Employment Status Prediction

Which is the most significant variable effecting the employment as well as unemployment?

- I have found importance of each variable using Random Forest Model to find out which one has the most effect on the employment status.
- From the result, I can see that the highest is of variable "**Weekly Hours Worked**", 0.35. Thus it is the most significant variable effecting both employment as well as unemployment.

```
Variable: Weekly Hours Worked  Importance: 0.35
Variable: Age                  Importance: 0.12
Variable: Sleeping             Importance: 0.06
Variable: Socializing & Relaxing Importance: 0.06
Variable: Job Searching        Importance: 0.05
Variable: Eating and Drinking  Importance: 0.05
Variable: Television           Importance: 0.05
Variable: Grooming             Importance: 0.04
Variable: Weekly Earnings      Importance: 0.03
Variable: Year                 Importance: 0.03
Variable: Housework            Importance: 0.03
Variable: Food & Drink Prep    Importance: 0.03
Variable: Shopping             Importance: 0.03
Variable: Education Level      Importance: 0.02
Variable: Children             Importance: 0.02
Variable: Caring for Children  Importance: 0.02
Variable: Gender               Importance: 0.01
Variable: Playing with Children Importance: 0.01
Variable: Volunteering         Importance: 0.01
Variable: Golfing              Importance: 0.0
Variable: Running              Importance: 0.0
```

Predict the values for employment variable in the test data set.

- I have used Neural Networks from Scikit learn to predict the employment status of people.
- The accuracy of the neural network when tested on the training dataset using k-cross validation was 94.3%.
- The prediction model is implemented using Python.
- I have used pandas library of Python for feature engineering, and numpy for data processing.
- I have used the following notations:
  - Gender:
    - Male: 1, Female: 2
  - Employment Status:
    - Unemployed: 3, Employed: 4, Not in labor: 5
  - Education:
    - 9th grade: 9, 10th grade: 10, 11th grade: 11, 12th grade: 12, High School: 13
    - Associate Degree: 14, Bachelor: 15, Master: 16, Doctoral Degree: 17
    - Prof. Degree: 18, Some College: 19

Summary in 2012(Average Time Spend on Activities)

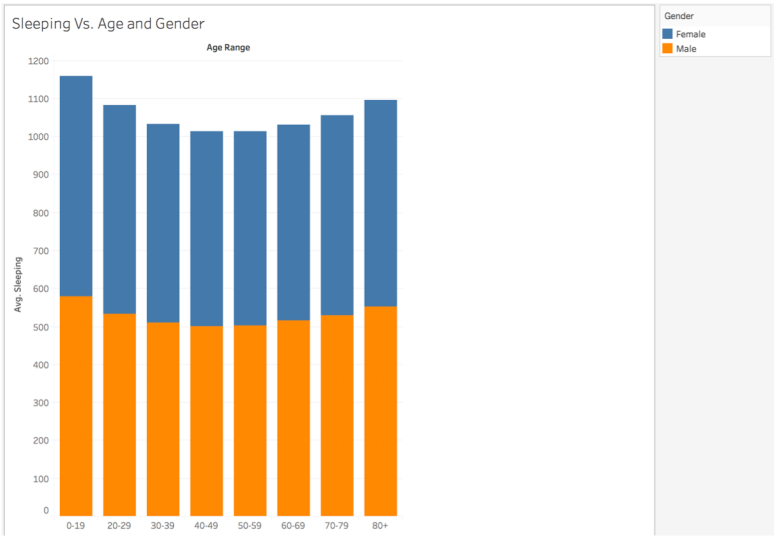
Activity	Average Time Spent
Sleeping	529.5290589
Grooming	40.39845019
Housework	41.43
Food & Drink Prep	33.16
Caring for Children	27.57
Playing with Children	8.82
Job Search	2.13
Shopping	22.67
Eating & Drinking	69.34
Socializing & Relaxing	298.22
Television	173.29
Golfing	1.49
Running	0.81
Volunteering	9.24

Summary in 2012(Standard Deviation in Time Spend on Activities)

Activity	Standard Deviation in Time Spent
Sleeping	136.8336947
Grooming	37.54
Housework	85.60
Food & Drink Prep	51.08
Caring for Children	73.77
Playing with Children	39.81
Job Search	24.88
Shopping	46.58
Eating & Drinking	54.12
Socializing & Relaxing	210.96
Television	176.24
Golfing	19.69
Running	8.45
Volunteering	45.98

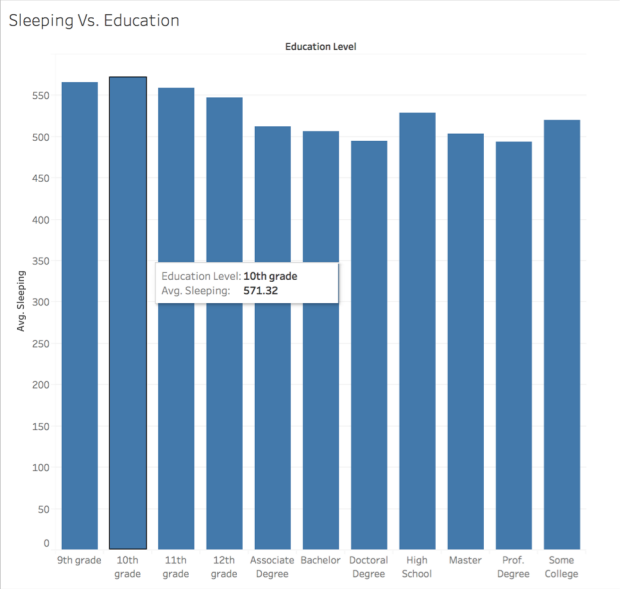
Change in sleeping based on age

- The age range 0-19 has the highest average sleeping time. Then I see a gradual decrease up till the age range 40-49 and then the time increases for the further ranges.



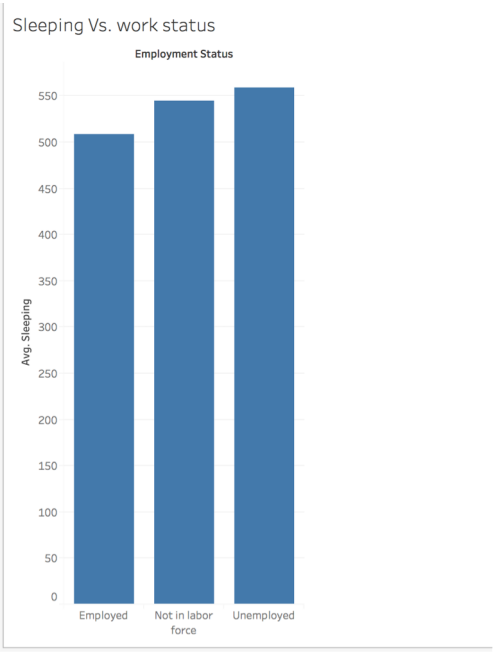
Change in sleeping based on education

- Higher averages for lower education and vice versa.



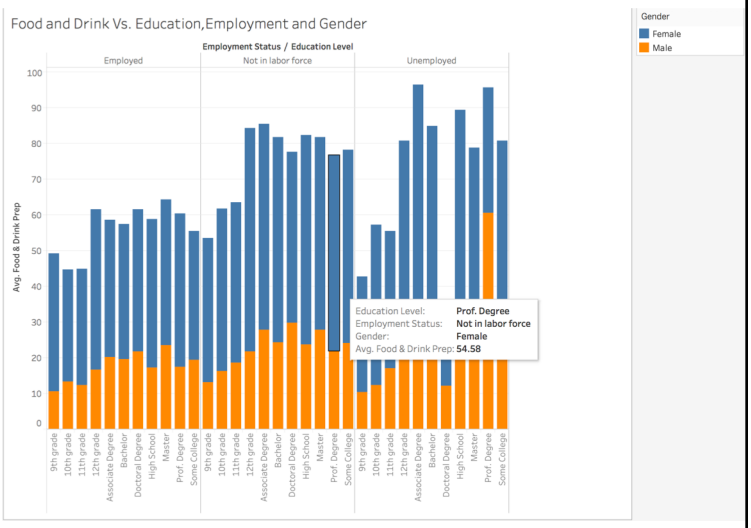
Change in sleeping based on work status

- Unemployed > Not in labor > Employed
- Employed people in the age range 30-60 have much lower average sleeping time than unemployed/not in labor



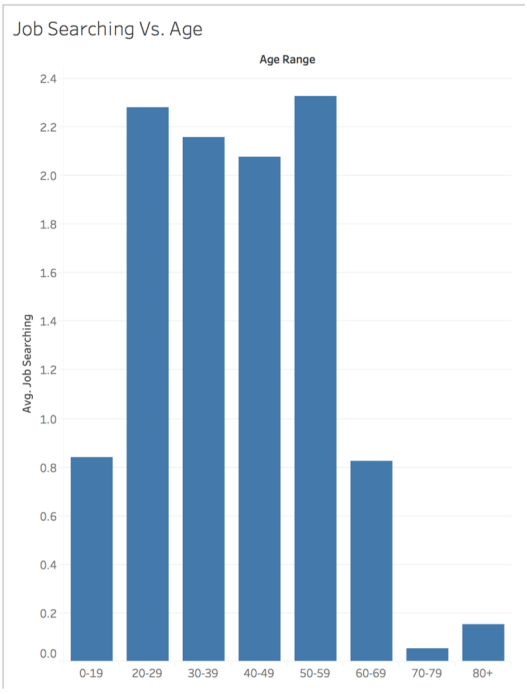
Change in food and drink based on education

- 12th grade students spend the most time on food and drink prep-38.79
- Male spend lot less time (20 mins) compared to female (45)
- Not in labor (44) > Unemployed (39) > Employed (29)



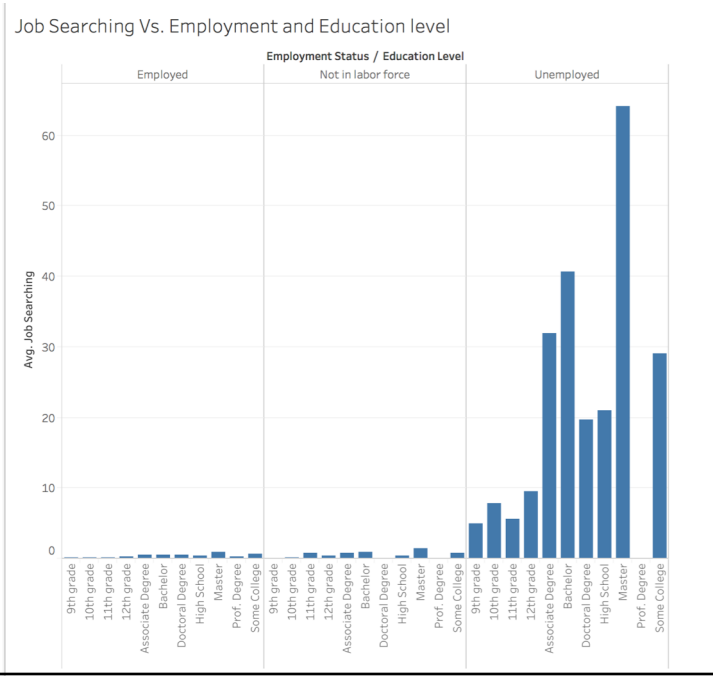
Change in job searching based on Age

- People in ages 0-19 spends on average 0.84 minutes in a day on job searching
- Time spend on job searching decreases as a person becomes 60 years older
- The max a person spends on job searching is on average 2.3 minutes a day



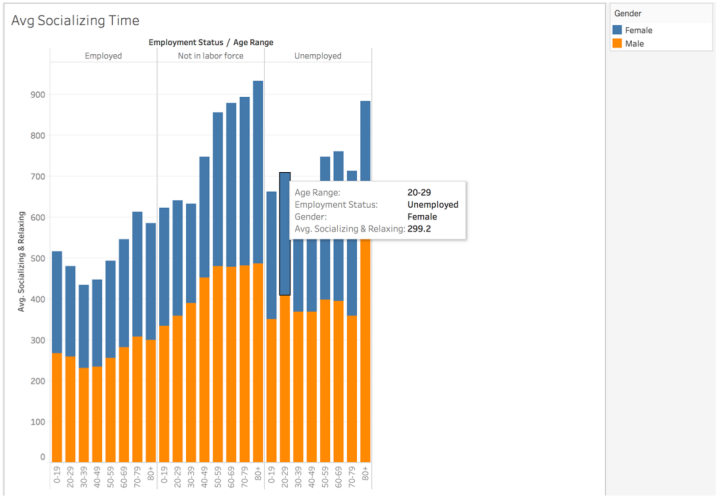
Change in job searching based on Employment and Education Level

- People with prof. degree spends the least time on job searching-0.18 minutes/day
- The next group is people having a doctorate degree-0.767 minutes/day
- Not a single person with prof. degree is unemployed.
- Out of people of all groups who are unemployed, Masters students spends the most on job searching-64 minutes
- Bachelors are spending more time than doctorate on job search.
- Job searching increased as time passed, except during 2007, where there was drop in time spent on job search



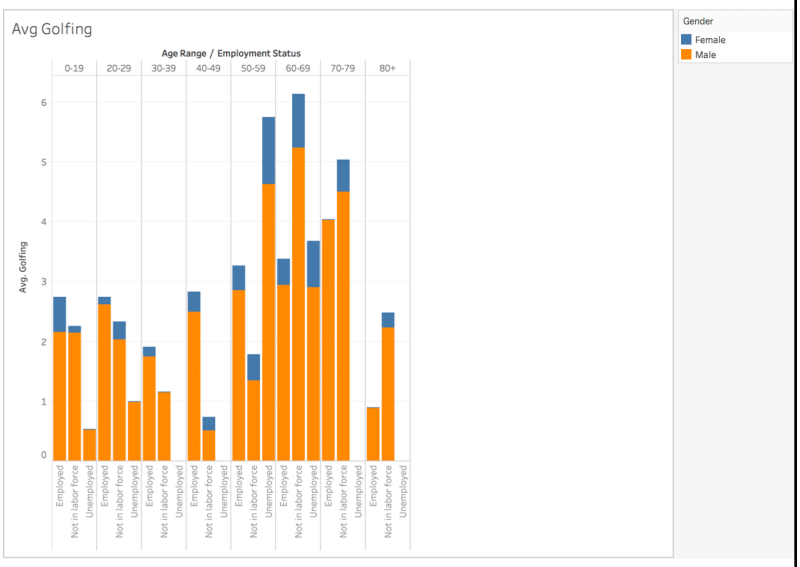
Change in socializing time

- On average, males spend more time on socializing compared to females
- 20-29 females who are unemployed spends more time on socializing than the one who are not in labor force



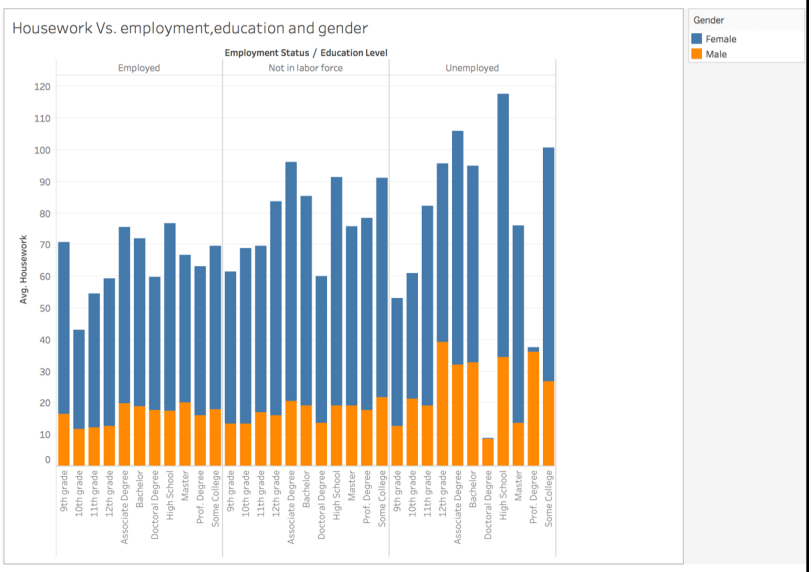
Change in golfing time

- People do a lot of golfing till their 20's, then it decrease till their 40's and then again time spent on golfing increases as age goes till 79
- After 79, it again decreases exponentially



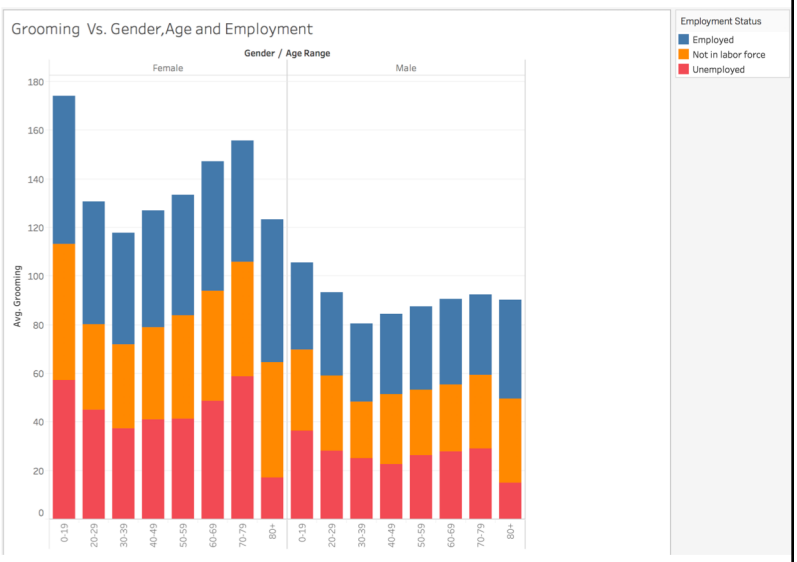
Change in housework time

- Male spends 18 minutes on housework whereas female spends 58 minutes on household.
- Not in labor force>unemployed>employed
- Doctoral spends the least time
- High School kids spend the most time



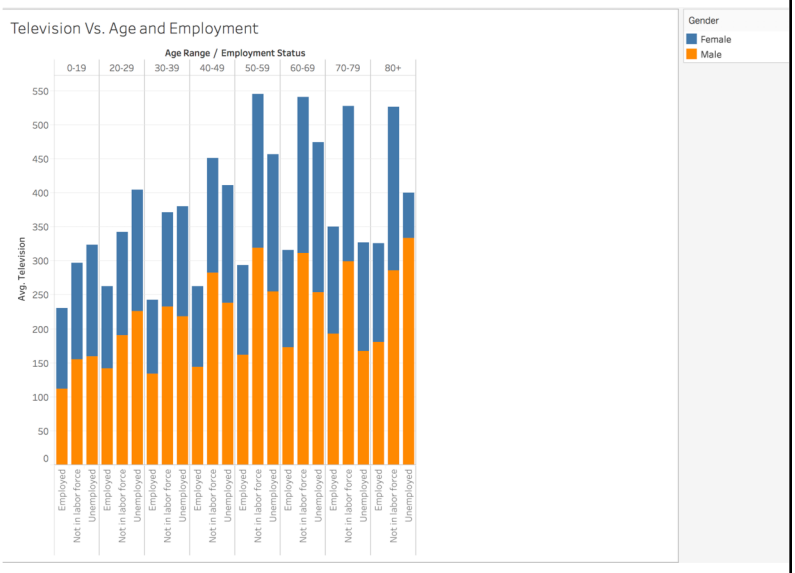
Change in grooming time

- On average, males spend less time on grooming compared to females
- People of Age 0-19 spend the most time on grooming no matter what their gender is



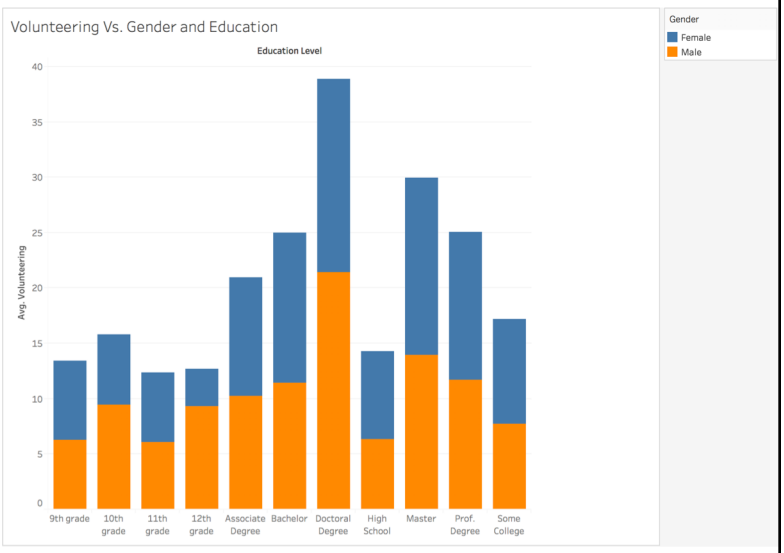
Change in television time

- On average, females spend less time on television compared to males
- Till Age of 40, people who are unemployed spends more time on watching television when compared to women who are not in labor force



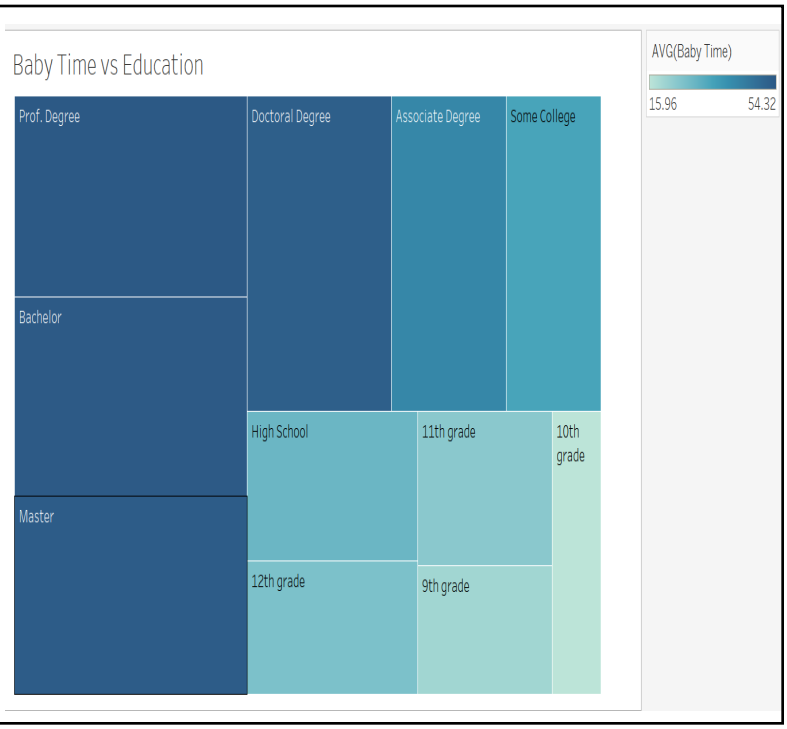
Change in volunteering time

- On average, females spend less time on volunteering compared to males
- People with doctorate degree spends the most time on volunteering



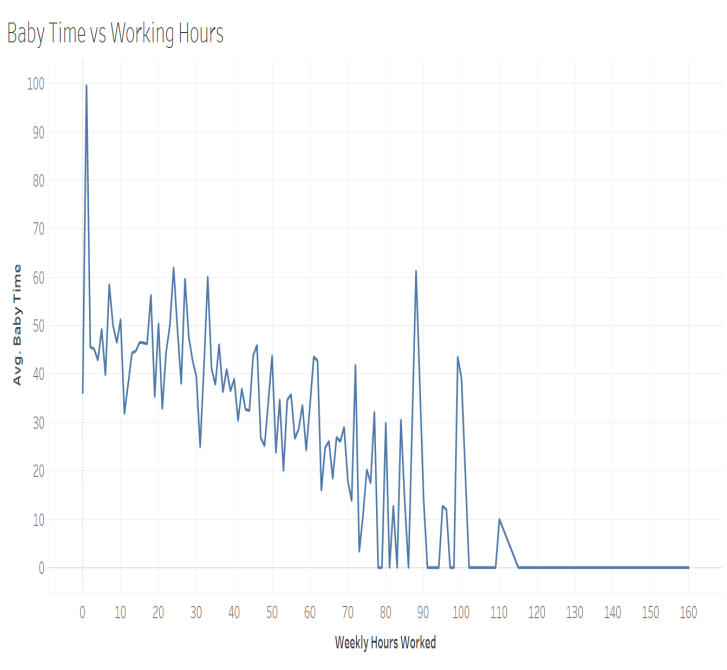
Change of baby care time based on Education.

- By looking at the treemap graph, I can see that people having Professional Degrees tend to spend the most amount of time with babies, on an average 54.32 mins per day, followed by people having Bachelors and Masters degree.
- The 10th grade students spend the least amount of time with babies.



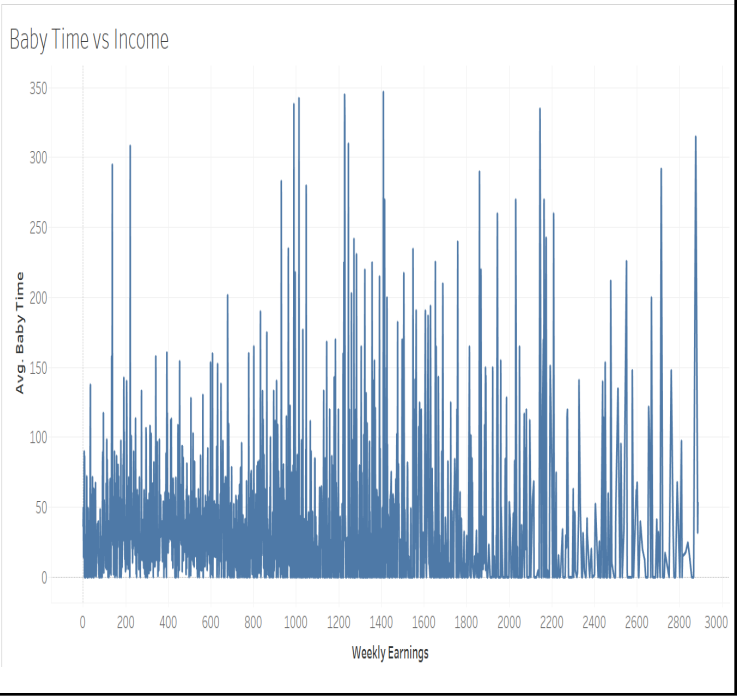
Change of baby care time based on Working Hours

- I can see that the time spent with babies is decreasing as the hours worked per week are increasing. Except for a few sudden peeks, the decrease is constant.
- This makes sense, as intuitively, people spending more hours working will have less time to spend with their babies.



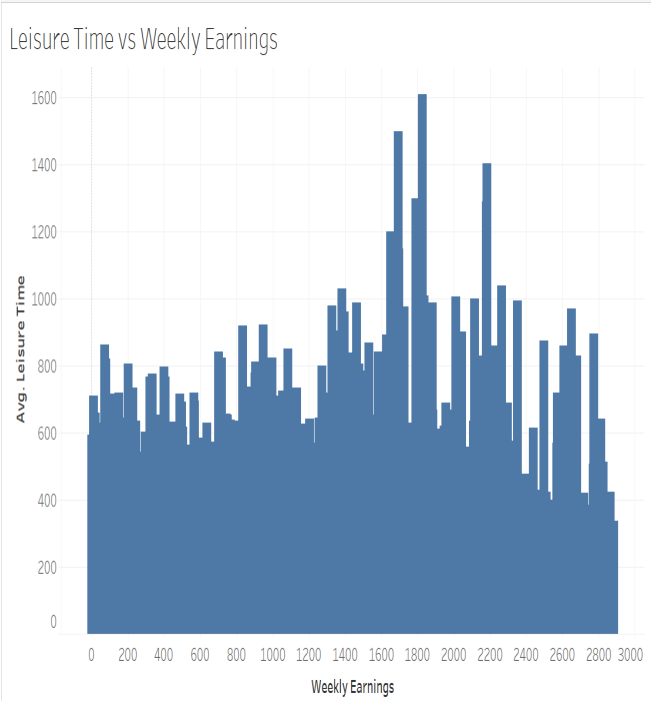
Change of baby care time based on income

- When I try to find the relation between income and time spent on baby care, I are not able to find any distinct pattern. **The values are all spread out**, but there are a few peaks in the middle region (1000-1600). This might mean that the people at both extremes of income level spend relatively less time with their babies. **The middle region shows the balanced people having substantial income and also spend enough time with their babies.**



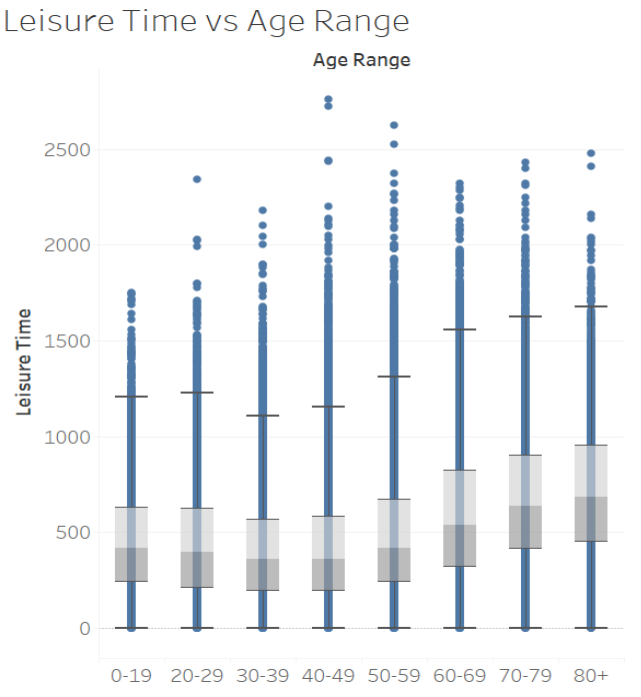
Change in leisure time based on Income

- I can see a sharp peak in the middle region, which tells us that **people having Weekly earnings between 1600 to 2200 tend to spend more time relaxing**.
- I have included the following in leisure time: Golfing, Playing with Children, Running, Shopping, Socializing & Relaxing, Television.
- Looking at the pattern, I believe that the people earning between 1600-2200 have enough money to spend on these activities, as Ill as enough time to indulge in leisure activities.
- **At one extreme I have people who probably don't earn enough to spend in the leisure activities, and at the other extreme I have very rich people who probably don't have enough time**, since they spend most of their time earning money.



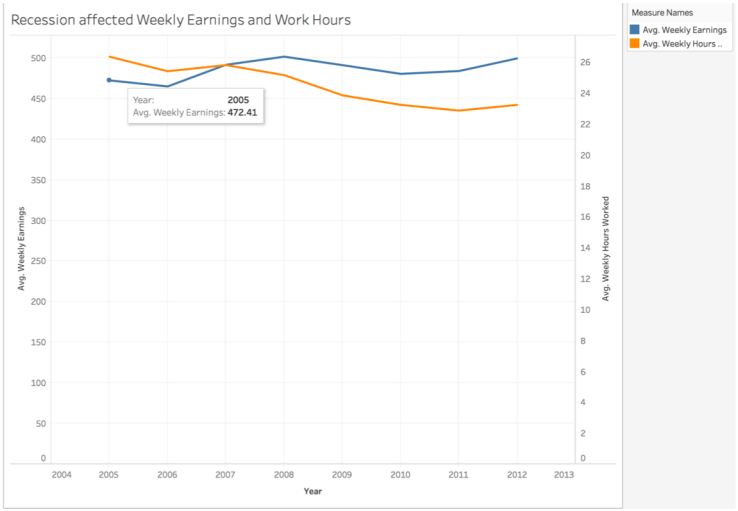
Change in leisure time based on Generations

- On analyzing the median, maximum and interquartile ranges of all the box plots, I see a pattern that **people above the age of 50 tend to spend more time in leisure activities**. The value decreases for people in their 30s and 40s, and then again increases for youngsters in their teens and 20s.
- This, again, is very intuitive. There is more time to spend in leisure for younger people, since they only have to worry about their studies; you build your career and work hard to earn money during your 30s and 40s, thus lower time left to spend in leisure activities. After reaching their 50s, people start retiring and thus have the most time to spend in leisure activities.

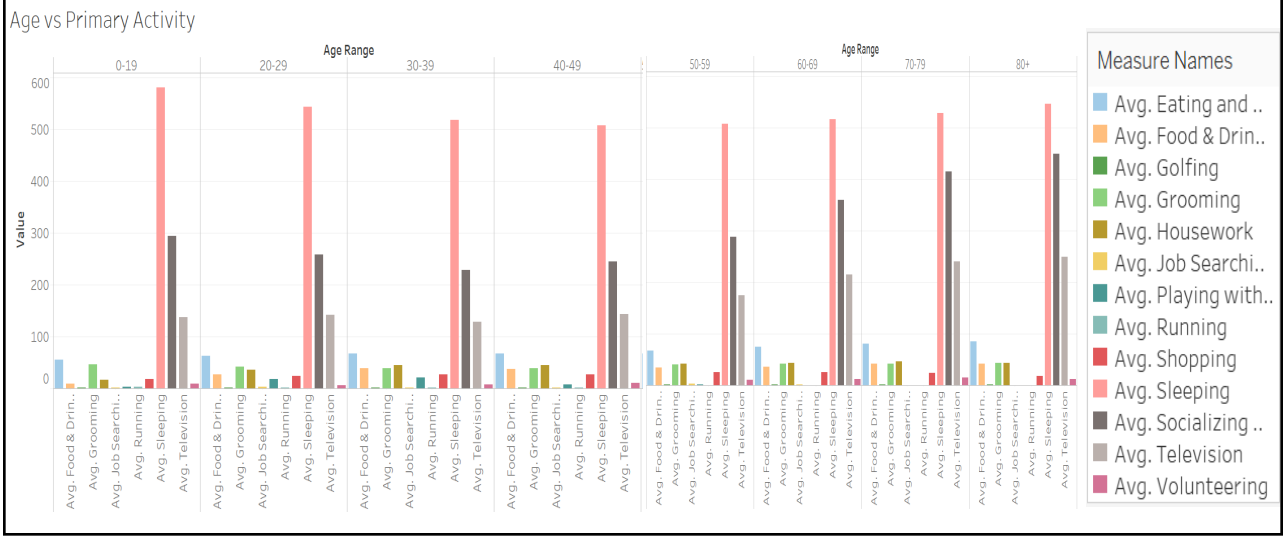


Recession affected Weekly Earnings and Work Hours

- From 2008 to 2009, the avg. Weekly hours worked and avg. weekly earnings decreased and the pattern remained the same for avg. Weekly hours worked, but avg. Weekly earnings started increasing after 2010.



Individual's Primary Activity based on Age



- On observing the bars for all age ranges, I can see that for all ranges, the primary activity is sleeping, followed by socializing and relaxing, followed by watching television, followed by eating and drinking.

- Now, let us see on which activity was the most time spent without considering the above mentioned activities.

- Age Range 0-19: Grooming
- Age Range 20-29: Grooming
- Age Range 30-39: Housework
- Age Range 40-49: Housework
- Age Range 50-59: Housework
- Age Range 60-69: Housework
- Age Range 70-79: Housework
- Age Range 80+: Housework and Grooming have almost the same height bars.

- Thus, I can see that no matter what the age of the person, their main primary activities remain the same. Sleeping, socializing, eating, drinking, watching television are the activities that everybody indulges in.

- Even the difference between grooming and housework is not a lot, thus these are also the primary activity that every individual indulges in.