My research question is if smoking, obesity, overworking or alcohol consumption are related to Victoria’s cancer incidence. It is an important question worth investigating because citizen health is the base for productivity. Finding a method to reduce cancer incidence is beneficial for Victoria.

Three datasets were used in this data wrangling and analysis. Adults Health Risk Factor Estimates for 2014-2015 in Local government areas in Victoria, which I will refer to as LGAs, Occupation by Hours Worked by Sex-Census for 2016, and LGA profile data 2015 for VIC. They were all found on Aurin website. These datasets were chosen because they all collect data for the LGAs in Victoria and the time periods are close to each other and are fairly recent. The Hours Worked by Sex-Census dataset was chosen for the possible difference of working hours for different genders.

Only relevant columns were chosen and the last row was deleted from two datasets in the data preprocessing as it was not contained in the ‘profile’ dataset and contained noise data. After data cleaning, the columns of the data frames are attributes for this investigation, and the rows are all the LGAs in Victoria.

翻页The first two plots shown are the hours worked distribution for male and female employees separately. It is noticeable that the percentage of males in Victoria that work for over 49 hours per week for this working time is overly high, which motivates the investigation between overworking for males and cancer incidence, as a concern for male health. The hours worked distribution for female employees in Victoria however, is moderate and reasonable.

翻页Male and female data were analyzed separately. Firstly, let’s look at the data related to the males. To distinguish each relation between pairs of attributes, these four scatter plots were made. Even though a large percentage of male employees are overworking, it shows that it does not have an obvious relation with cancer incidence. The strongest positive relation among these datasets is between smoking and obesity. There are also moderate positive relations between obesity or smoking and cancer.

翻页To be more precise about these correlations, this heatmap for correlation coefficients for male data was created. The row pointed by the red arrow shows that smoking and obesity both have around 0.5 correlation coefficients with cancer incidence, implying a noticeable positive relation. Working for more than 49 hours a week also has a small but certain positive relation with cancer incidence for males in Victoria.

翻页Now let’s look at females related data instead.

These four scatter plots corresponding to female shows that the relations have similar behaviors as the male data, except for the overworking attribute is replaced here by females who do not accomplish enough physical exercises since overworking females take very small percentage. And this attribute also does not have an obvious relation with cancer incidence, as shown in the fourth plot.

翻页 Here is the heatmap for correlation coefficients for female data. The row pointed by the red arrow implies the correlation coefficients between all the attributes and cancer incidents. The results are consistent with other plots, and are very similar to male data results.

翻页 The cancer incidence for population in Victoria is obtained by combining male and female columns. The correlation coefficients between risky alcohol consumption, overworking and cancer incidence are both around 0.5, implying positive relations. 翻页 And the scatter plot shows the same results. This could be because when employees work for too many hours a week, they use alcohol as a release for stress and tiredness. Therefore even though overworking does not have an obvious relation with cancer incidence, it could still have an indirect relation.

In conclusion, the factors positively relate to Victoria’s cancer incidence are smoking, obesity and alcohol consumption. Since smoking and obesity has a strong relation between each other, the government should try to control at least one of these two factors in order to reduce cancer incidence, such as raising the pricing for cigarettes. Reducing male employee working hours could also indirectly reduce cancer incidence.