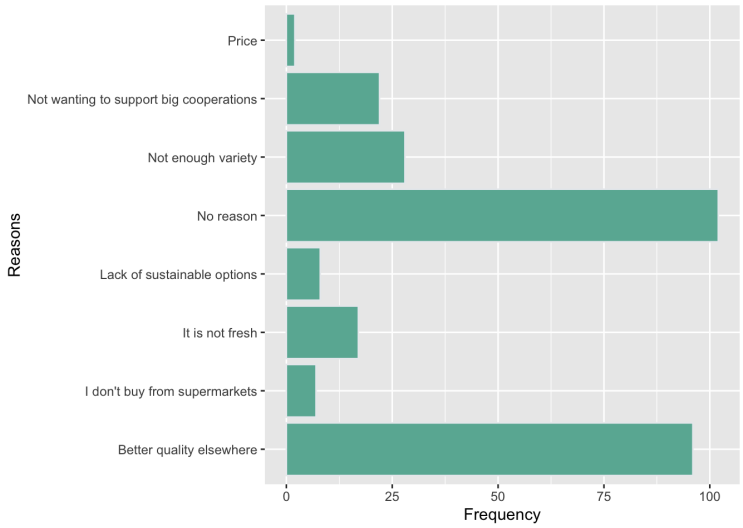
Reasons buying from the supermarket

Reasons	Frequency
I am satisfied with the product	90
Price	71
Time-saving	56
Convenience	53
I do not purchase coffee from the supermarket	45
I do not have specialty stores near where I live	16
Other	3



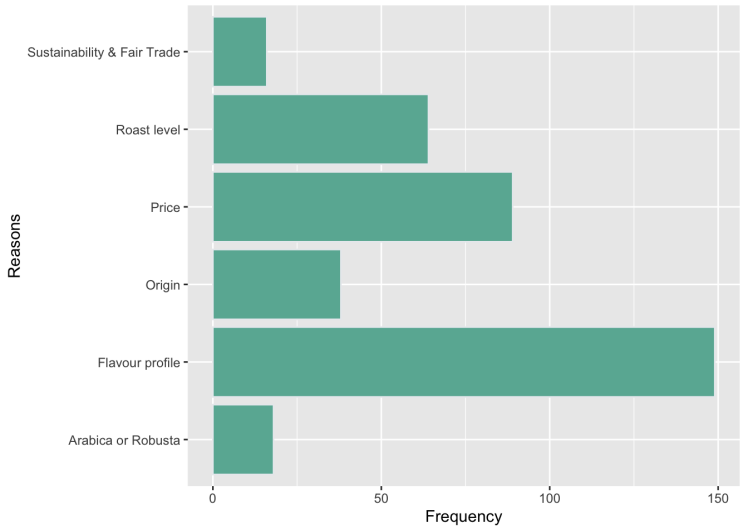
Reasons for not buying from the supermarket

Reasons	Frequency
No reason	102
Better quality elsewhere	96
Not enough variety	28
Not wanting to support big cooperations	22
It is not fresh	17
Lack of sustainable options	8
I don’t buy from supermarkets	7
Price	2



Criteria for choosing the type of coffee

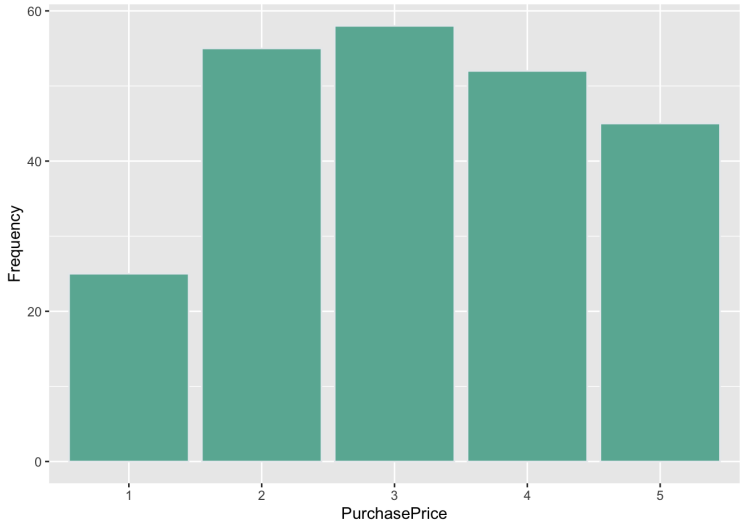
Reasons	Frequency
Flavour profile	149
Price	89
Roast level	64
Origin	38
Arabica or Robusta	18
Sustainability & Fair Trade	16

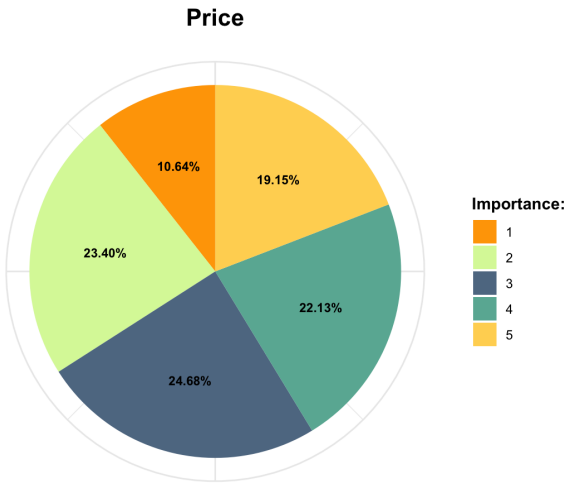


Purchase decisions 1-5

Price

Purchase decision - price	Absolute	Relative
1	25	10.64%
2	55	23.40%
3	58	24.68%
4	52	22.13%
5	45	19.15%

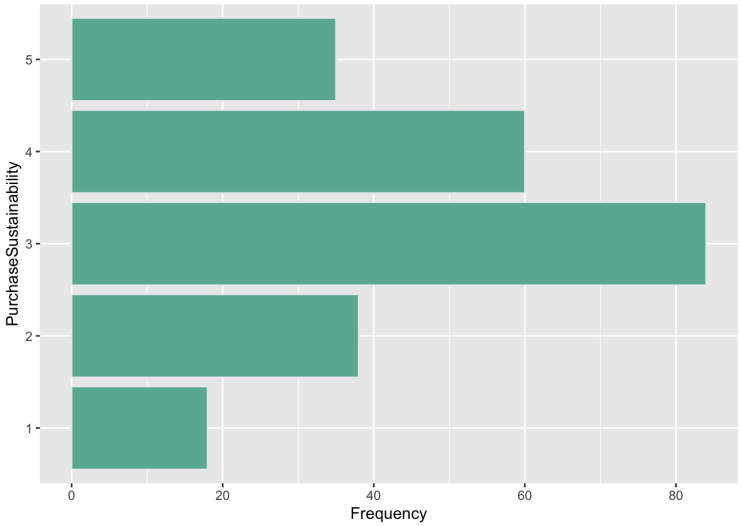


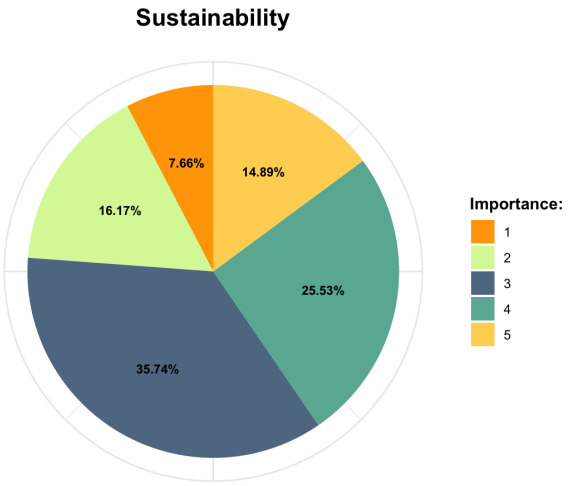


Sustainability

Purchase decision - sustainability	Absolute	Relative
1	18	7.66%

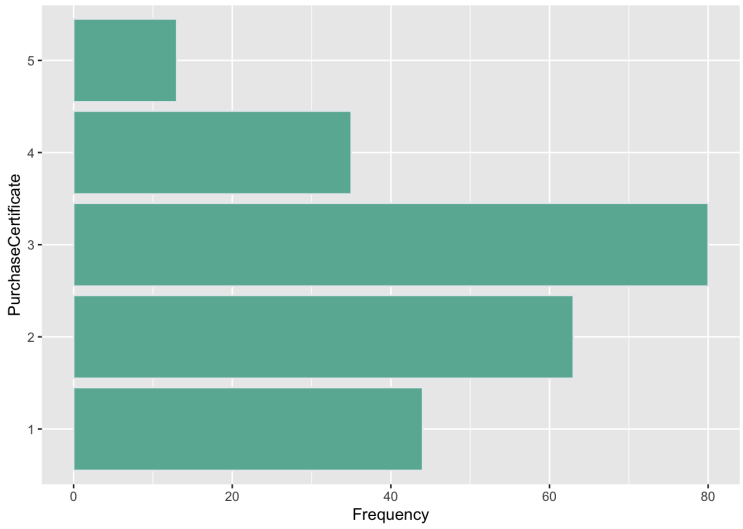
2	38	16.17%
3	84	35.74%
4	60	25.53%
5	35	14.89%



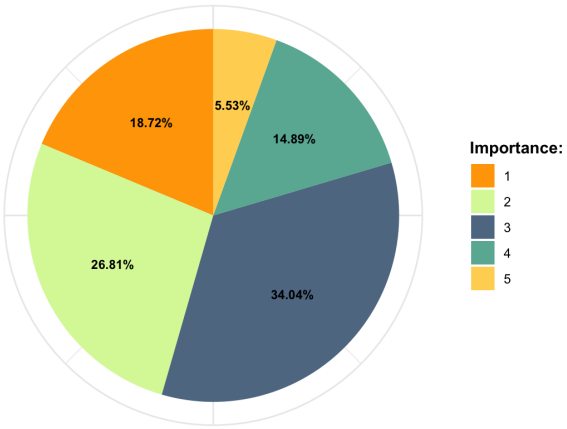


Certificates

Purchase decision - certificate	Absolute	Relative
1	44	18.72%
2	63	26.81%
3	80	34.04%
4	35	14.89%
5	13	5.53%

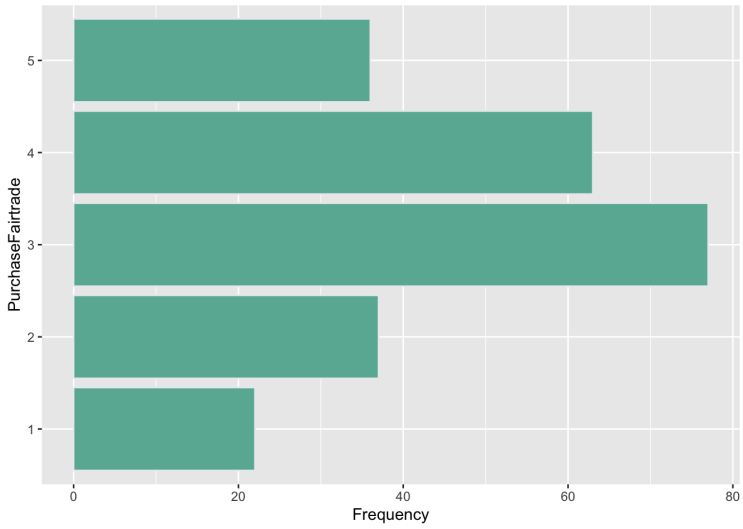


Certificate

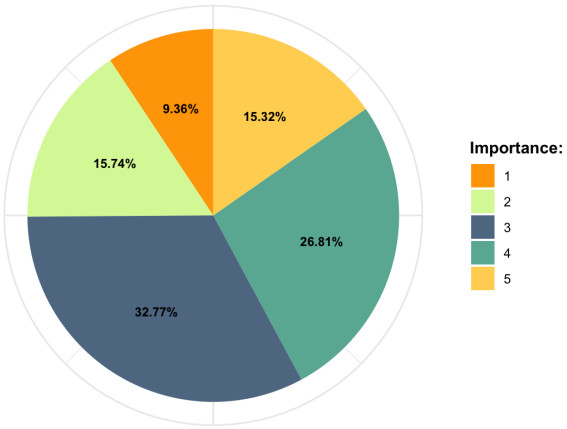


Fairtrade

Purchase decision - fairtrade	Absolute	Relative
1	22	9.36%
2	37	15.74%
3	77	32.77%
4	63	26.81%
5	36	15.32%

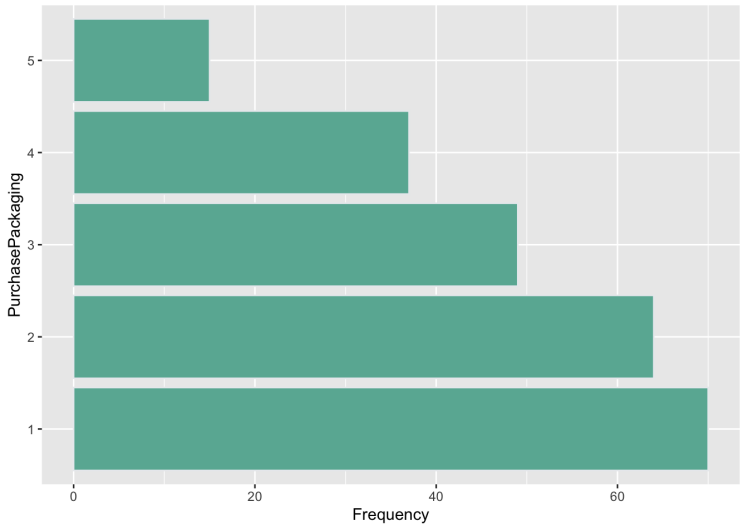


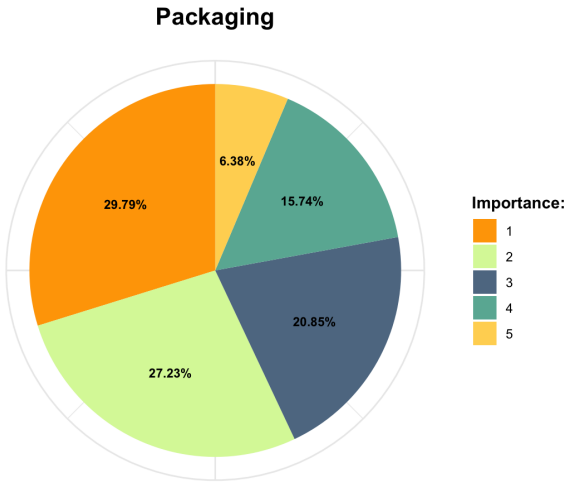
Fair trade



Packaging

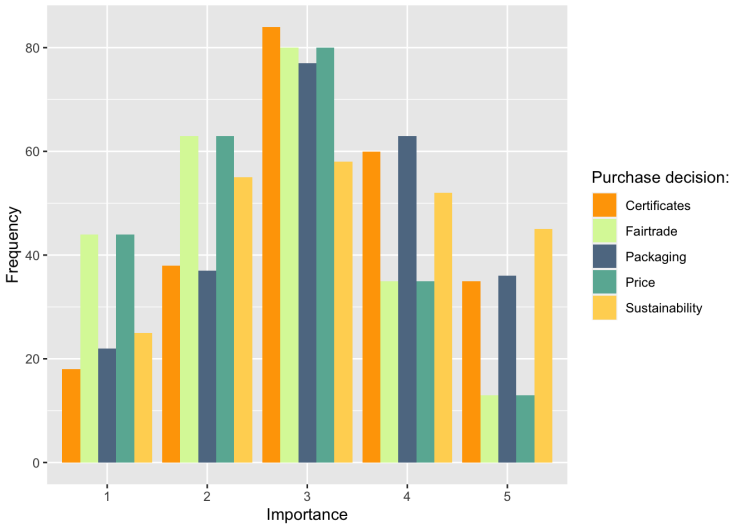
Purchase decision - packaging	Absolute	Relative
1	70	29.79%
2	64	27.23%
3	49	20.85%
4	37	15.74%
5	15	6.38%

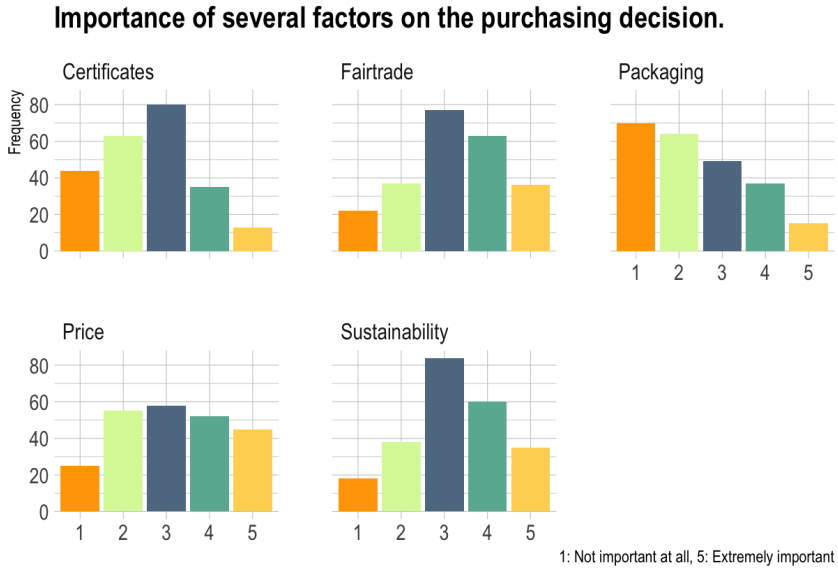

PurchasePackaging
5
4
3
2
1
0
20
40
60
Frequency


Packaging
6.38%
15.74%
29.79%
20.85%
27.23%
Importance:
1
2
3
4
5

Combined data

Importance	Price	Sustainability	Certificates	Fairtrade	Packaging
1	44	25	18	44	22
2	63	55	38	63	37
3	80	58	84	80	77
4	35	52	60	35	63
5	13	45	35	13	36

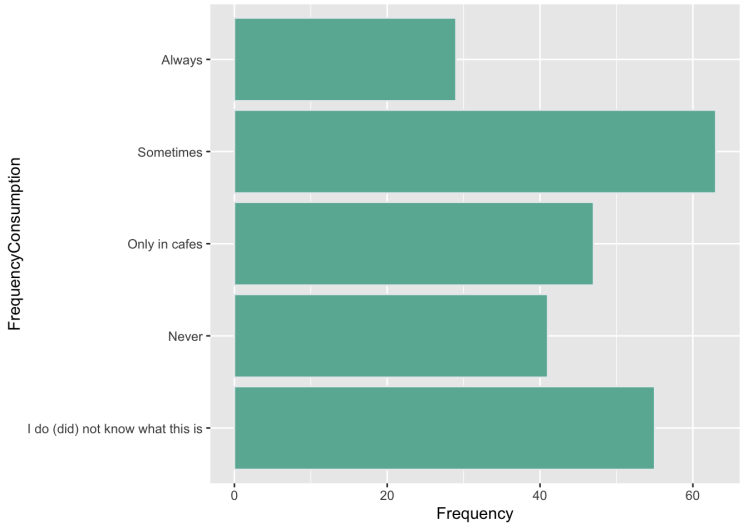




Frequency specialty coffee consumption

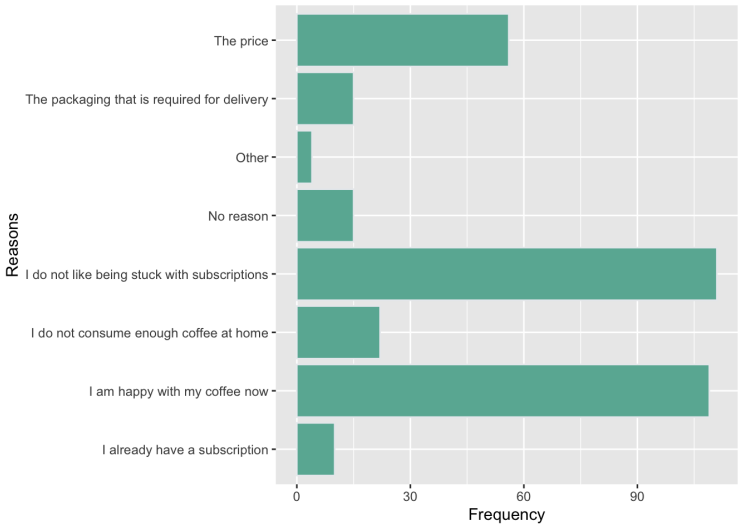
Frequency coffee consumption	Absolute	Relative

I do (did) not know what this is	55	23.40%
Never	41	17.45%
Only in cafes	47	20.00%
Sometimes	63	26.81%
Always	29	12.34%



Reasons for not being likely to set up a subscription

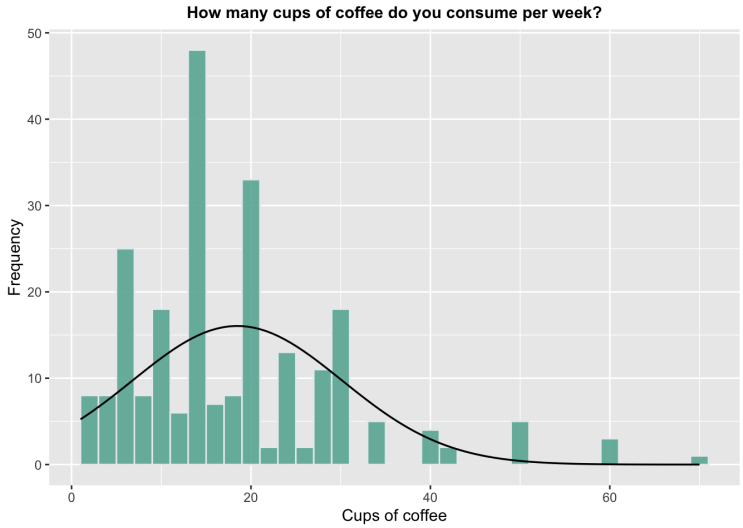
Reasons	Frequency
I do not like being stuck with subscriptions	111
I am happy with my coffee now	109
The price	56
I do not consume enough coffee at home	22
The packaging that is required for delivery	15
No reason	15
I already have a subscription	10
Other	4


The price
The packaging that is required for delivery
Other
No reason
I do not like being stuck with subscriptions
I do not consume enough coffee at home
I am happy with my coffee now
I already have a subscription
Reasons
0
30
60
90
Frequency

Univariate descriptions - Numerical variables

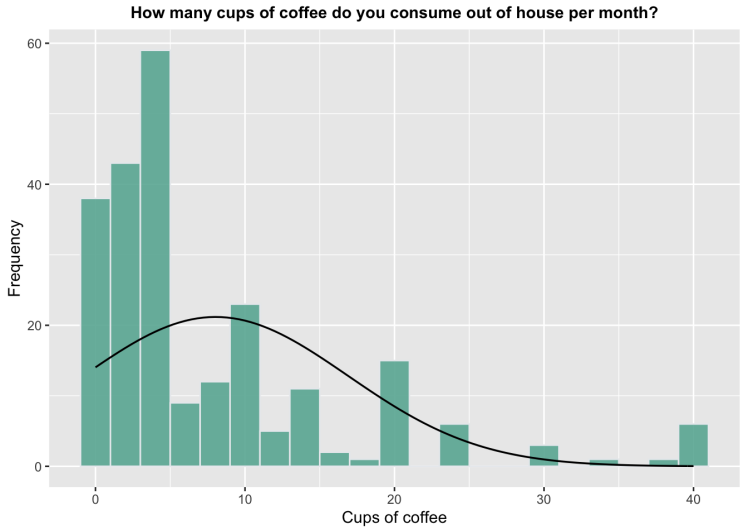
Amount coffe consumed weekly

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00   10.00   15.00   18.48   25.00   70.00
```



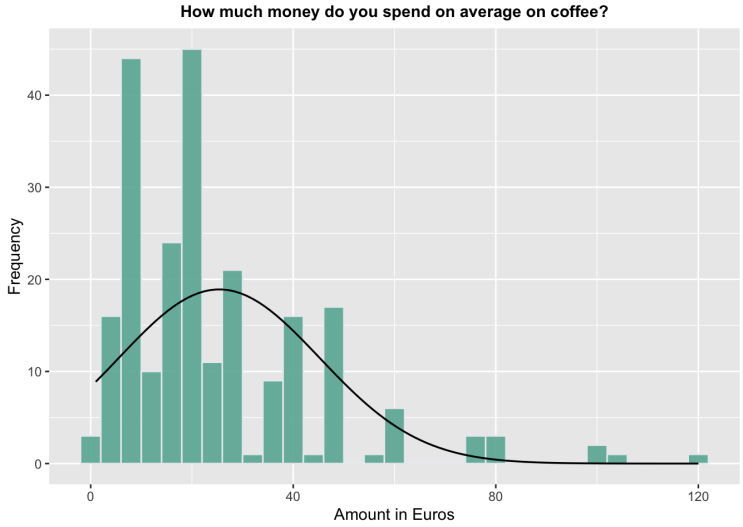
Amount per month out of house

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    2.00    5.00    8.03   10.00   40.00
```



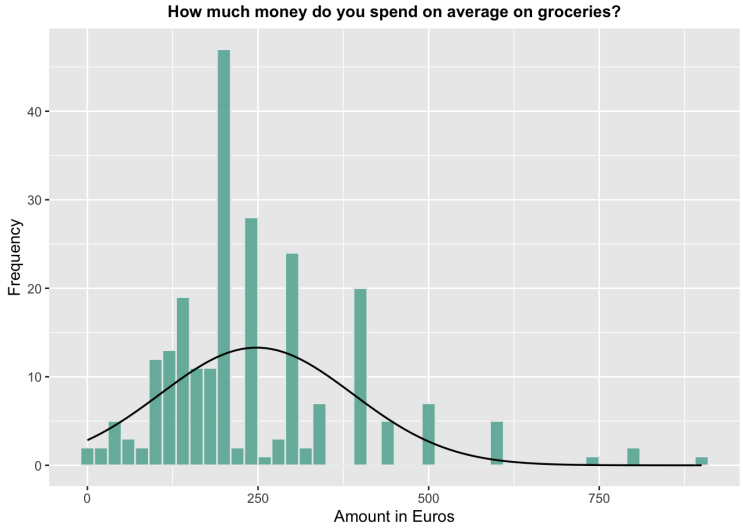
Money coffee

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00   10.00   20.00   25.38   35.00  120.00
```



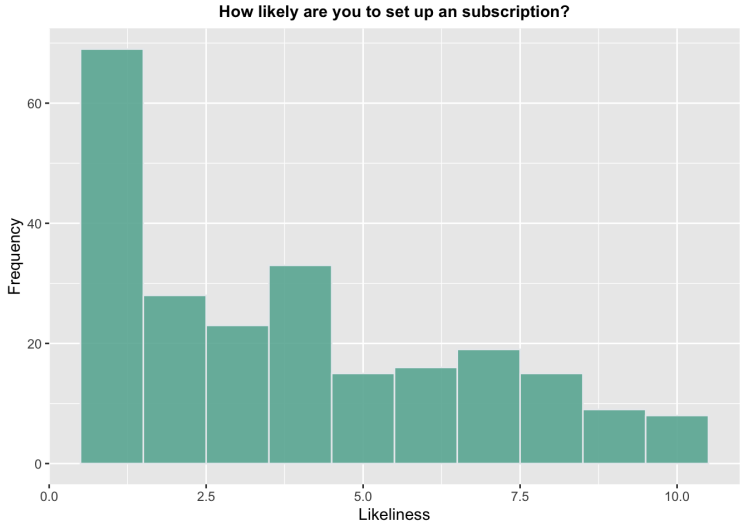
Money groceries

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    0.0   160.0   200.0   247.8   300.0   900.0 
```



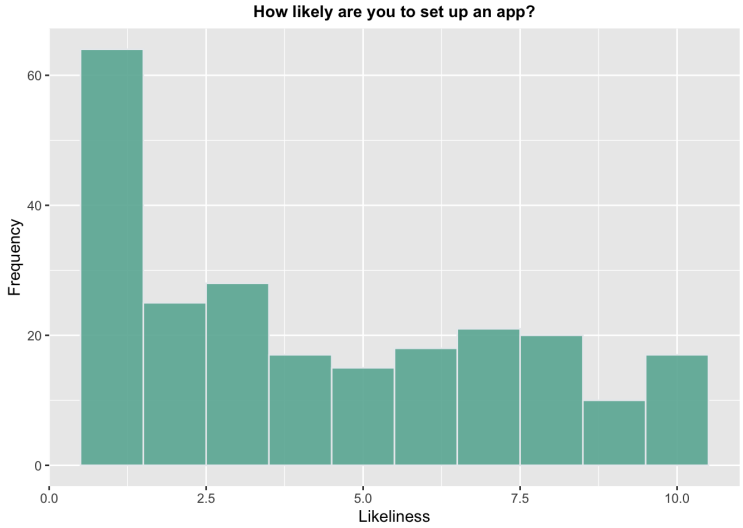
Subscription likely

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   1.000   3.000   3.877   6.000  10.000 
```

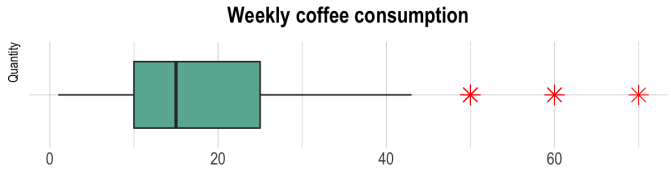


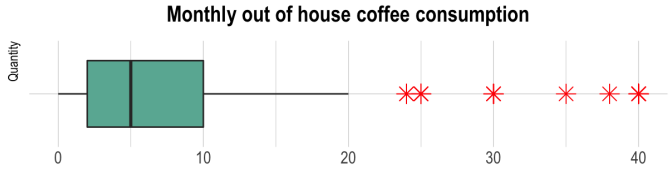
App likely

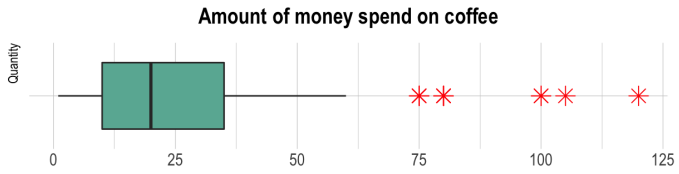
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   4.000   4.323   7.000  10.000
```

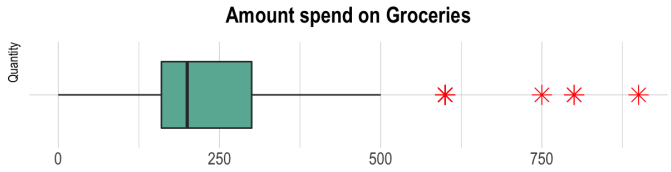


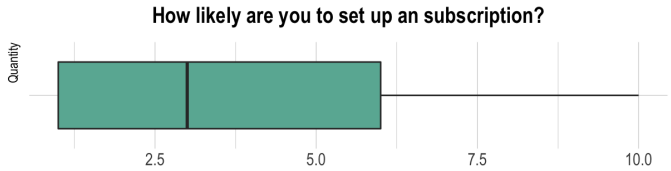
Boxplots

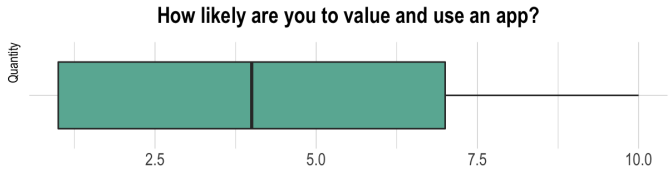

Weekly coffee consumption
Quantity
0
20
40
60


Monthly out of house coffee consumption
Quantity
0
10
20
30
40


Amount of money spend on coffee
Quantity
0
25
50
75
100
125


Amount spend on Groceries
Quantity
0
250
500
750


How likely are you to set up an subscription?
Quantity
2.5
5.0
7.5
10.0


How likely are you to value and use an app?
Quantity
2.5
5.0
7.5
10.0

Parametric testing

H_0 <- There is no association between the two variables.
H_a <- There is a association.

Age - Amount coffee drank

```
    Pearson's Chi-squared test

data:  AmountWeek and AgeCategory
X-squared = 241.68, df = 136, p-value = 0.00000006432


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  AmountWeek and AgeCategory
X-squared = 241.68, df = NA, p-value = 0.02595
```

Education - Amount coffee drank

```
    Pearson's Chi-squared test

data:  AmountWeek and Education
X-squared = 229.99, df = 170, p-value = 0.001491


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  AmountWeek and Education
X-squared = 229.99, df = NA, p-value = 0.06786
```

Gender - Amount coffee drank

```
    Pearson's Chi-squared test

data:  AmountWeek and Gender
X-squared = 69.019, df = 68, p-value = 0.4427


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  AmountWeek and Gender
X-squared = 69.019, df = NA, p-value = 0.3653
```

Home - Amount coffee drank

```
    Pearson's Chi-squared test

data:  AmountWeek and Home
X-squared = 66.506, df = 68, p-value = 0.5286


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)
```

```
data:  AmountWeek and Home
X-squared = 66.506, df = NA, p-value = 0.5469
```

App - Age

```
    Pearson's Chi-squared test

data:  App_Likely and AgeCategory
X-squared = 58.189, df = 36, p-value = 0.01103


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  App_Likely and AgeCategory
X-squared = 58.189, df = NA, p-value = 0.01597
```

Coffee knowledge - Age

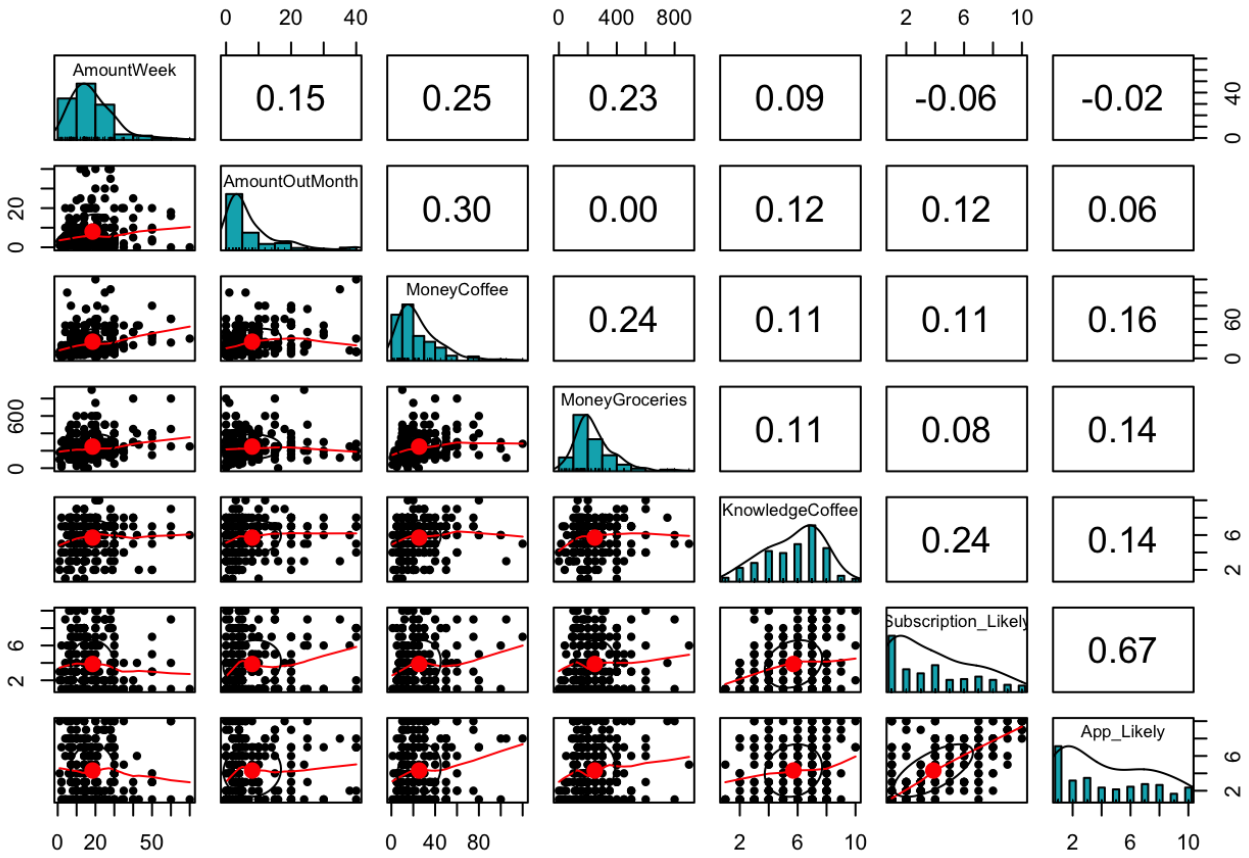
```
    Pearson's Chi-squared test

data:  KnowledgeCoffee and AgeCategory
X-squared = 154.32, df = 36, p-value < 0.00000000000000022


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  KnowledgeCoffee and AgeCategory
X-squared = 154.32, df = NA, p-value = 0.001996
```

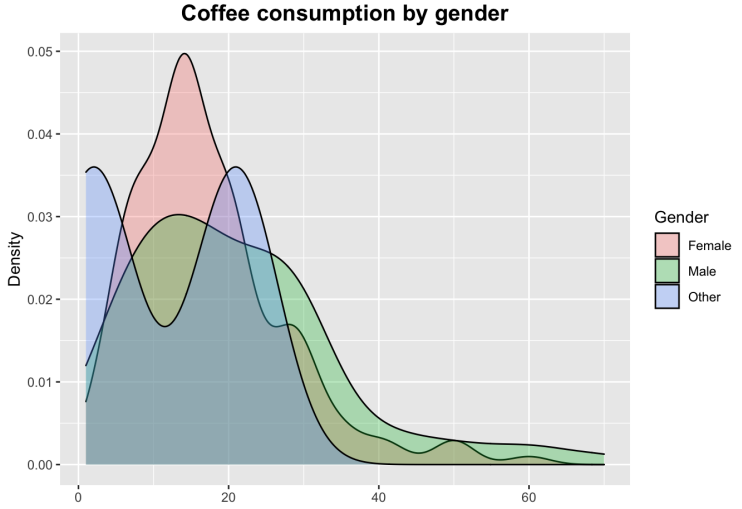
Coffee knowledge - Purchase location

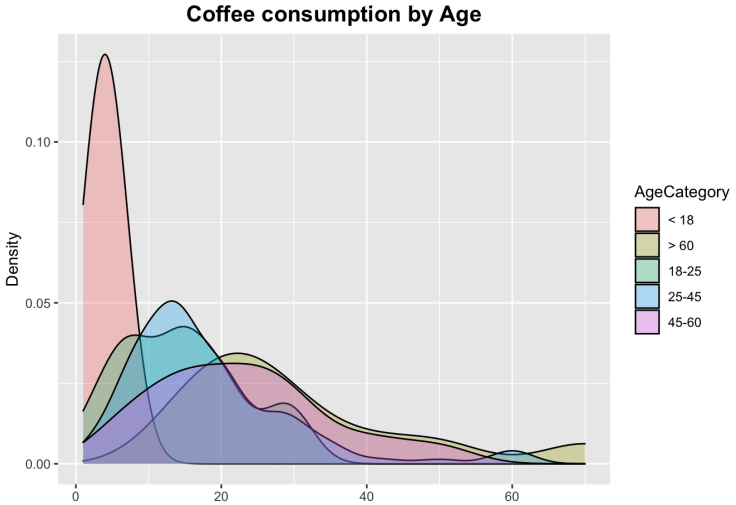
```
    Pearson's Chi-squared test

data:  KnowledgeCoffee and PurchaseLocation
X-squared = 34.489, df = 27, p-value = 0.1523


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  KnowledgeCoffee and PurchaseLocation
X-squared = 34.489, df = NA, p-value = 0.1597


    Pearson's Chi-squared test

data:  Subscription_Likely and App_Likely
X-squared = 347.04, df = 81, p-value < 0.00000000000000022


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  Subscription_Likely and App_Likely
X-squared = 347.04, df = NA, p-value = 0.001996


    Pearson's Chi-squared test
```

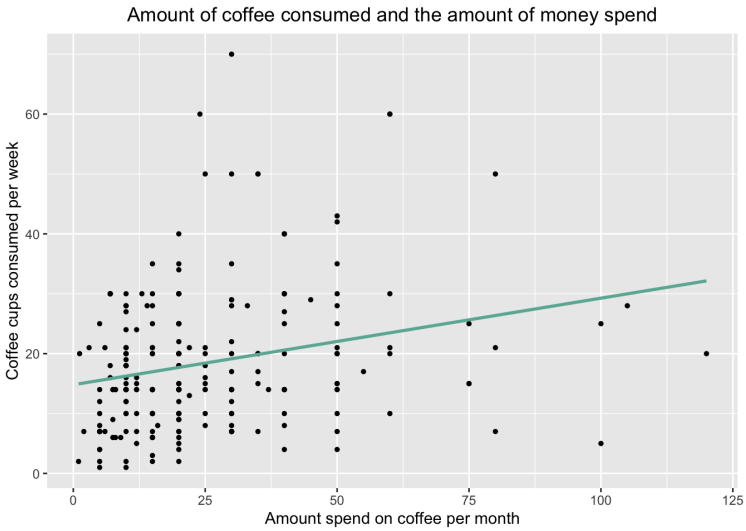
```
data:  Subscription_Likely and KnowledgeCoffee
X-squared = 109.94, df = 81, p-value = 0.01789


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  Subscription_Likely and KnowledgeCoffee
X-squared = 109.94, df = NA, p-value = 0.01397


    Pearson's Chi-squared test

data:  Subscription_Likely and AmountWeek
X-squared = 311.13, df = 306, p-value = 0.4078


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  Subscription_Likely and AmountWeek
X-squared = 311.13, df = NA, p-value = 0.4411


    Pearson's Chi-squared test

data:  Subscription_Likely and Frequency_Specialty
X-squared = 102.57, df = 36, p-value = 0.00000002601


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  Subscription_Likely and Frequency_Specialty
X-squared = 102.57, df = NA, p-value = 0.001996


    Pearson's Chi-squared test

data:  Subscription_Likely and BrandChange
X-squared = 38.718, df = 27, p-value = 0.06719


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  Subscription_Likely and BrandChange
X-squared = 38.718, df = NA, p-value = 0.0978


    Pearson's Chi-squared test

data:  Subscription_Likely and PurchaseLocation
X-squared = 61.31, df = 27, p-value = 0.0001772


    Pearson's Chi-squared test with simulated p-value (based on 500
    replicates)

data:  Subscription_Likely and PurchaseLocation
X-squared = 61.31, df = NA, p-value = 0.001996
```

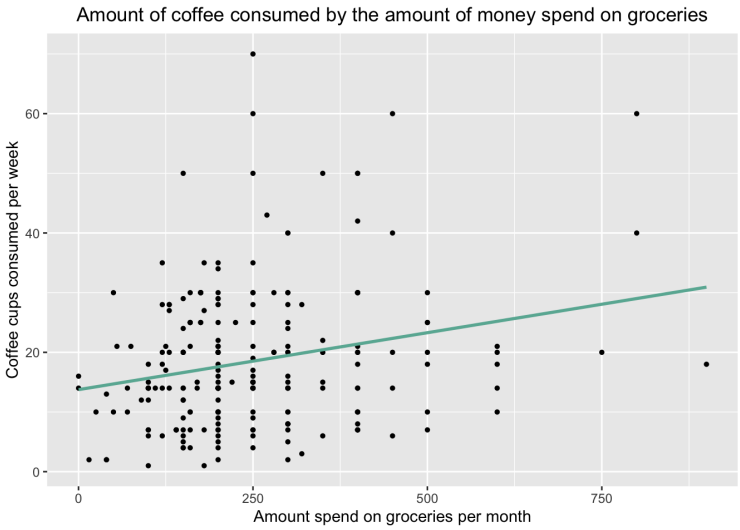
Relationships






Coffee consumption by Age
Density
0.10
0.05
0.00
0
20
40
60
AgeCategory
< 18
> 60
18-25
25-45
45-60


Amount of coffee consumed and the amount of money spend
Coffee cups consumed per week
60
40
20
0
0
25
50
75
100
125
Amount spend on coffee per month


Amount of coffee consumed by the amount of money spend on groceries
Coffee cups consumed per week
60
40
20
0
0
250
500
750
Amount spend on groceries per month

Regressions

```
=================================================================
                              Dependent variable:
                   ---------------------------------------------
                              Subscription_Likely
                            (1)                     (2)
-----------------------------------------------------------------
KnowledgeCoffee           0.324***                0.325***
                           (0.087)                (0.088)

Purchase_Fairtrade        0.384***
                           (0.147)

AmountWeek                                        -0.027*
                                                  (0.015)

MoneyCoffee                                        0.016*
                                                  (0.009)

Constant                    0.787                 2.108***
                           (0.700)                (0.586)

-----------------------------------------------------------------
Observations                 235                    235
R2                          0.084                  0.077
Adjusted R2                 0.076                  0.065
Residual Std. Error    2.633 (df = 232)        2.649 (df = 231)
F Statistic         10.601*** (df = 2; 232) 6.406*** (df = 3; 231)
=================================================================
Note:                                  *p<0.1; **p<0.05; ***p<0.01

   KnowledgeCoffee Purchase_Fairtrade
          1.000477           1.000477

KnowledgeCoffee      AmountWeek     MoneyCoffee
       1.016549        1.068581        1.072962
```