Project 1 - Travel Insurance Data

Kevin Lopez, Clarisse Sarah, and Mayank Thacker

2024-02-18

# Explain the dataset

The dataset is a travel insurance dataset. It contains information on travelers and whether they purchased travel insurance or not and other information (age, income, Employment, etc). The dataset contains 10 columns and 1986 rows of information.

### The columns are:

**Age**- Age Of The Customer

**Employment Type**- The Sector In Which Customer Is Employed

**GraduateOrNot**- Whether The Customer Is College Graduate Or Not

**AnnualIncome**- The Yearly Income Of The Customer In Indian Rupees[Rounded To Nearest 50 Thousand Rupees]

**FamilyMembers**- Number Of Members In Customer’s Family

**ChronicDisease**- Whether The Customer Suffers From Any Major Disease Or Conditions Like Diabetes/High BP or Asthama,etc.

**FrequentFlyer**- Derived Data Based On Customer’s History Of Booking Air Tickets On Atleast 4 Different Instances In The Last 2 Years[2017-2019].

**EverTravelledAbroad**- Has The Customer Ever Travelled To A Foreign Country[Not Necessarily Using The Company’s Services]

**TravelInsurance**- Did The Customer Buy Travel Insurance Package During Introductory Offering Held In The Year 2019.

# Load the Dataset

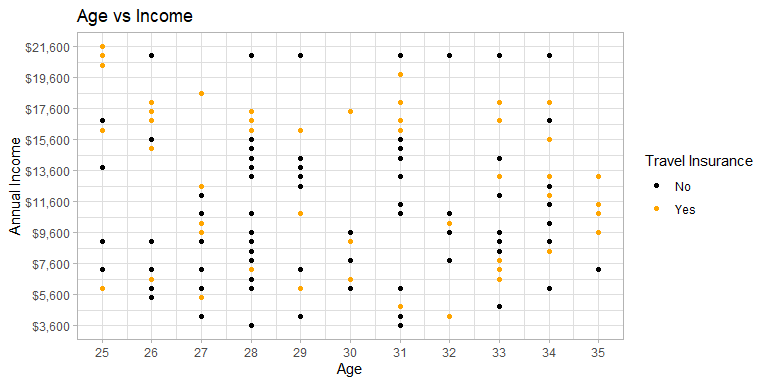
| **X** | **Age** | **Employment.Type** | **GraduateOrNot** | **AnnualIncome** | **FamilyMembers** | **ChronicDiseases** | **FrequentFlyer** | **EverTravelledAbroad** | **TravelInsurance** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 31 | Government Sector | Yes | 400,000 | 6 | 1 | No | No | 0 |
| 1 | 31 | Private Sector/Self Employed | Yes | 1,250,000 | 7 | 0 | No | No | 0 |
| 2 | 34 | Private Sector/Self Employed | Yes | 500,000 | 4 | 1 | No | No | 1 |
| 3 | 28 | Private Sector/Self Employed | Yes | 700,000 | 3 | 1 | No | No | 0 |
| 4 | 28 | Private Sector/Self Employed | Yes | 700,000 | 8 | 1 | Yes | No | 0 |
| 5 | 25 | Private Sector/Self Employed | No | 1,150,000 | 4 | 0 | No | No | 0 |

## reshaping data

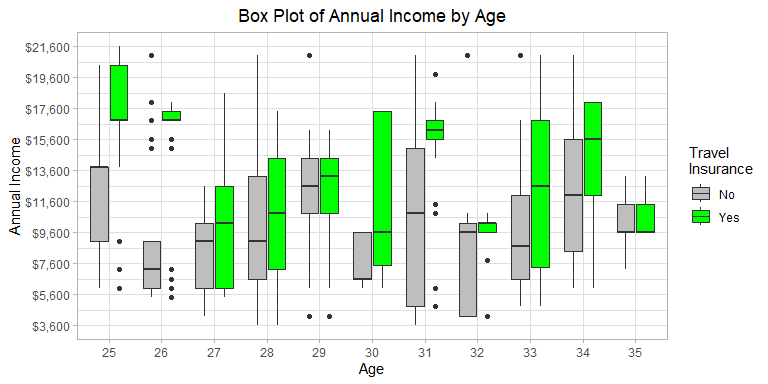
Data is already in long format. Data for income is given in Indian Rupees. We will create a new column to convert it to USD. Column name will be AnnualIncomeUSD

| **X** | **Age** | **Employment.Type** | **GraduateOrNot** | **AnnualIncome** | **FamilyMembers** | **ChronicDiseases** | **FrequentFlyer** | **EverTravelledAbroad** | **TravelInsurance** | **AnnualIncomeUSD** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 31 | Government Sector | Yes | 400,000 | 6 | 1 | No | No | 0 | 4,800 |
| 1 | 31 | Private Sector/Self Employed | Yes | 1,250,000 | 7 | 0 | No | No | 0 | 15,000 |
| 2 | 34 | Private Sector/Self Employed | Yes | 500,000 | 4 | 1 | No | No | 1 | 6,000 |
| 3 | 28 | Private Sector/Self Employed | Yes | 700,000 | 3 | 1 | No | No | 0 | 8,400 |
| 4 | 28 | Private Sector/Self Employed | Yes | 700,000 | 8 | 1 | Yes | No | 0 | 8,400 |
| 5 | 25 | Private Sector/Self Employed | No | 1,150,000 | 4 | 0 | No | No | 0 | 13,800 |

# Graph 2 - Age/Income

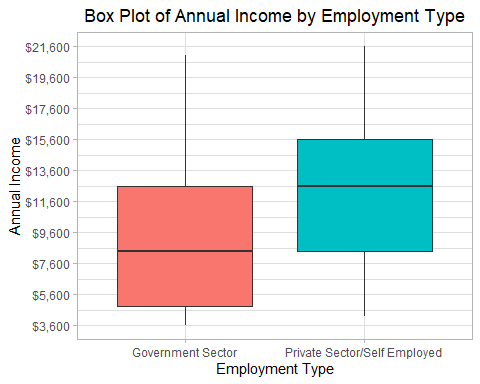
Age and income is a cool graph to see. We can see how income is distributed by age. - Generally it is believed that income increases with age. - Peak around 40-50 since that is when people are at their peak in their careers. ## What does the data tell us - There is not a big difference in the age and the income that these people make. ## By looking at a barplot for each age, - There is a slight increase of people who buy insurance as they make more money. - It is easier to tell in box plot compared to scatter plot. 

## Graph 2 - Age/Income as boxplot



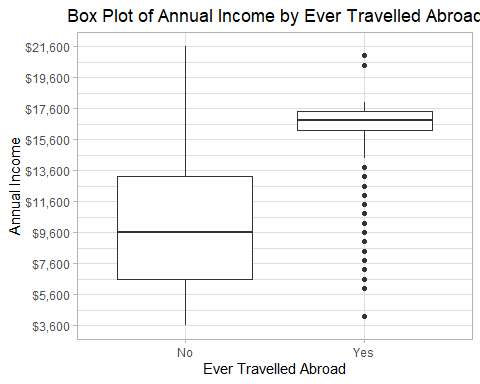
# Graph 3 - Sector/Income

* We can see how the income is distributed by Employment Type (Goverment vs Private).
* We can see that the income is slightly higher in the private sector.



# Ever Travelled Abroad and anual income

#scatter plot



# Graph 4 - graduate or not and anual income

