The relationship between gender and seasonal allergies: influence by socioeconomic factors

- Konstantinos Paganopoulos |CID: 01769789 |July 2020
Healthcare & Medical Analytics, MSc Business Analytics, Imperial College London, UK

Abstract

The prevalence as well as the severity of seasonal allergy or hay fever or allergic rhinitis symptoms in the population changes by living conditions and other factors. In this report, we examine the relationship between seasonal allergies and gender, and how this relationship is affected by several socioeconomic factors. The objective is to examine the true effect of gender in the probability of suffering from seasonal allergies while holding for other socioeconomic variables. Apart from academic purposes, the report is also conducted to examine possible reasons why the writer suffers from allergic rhinitis.

Introduction

The relationship between seasonal allergies and gender has been examined by several studies and association has been proved to exist (Muhammad Khan et al., 2013). For our purposes we focus on the correlation between female gender and hay fever. As seen in previous studies, (Joanne K. Fagan et al., 2001), female population showed more intense symptoms of seasonal allergies. Moreover, as B. Kjellman and P. M. Gustafsson state (2000:p.464), "while asthma continuous to improve after adolescence in the males, it seems to worsen among the females". The above-mentioned relationship between female gender and seasonal allergies is explained in this study too.

Apart from examining how gender affects the answer of our respondents, it is examined how other socioeconomic factors influence that relationship. More specifically, it is assessed how the relationship between gender and seasonal allergies is affected by respondent's stress, if he or she has ever been arrested or raped, his or her nutrition, total asset valuation and ethnicity.

Descriptive Statistics

In this report, a population sample of 4557 people, responding if they had active seasonal allergies within the last four weeks or not, is used. As explained above, for our analysis 1 health measure (allergies) as well as 11 variables were used in total. Our dimensions were gender, lifestyle (stress and nutrition), social abuse (arrested or not and raped or not), income (assets) and ethnicity. Note that all of our variables were qualitative, and they can only change from one category or grouping transformation to another (e.g. healthy nutrition can change to unhealthy).

In terms of our exploratory data analysis, static graphs were used, and because our data were categorical, mainly segmented bar charts were chosen to visualize multiple variables together. In all of our segmented bar charts proportions were depicted, so as to easily compare our variables. Furthermore, mosaic plots were drawn to give statistical highlighting for the variances. Depicting our qualitative variables via static mosaic plots enabled us recognize relationships between different variables, as boxes across categories with different areas showed dependence.

In order to measure the prevalence of seasonal allergies by gender, controlling for variables that give males and females equal probabilities is vital. Apart from the relationship between allergies and gender that was explained above and was also observed in our explanatory data analysis (*Figure 1*), the reason that the variables stress, nutrition, arrested, rape, assets and ethnicity were chosen, was that all those variables affect the relationship that we investigate. Some of them affect the prevalence of seasonal allergies not only indirectly but also directly, such as stress (Hans Oh et al., 2018), while others differentiate by gender and thus affect our target variable (allergies) only indirectly, such as if someone has ever been arrested or not (G. Schwartz et al., 2010). As for nutrition and total assets, it is imperative to control for them, as females may have a healthier nutrition than males (*Figure 3*) and females may own less assets due to gender discrimination (*Figure 4*).

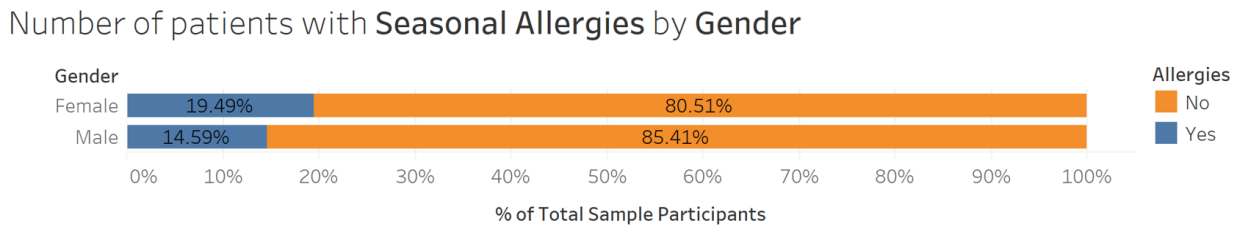


Figure 1: Seasonal allergies by gender, segmented bar chart.

More specifically, from our analysis we can see that many people suffering from allergies suffer from stress too (*Figure 5*). However, in terms of how stressful respondents suffering from seasonal allergies are split by gender (*Figure 6*), when comparing females who do not have stress (18.84%) with those who have (35.25%), we see that the latter ones are almost double. Therefore, it is important to control for stress, and capture the true effect of gender on allergies.

Similarly, as seen in (*Figure 8*), from females who have allergies symptoms, those who have been raped are more (23.35%) than those who have not (18.49%). As for the arrested variable, from (*Figure 7)* we see that males who have been arrested and experienced seasonal allergies symptoms at least once in the last month, are slightly less than those who have never been arrested, 13.94% and 15.05% respectively. As a result, arrested or not may have a minor significance in the relationship between gender and seasonal allergies. Hence, in order to gauge the true effect of gender on the dependent variable, we control both for rape (Lianne M. Tomfohr-Madsen et al., 2016) and arrested variable.

In terms of nutrition effect on the examined relationship, in (*Figure 9)* even though in "healthy" and "unhealthy" category there are no differences among the two genders, we see that for "normal" nutrition, females experience seasonal allergies with a slightly higher percentage (20.58%), so we have to examine any possible significance. Regarding assets variable, again, whereas in males no association between the variable and allergies is observed, females that belong to the "poor" category of assets, are slightly less than those that belong to the other categories (*Figure 10*). Thus, we control for nutrition and assets.

Lastly, regarding ethnicity and its net effect on the relationship between gender and seasonal allergies (*Figure 11*), we can see that Indian males are much more prone to hay fever (28.56%) than Asian (7.85%), which makes controlling for ethnicity necessary.

Regressions

In order to examine how gender and prevalence of seasonal allergies are associated with multiple socioeconomic variables we construct multiple regression models. Since our dependent variable is a factor, the prediction would be between 0 and 1, so we use logistic regression models. The final model, containing the marginal effects (coefficients) of all variables and how these affect the prevalence of seasonal allergies is the following:

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.86671    0.27280  -6.843 7.76e-12 ***
gender_Male         -0.27125    0.08865  -3.060  0.00221 **
stress_Yes           0.56687    0.19275   2.941  0.00327 **
arrested_Yes        -0.17747    0.09602  -1.848  0.06458 .
rape_Yes             0.30564    0.11381   2.686  0.00724 **
nutrition_normal     0.12627    0.10093   1.251  0.21091
nutrition_unhealthy  0.06957    0.11026   0.631  0.52806
assets_poor         -0.11955    0.09928  -1.204  0.22852
assets_wealthy       0.01774    0.09459   0.188  0.85127
ethnicity_black      0.16773    0.26980   0.622  0.53415
ethnicity_indian     0.92787    0.46982   1.975  0.04827 *
ethnicity_white      0.40188    0.25781   1.559  0.11904
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 1: Logistic regression analysis, significance of coefficients.

From (*Table 1*), we can clearly see that the gender of someone, the fact that she/he has been raped or not, as well as if the respondent has stress, are the three most significant factors that determine the frequency of his/her seasonal allergies occurrence. The last two were expected, since women are more likely to be abused and raped due to less strength, and because the proportion of women feeling stressed is higher than that of men (Carolyn M. Mazure et al., 2014).

Moreover, Indian ethnicity seems to play a role in determining the frequency of seasonal allergies, whereas the other ethnicity categories not. As for the variable arrested or not,

we could say that is quite significant, probably because it is associated with gender as men are more likely to be arrested due to increased violence.

Other socioeconomic factors, such as total assets that somebody owns, black or white ethnicity, or nutrition, don't affect the frequency of seasonal allergies between genders.

Limitations

In that study female respondents were slightly more than male ones, 2468 and 2089 respectively (*Figure 2*), which raises the question of potential confounding. However, this issue was addressed by using the proportions of total population in the segmented bar charts. Moreover, even though in the study only data from Add Health Wave IV were used, in which respondents were 24 to 32 years old, age of the respondents seems that has not affected the results, as the sample consists of neither a very young nor of a very old population, which makes it appropriate for this study.

Weights

As for weights, those who participated in the study might differ from those that dropped out and thus introduce bias in the results (e.g. those claiming that they have been arrested at least once might be more, since many left the study in fear of social discrimination, or vice versa many that have never been arrested thought that this type of study will not add any value to them). As a result, attrition problem might exist, which is inevitable to some extent, and that is the reason why weights were not used in the descriptive statistics.

Recommendations

This study has shown that gender affects the prevalence of seasonal allergies. While stress and if someone has been raped or not affect that relationship, assets and nutrition have a negligible effect on the association between gender and seasonal allergies. Further research is needed to examine the validity of the above results in a wider scale.

References

B. Kjellman and P. M. Gustafsson (2000) Asthma from childhood to adulthood: asthma severity, allergies, sensitization, living conditions, gender influence and social consequences. *Respiratory Medicine.* 94, 454-465. Available from: doi:10.1053/rmed.1999.0764

Carolyn M. Mazure, Andrea H. Weinberger, Brian Pittman, Igor Sibon, Joel Swendsen (2014) Gender and Stress in Predicting Depressive Symptoms Following Stroke. *Cerebrovasc Dis.* 38, 240–246. Available from: doi:10.1159/000365838

Dennis Shusterman, Mary Alice Murphy, John Balmes (2003) Differences in nasal irritant sensitivity by age, gender, and allergic rhinitis status. *Int Arch Occup Environ Health.* 76, 577–583. Available from: doi:10.1007/s00420-003-0459-0

Fredrik Barrenäs, Bengt Andersson, Lars Olaf Cardell, Michael Langston, Reza Mobini, Andy Perkins, Juhani Soini, Arne Stahl, Mikael Benson (2008) Gender differences in inflammatory proteins and pathways in seasonal allergic rhinitis. *The International Cytokine & Interferon Society*. 42 (3), 325-329. Available from: doi:10.1016/j.cyto.2008.03.004

G. Schwartz, R. S. Gupta, E. Springston, X. Zhang, L. C. Grammer (2010) The Association between Crime and Adult Asthma Severity. *The Journal of Allergy and Clinical Immunology.* 125 (2), S1, AB32. Available from: doi:10.1016/j.jaci.2009.12.157

Hans Oh, Ai Koyanagi, Jordan E. DeVylder, Andrew Stickley (2018) Seasonal Allergies and Psychiatric Disorders in the United States. *International Journal of Environmental Research and Public Health.* 15, 1965. Available from: doi:10.3390/ijerph15091965

Joanne K. Fagan, Peter A. Scheff, Dan Hryhorczuk, Viswanathan Ramakrishnan, Mary Ross, Victoria Persky (2001) Prevalence of asthma and other allergic diseases in an adolescent population: association with gender and race. *Annals of Allergy, Asthma & Immunology.* 86 (2), 177-184. Available from: doi:10.1016/S1081-1206(10)62688-9

Lianne M. Tomfohr-Madsen, Hamideh Bayrampour, Suzanne Tough (2016) Maternal History of Childhood Abuse and Risk of Asthma and Allergy in 2-Year-Old Children. *Psychosomatic Medicine.* 78, 1031-1042. Available from: doi:1097/PSY.0000000000000419

M. R. Sears, B. Burrows, E. M. Flannery, G. P. Herbison, M. D. Holdaway (1993) Atopy in childhood. I. Gender and allergen related risks for development of hay fever and asthma. *Clinical and Experimental Allergy.* 23 (11), 941-948. Available from: doi:10.1111/j.1365-2222.1993.tb00279.x

Muhammad Khan, Muhammad Alamgir Khan, Faizania Shabbir, Tausif Ahmed Rajput (2013) Association of Allergic Rhinitis with gender and asthma. *Journal of Ayub Medical College Abbottabad-Pakistan*. 25 (1-2), 120-122. Available from: https://jamc.ayubmed.edu.pk/index.php/jamc/article/view/1869/671 [Accessed 1st June 2013].

Appendix

First, we show how data were manipulated and cleaned, as well as some descriptive statistics with Python.

Healthcare & Medical Analytics

Individual Assignment - Konstantinos Paganopoulos

"The relationship between gender and seasonal allergies; influence by socioeconomic factors"

First we load the necessary libraries.

In [1]:
```
import string
import pyreadr
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.mosaicplot import mosaic
```

Then we load and visually inspect our data set.

In [2]:
```
data_orig = pyreadr.read_r('21600-0022-Data.rda')
```

In [3]:
```
data_orig = data_orig['da21600.0022']
data_orig
```

Out[3]:

	AID	IMONTH4	IDAY4	IYEAR4	BIO_SEX4	VERSION4	BREAK_Q	PRYEAR4	PRETEST4	PRISON4	...	H4EO5C	H4EO5D	H4EO5E	H4EO5F	H4EC
0	57101310	(05) (5) May	6.0	2008.0	(2) (2) Female	V5.4	NO	2001.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele

1	57103869	(05) (5) May	22.0	2008.0	(1) (1) Male	V5.4	NO	2002.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele
2	57109625	(11) (11) November	2.0	2008.0	(1) (1) Male	V5.5	NO	2002.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	NaN	NaN	NaN	NaN	I
3	57111071	(06) (6) June	29.0	2008.0	(1) (1) Male	V5.4	NO	2001.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele
4	57113943	(11) (11) November	11.0	2008.0	(1) (1) Male	V5.5	NO	2002.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
5109	99719930	(06) (6) June	7.0	2008.0	(2) (2) Female	V5.4	NO	2001.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(1) (1) Selected	(0 sele
5110	99719939	(02) (2) February	13.0	2008.0	(1) (1) Male	V5.1	NO	2001.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele
5111	99719970	(03) (3) March	22.0	2008.0	(1) (1) Male	V5.2	NO	1996.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(1 Sele
5112	99719976	(04) (4) April	1.0	2008.0	(2) (2) Female	V5.2	NO	2001.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele
5113	99719978	(04) (4) April	16.0	2008.0	(1) (1) Male	V5.3	NO	1995.0	(0) (0) Not a pretest interview	(0) (0) Not a prison interview	...	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0) (0) Not selected	(0 sele

5114 rows × 920 columns

We then choose the variables that we are interested in.

In [4]:
```
data_new = data_orig[['H4ID9F','BIO_SEX4','H4ID5I','H4CJ1','H4SE32','H4GH8','H4EC7','H4IR4']]
```

After that we give our variables appropriate names.

In [5]:
```
data_new = data_new.rename(columns ={'H4ID9F':'allergies','BIO_SEX4':'gender','H4ID5I':'stress','H4CJ1':'arrested','H4SE32':'rape','H4GH8':'nutrition','H4EC7':'assets','H4IR4':'ethnicity'})
```

Then we clean our data by removing any special characters from the columns needed.

In [6]:
```
columns = ['allergies','gender','stress','arrested','rape','ethnicity']

for i in columns:
    for j in string.punctuation:
        data_new[i] = data_new[i].str.lstrip(' ')
        data_new[i] = data_new[i].str.replace('\d+','')
        data_new[i] = data_new[i].astype(str).str.replace(j,'')
```

Afterwards, we prepare our data. First, we pick the variables that we need to use in our model and drop "NaN" values.

In [7]:
```
data = data_new[['allergies','gender','stress','arrested','rape','nutrition','assets','ethnicity']]
data = data.dropna()
```

Let's now reduce the number of categories of some of our categorical variables. We reduce the dimension of the fast-food consumption variable from 8 to 3, and classify the nutrition as "healthy", "normal" and "unhealthy". We also reduce the dimension of total assets besides homes variable from 9 to 3, "poor", "normal" and "wealthy". Moreover, we rename our categories of our ethnicity variable so as to have shorter names.

In [8]:
```
for i in range(len(data['nutrition'])):
    if data['nutrition'].iloc[i] == 0:
        data['nutrition'].iloc[i] = 'healthy'
    elif data['nutrition'].iloc[i] == 1 or data['nutrition'].iloc[i] == 2:
        data['nutrition'].iloc[i] = 'normal'
    else:
        data['nutrition'].iloc[i] = 'unhealthy'

mapping = {'(1) (1) Less than $5,000': 'poor', '(2) (2) $5,000 to $9,999': 'poor', '(3) (3) $10,000 to $24,999': 'normal', '(4)
(4) $25,000 to $49,999': 'normal', '(5) (5) $50,000 to $99,999': 'wealthy','(6) (6) $100,000 to $249,999': 'wealthy','(7) (7) $2
50,000 to $499,999': 'wealthy','(8) (8) $500,000 to $999,999': 'wealthy','(9) (9) $1,000,000 or more': 'wealthy'}
data['assets'] = data.assets.map(mapping)

mapping2 = {'White': 'white', 'Black or African American': 'black', 'Asian or Pacific Islander': 'asian', 'American Indian or Al
aska Native': 'indian'}
data['ethnicity'] = data.ethnicity.map(mapping2)
```

```
C:\Users\fdt\Anaconda3\lib\site-packages\pandas\core\indexing.py:205: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy
  self._setitem_with_indexer(indexer, value)
```

Now, in order to deal with perfect multicolinearity issues, after we create dummy variables for each of the categories of our categorical variables, we drop one column.

In [9]:
```
# remove allergies from our list so as not to create a dummy for it
columns = list(data.columns)
columns.pop(0)

# deal with perfect multicolinearity
model = pd.get_dummies(data, columns = columns, drop_first = True)
```

We remove "NaN" dummy columns and we visually inspect our final data for our model.

In [10]:
```
model = model.drop(['arrested_nan'], axis = 1)
model = model.drop(['rape_nan'], axis = 1)
model
```

Out[10]:

	allergies	gender_Male	stress_Yes	arrested_Yes	rape_Yes	nutrition_normal	nutrition_unhealthy	assets_poor	assets_wealthy	ethnicity_black	ethnicity_i
0	No	0	0	0	0	0	1	0	0	1	
1	No	1	0	1	1	0	1	1	0	1	
2	No	1	0	1	0	1	0	0	0	0	
3	No	1	0	0	0	1	0	0	0	0	
4	No	1	0	1	0	1	0	1	0	1	
...	...	...	...	...	...	...	...	...	...	...	
5108	No	0	0	0	0	1	0	0	0	0	
5109	No	0	0	0	0	0	1	1	0	1	
5110	Yes	1	0	0	0	0	1	1	0	1	
5112	No	0	0	0	0	0	1	1	0	0	
5113	No	1	0	1	0	1	0	0	1	0	

4562 rows × 12 columns

We store the final data for our model in a csv file for future purposes.

In [11]:
```
model.to_csv('data_model.csv')
```

Lastly, we visually inspect our final data for our descriptive statistics and store them in a csv file for future analysis.

In [12]:
```
data.to_csv('data_eda.csv')
data
```

Out[12]:

	allergies	gender	stress	arrested	rape	nutrition	assets	ethnicity
0	No	Female	No	No	No	unhealthy	normal	black
1	No	Male	No	Yes	Yes	unhealthy	poor	black
2	No	Male	No	Yes	No	normal	normal	white
3	No	Male	No	No	No	normal	normal	white
4	No	Male	No	Yes	No	normal	poor	black
...	...	...	...	...	...	...	...	...
5108	No	Female	No	No	No	normal	normal	white
5109	No	Female	No	No	No	unhealthy	poor	black
5110	Yes	Male	No	No	No	unhealthy	poor	black
5112	No	Female	No	No	No	unhealthy	poor	white
5113	No	Male	No	Yes	No	normal	wealthy	white

4562 rows × 8 columns

Let's now draw some descriptive statistics.

We choose the Mosaic plot from statsmodels, since it gives us statistical highlighting for the variances. In other words, it is a graphical method for visualizing data from two or more qualitative variables. It is the multidimensional extension of spineplots, which graphically display the same information for only one variable. It gives an overview of the data and makes it possible to recognize relationships between different variables. For example, independence is shown when the boxes across categories all have the same areas.

We now draw mosaic plots for several variables.

In [15]:
```
data.gender.value_counts()
```

Out[15]:
```
Female    2468
Male      2089
Name: gender, dtype: int64
```

In [16]:
```
plt.rcParams['font.size'] = 16.0
pd.crosstab(data.allergies, data.gender)
ct = pd.crosstab(data.allergies, data.gender)
mosaic(ct.unstack(), gap = 0.006675, title = 'Allergies per Gender')
plt.show()
```

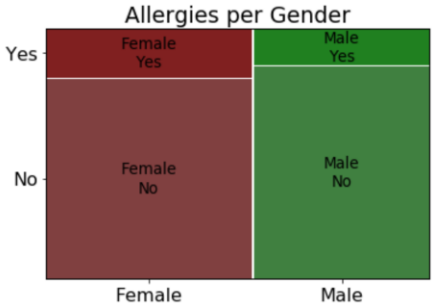


Figure 2: Allergies split by gender, mosaic plot.

In [13]:
```
plt.rcParams['font.size'] = 16.0
pd.crosstab(data.gender, data.nutrition)
ct = pd.crosstab(data.gender, data.nutrition)
mosaic(ct.unstack(), title = 'Nutrition type per Gender')
plt.show()
```

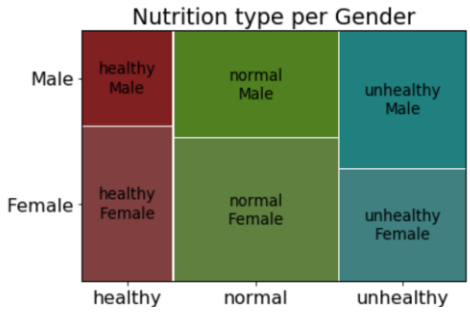
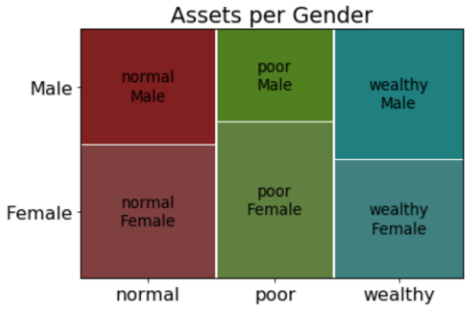


Figure 3: Nutrition type split by gender, mosaic plot.

Hence, since the area for unhealthy nutrition for men is larger than that of women and vice versa, we can conclude that men have on average an unhealthier nutrition in comparison with that of women.

We can see that female are more prone to seasonal allergies than men.

In [18]:
```
plt.rcParams['font.size'] = 16.0
pd.crosstab(data.gender, data.assets)
ct = pd.crosstab(data.gender, data.assets)
mosaic(ct.unstack(), gap = 0.006675, title = 'Assets per Gender')
plt.show()
```



Lastly, we can clearly see that females have on average a lower number of total assets (bank accounts, retirement plans and stocks) in comparison with men, who own more.

Figure 4: Assets owned split by gender, mosaic plot.

Then a presentation of some of the logistic regression models that were built is given.

Healthcare and Medical Analytics

Individual Assignment

Konstantinos Paganopoulos

First, we load all the libraries that we need.

After that, we load the data that we need for our model:

```
data_model = read.csv("data_model.csv")
head(data_model)
```

```
##   X allergies gender_Male stress_Yes arrested_Yes rape_Yes nutrition_normal
## 1 0        No           0          0            0        0                 0
## 2 1        No           1          0            1        1                 0
## 3 2        No           1          0            1        0                 1
## 4 3        No           1          0            0        0                 1
## 5 4        No           1          0            1        0                 1
## 6 5        No           1          0            1        0                 0
##   nutrition_unhealthy assets_poor assets_wealthy ethnicity_black
## 1                   1           0              0               1
## 2                   1           1              0               1
## 3                   0           0              0               0
## 4                   0           0              0               0
## 5                   0           1              0               1
## 6                   1           0              0               0
##   ethnicity_indian ethnicity_white
## 1                0               0
## 2                0               0
## 3                0               1
## 4                0               1
## 5                0               0
## 6                0               1
```

Let's now build several logistic regression models to examine how each factor affects the relationship between gender and seasonal allergies.

First, we start by examining the relationship between gender and seasonal allergies:

```
model <- glm(allergies ~ gender_Male, family = binomial(link = 'logit'), data_model)
summary(model)
```

```
##
## Call:
## glm(formula = allergies ~ gender_Male, family = binomial(link = "logit"),
##     data = data_model)
##

## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6583  -0.6583  -0.5593  -0.5593   1.9660
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.41902    0.05081 -27.926  < 2e-16 ***
## gender_Male -0.35722    0.08026  -4.451 8.55e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4186.2  on 4561  degrees of freedom
## Residual deviance: 4166.1  on 4560  degrees of freedom
## AIC: 4170.1
##
## Number of Fisher Scoring iterations: 4
```

Now we combine all of the above factors and check how all of them affect the relationship between gender and seasonal allergies:

```
model <- glm(allergies ~ gender_Male + stress_Yes + arrested_Yes + rape_Yes
             + nutrition_normal+ nutrition_unhealthy + assets_poor
             + assets_wealthy + ethnicity_black + ethnicity_indian
             + ethnicity_white, family = binomial(link = 'logit'), data_model)
summary(model)
```

```
##
## Call:
## glm(formula = allergies ~ gender_Male + stress_Yes + arrested_Yes +
##     rape_Yes + nutrition_normal + nutrition_unhealthy + assets_poor +

##     assets_wealthy + ethnicity_black + ethnicity_indian + ethnicity_white,
##     family = binomial(link = "logit"), data = data_model)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1788  -0.6468  -0.5796  -0.5125   2.1663
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.86671    0.27280  -6.843 7.76e-12 ***
## gender_Male         -0.27125    0.08865  -3.060  0.00221 **
## stress_Yes           0.56687    0.19275   2.941  0.00327 **
## arrested_Yes        -0.17747    0.09602  -1.848  0.06458 .
## rape_Yes             0.30564    0.11381   2.686  0.00724 **
## nutrition_normal     0.12627    0.10093   1.251  0.21091
## nutrition_unhealthy  0.06957    0.11026   0.631  0.52806
## assets_poor         -0.11955    0.09928  -1.204  0.22852
## assets_wealthy       0.01774    0.09459   0.188  0.85127
## ethnicity_black      0.16773    0.26980   0.622  0.53415
## ethnicity_indian     0.92787    0.46982   1.975  0.04827 *
## ethnicity_white      0.40188    0.25781   1.559  0.11904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4186.2  on 4561  degrees of freedom
## Residual deviance: 4130.1  on 4550  degrees of freedom
## AIC: 4154.1
##
## Number of Fisher Scoring iterations: 4
```

R Markdown Code, logistic regression models.

Finally, further descriptive statistics are provided.

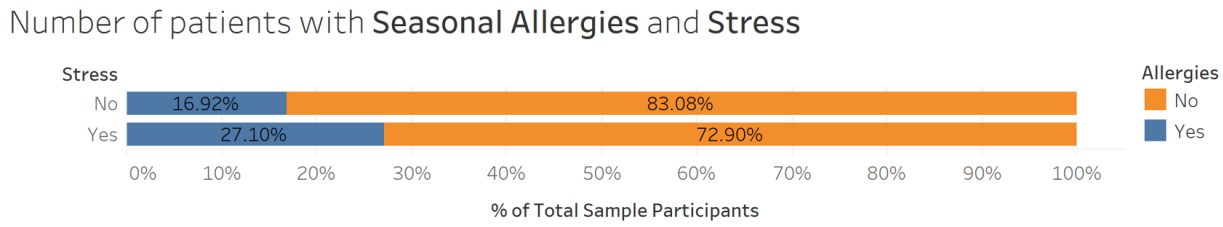


Figure 5: Seasonal allergies per stress: segmented bar chart.

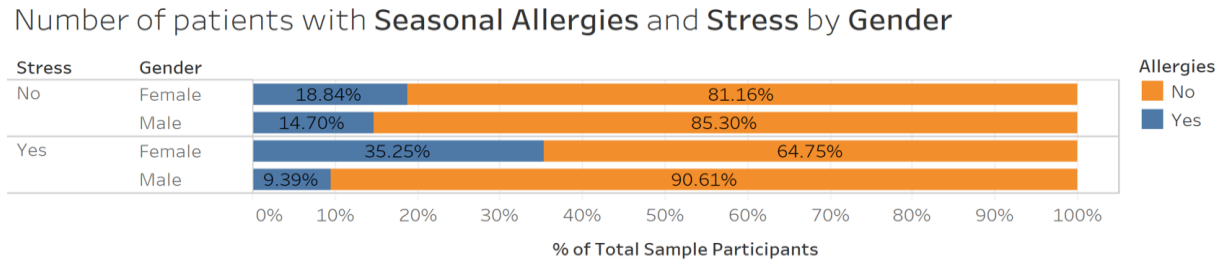


Figure 6: Seasonal allergies by stress split by gender: segmented bar chart.

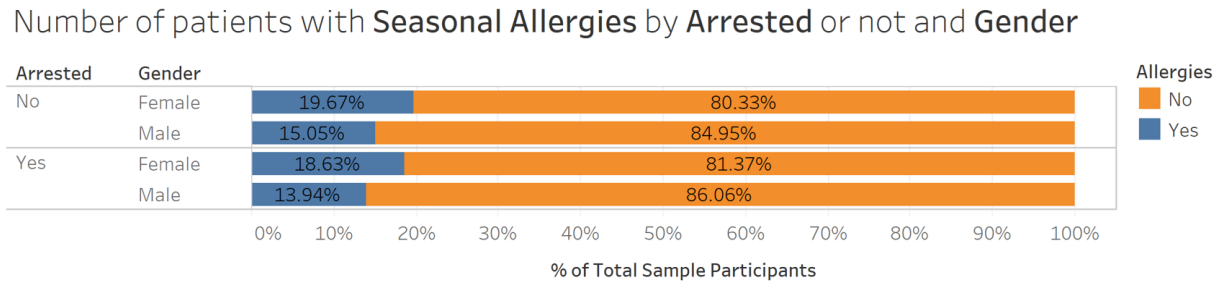


Figure 7: Seasonal allergies by arrested or not split by gender: segmented bar chart.

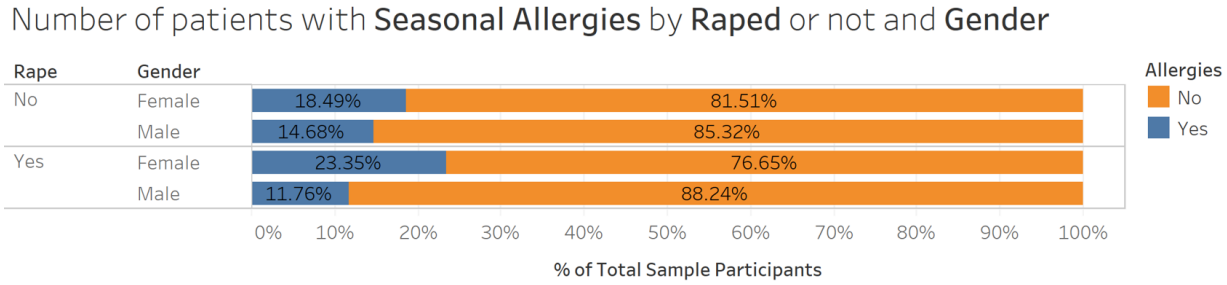


Figure 8: Seasonal allergies by raped or not split by gender: segmented bar chart.

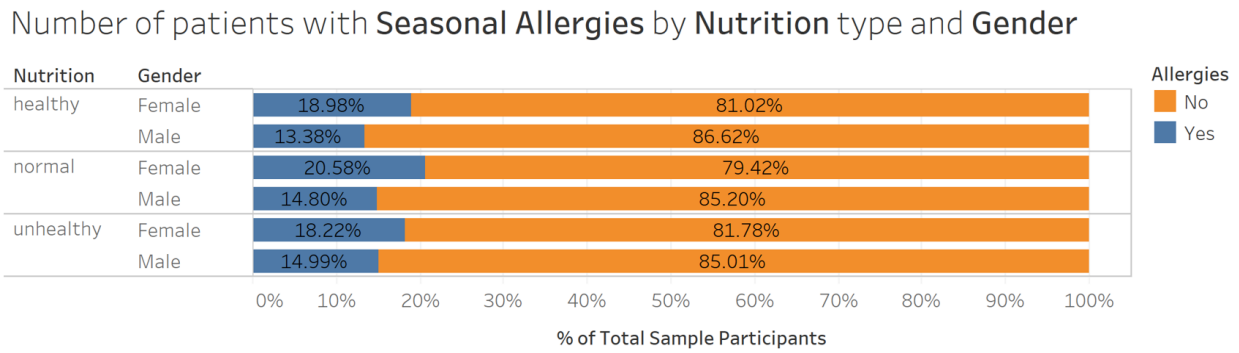


Figure 9: Seasonal allergies by nutrition type split by gender: segmented bar chart.

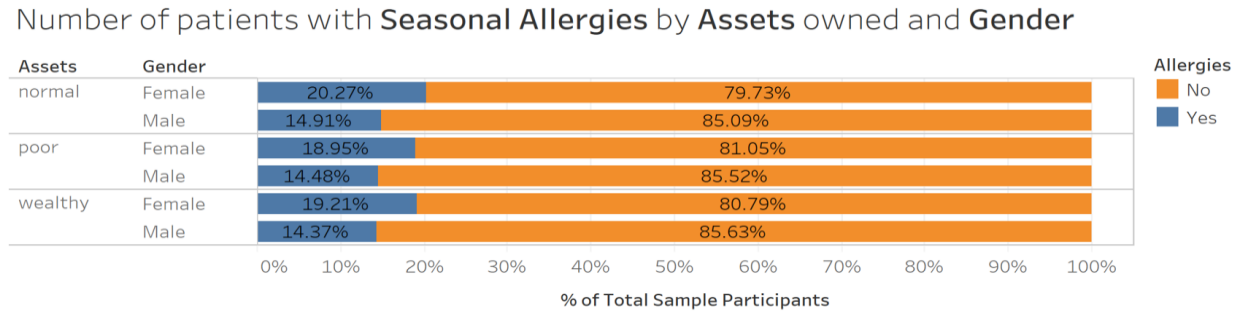


Figure 10: Seasonal allergies by total assets owned split by gender: segmented bar chart.

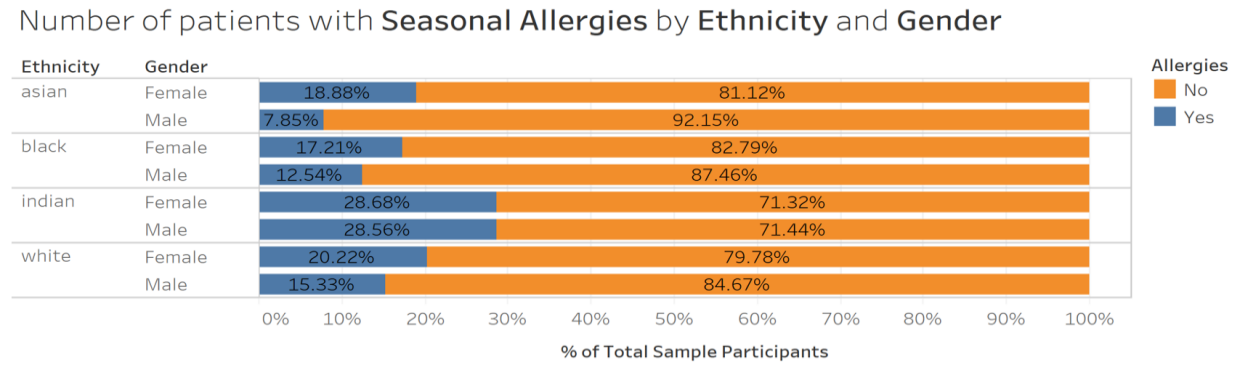


Figure 11: Seasonal allergies by ethnicity split by gender: segmented bar chart.