CHAMATH DE SILVA

2017/s/16156

S13825

**dataset 1: bank client contact**

**EXPLORATORY DATA ANALYSIS (EDA)**

**ST 3011 STATISTICAL PROGAMMING**

**assignment 01**

* **Data**

**Response variable:**

* **duration**: last contact duration, in seconds **(numeric)**

**Explanatory variables:**

1. **age**: ages of clients (**numeric**)
2. **job**: type of job (**categorical**: "admin.","blue-collar","entrepreneur","housemaid","management","retired","self-employed","services","student","technician","unemployed","unknown")
3. **marital**: marital status (**categorical**: "divorced","married","single","unknown")
4. **education:**  (**categorical**: "basic.4y","basic.6y","basic.9y","high. school","illiterate","professional. course","university. degree","unknown")
5. **default**: has credit in default (**categorical**: "no","yes","unknown")
6. **housing:** has housing loan (**categorical**: "no","yes","unknown")
7. **contact**: contact communication type (**categorical**: "cellular","telephone")
8. **month**: last contact month of year (**categorical:** "jan", "feb", "mar", ..., "nov", "dec")

* **Data Preparation**

All the data were carefully studied before being used for analysis to identify whether there is any missing values or unusual observations. Since missing values or unusual observations did not include, data preparation processors were not conducted.

* **Univariate Analysis**

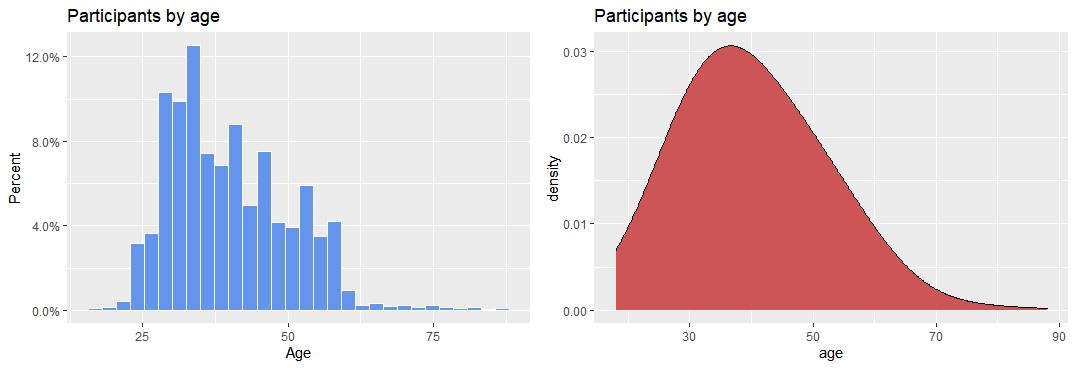
1. **Age:**

* **Summary Statistics for Age:**

****

According to this descriptive statistics summary it is clear that mean of Age is 40.11 years. All Ages are lies between 18 years and 88 years

* **Histogram and Density graph of age:**

**** This is the histogram and density graph for age of bank clients. When we consider the distribution of the age of bank clients, it is noticed as a unimodal and positive skewness. So, most of the ages of clients below 50 years.

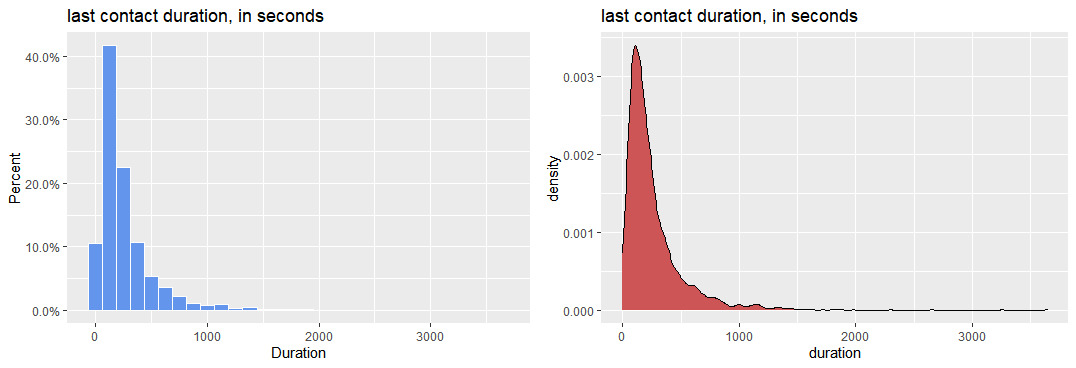
1. **Duration:**

* **Summary Statistics for Duration:**

****

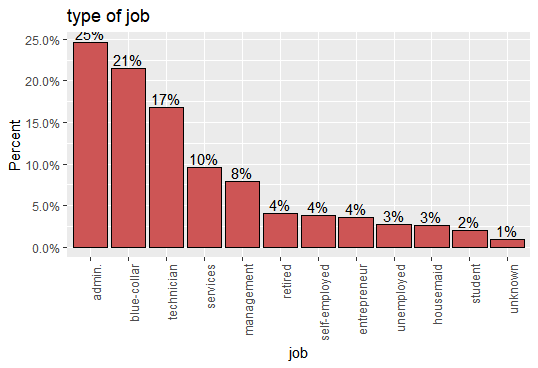
According to this descriptive statistics summary it is clear that mean of Duration is 256.8 seconds. All duration times are below the 3643 seconds.

* **Histogram and Density graph of duration**:



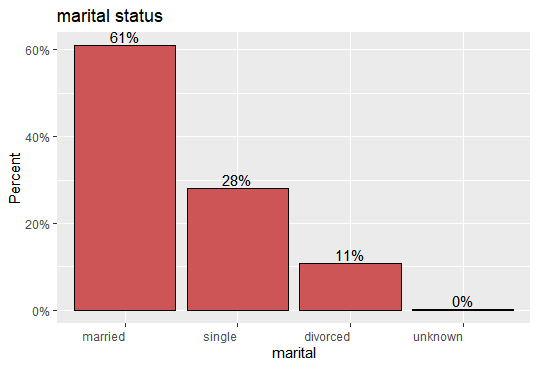
This is the histogram and density graph for last contact duration of bank clients. When we consider the distribution of the last contact duration of bank clients, it is noticed as a unimodal and highly positive skewness. So, most of the ages of clients below 1000 seconds.

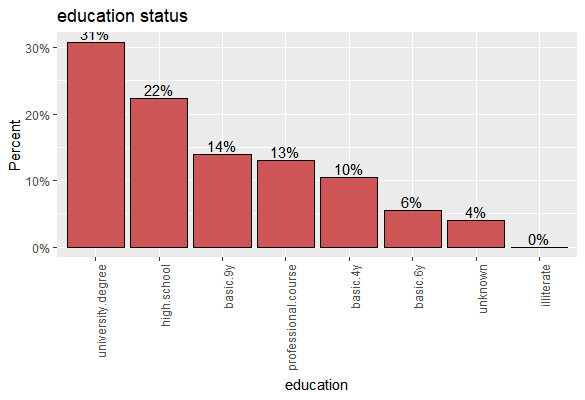
1. **Type of job**

****

This is the Bar Graph of job type of bank clients. According to a graph it is shown that most clients have an admin job. It is 25% of the total population. The lowest number are students. It is 2%. 1% of clients do not know their job.

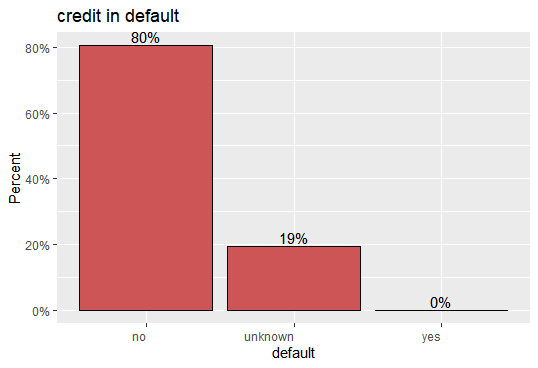
1. **marital status**

****

This is the Bar Graph of marital status of bank clients. According to a graph it is shown that most clients are married. It is 61% of the total population. The lowest number of clients are divorced. It is 11%. 5. **5. Education background **

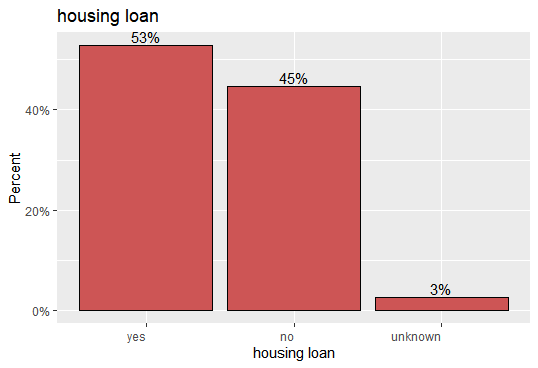
This is the Bar Graph of Education background of bank clients. According to a graph it is shown that most of clients have university degree. It is 31% of the total population. No one here is they have illiterate education background. 4% of clients do not know their educational status.

1. **Has credit in default**

****

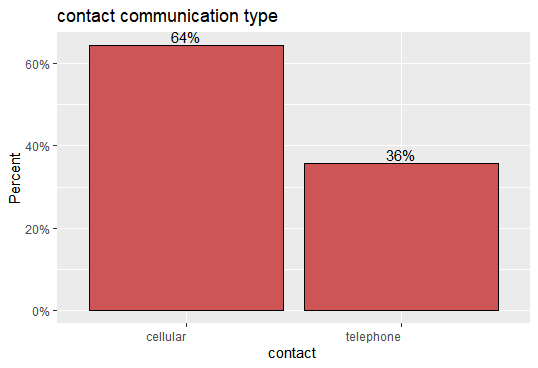
This is the Bar Graph of Default deposit of bank clients. According to a graph it is shown that most of clients have not default deposit. It is 80% of the total population. 19% of clients do not know their default credit. No one here has default credit.

1. **Housing loan**

****

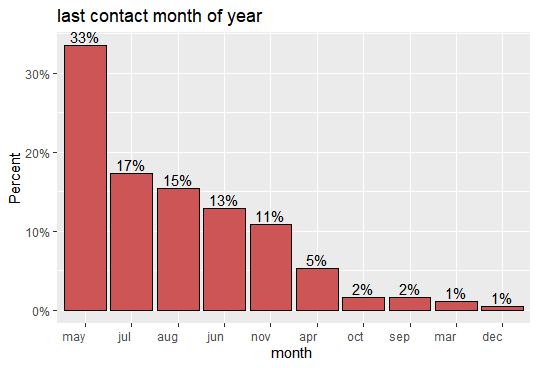
This is the Bar Graph of Housing loan of bank clients. According to a graph it is shown that most of clients have housing loan. 3% of clients do not know they Have housing loan or any other loan.

1. **Contact communication type**

****

This is the Bar Graph of communication contact type of bank clients. According to a graph it is shown that most of clients use cellular for contact to bank. It is 64% of the total population. 36% of clients use telephone for contact bank.

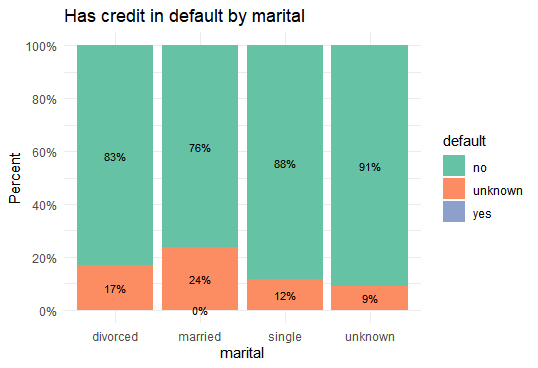
1. **last contact month of year**

****

This is the Bar Graph of last contact month of bank clients. According to a graph it is shown that most of clients contact with bank in May. It is 33% of the total population. Lowest percentage of customers contact in bank at December.

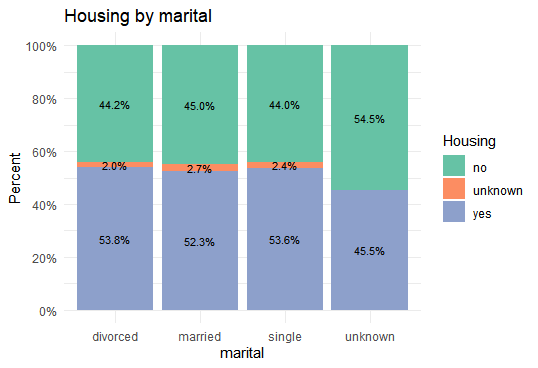
* **Bivariate Analysis**
* **Categorical vs categorical**

1. **Credit in default by marital**

****

According to the stack bar graph 82 percentage of the divorced does not have default credit while 76% of the married also does not have default credit.

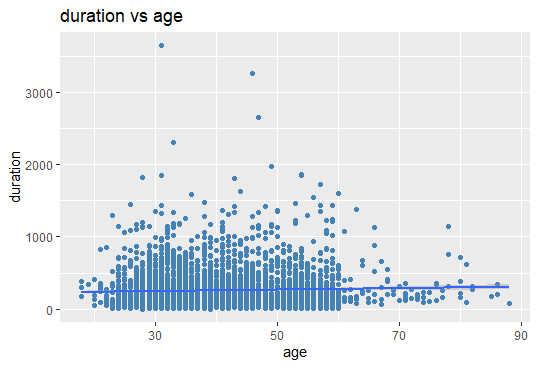
1. **Housing loan by marital**

****

And according to the stack bar graph 53% of the category who are divorced has a housing loan and 52.3 percent of the married do not have a housing loan.

* **Numerical vs numerical**

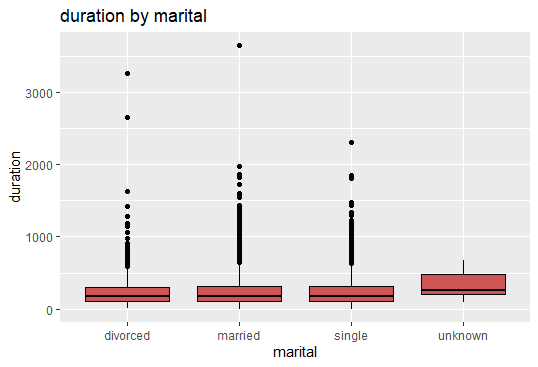
1. **Duration vs Age**

****

According to scatterplot, clear relationship cannot be observed within the duration Vs age scatterplot although a slight positive relationship can be noted. And also, most of the age limits are under 60.

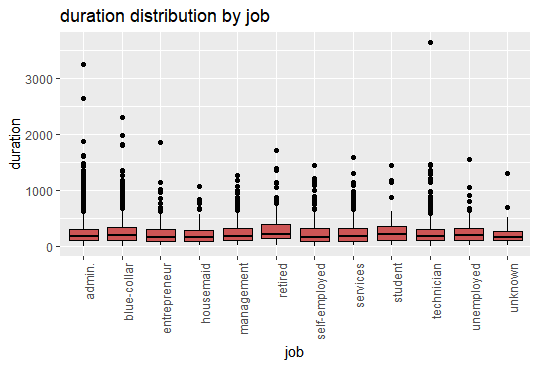
* **Numerical vs Categorical**

1. **Duration by marital**

****

When you consider the median of the people who aren't familiar with the marital status tend to have a higher median than the others. And the values of the medians of the categories divorced, married and single seems to be the same. Certain outliers can be noted in duration of the categories divorced, married and single.

1. **Duration by type of job**

****

When you consider the median duration of the retired it is slightly higher than the other groups. And the median duration of the other jobs seems to be a bit same. In almost every job there seems to be outliers.