

Tobacco -Related Deaths in the City of Austin 2006-2018

Dorothy Hewitt-Sanchez

# Introduction

This analysis examines the tobacco-related deaths in the City of Austin between 2006 and 2018. The dataset includes the year of the death, person’s gender, age, race/ethnicity and whether tobacco contributed to the cause of death.

The dataset is available for viewing through Austin open data portal. The site location is <https://data.austintexas.gov/Health-and-Community-Services/Tobacco-Related-Deaths-in-the-City-of-Austin-2006-/fc6z-ynb8>

# The Dataset

The data set has 8998 entries, numbered 0 to 8997 that contain 6 columns. The columns are:

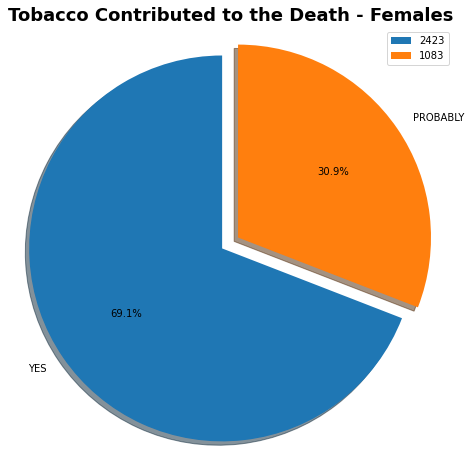
* Year – (Year of the death)
* Contributed – (Tobacco Contributed to the Death) the options are YES or PROBABLY
* Age – (Age of death)
* Sex – (MALE, FEMALE, UNKNOWN)
* Race – (Race/Ethnicity)
* Age\_Group – (Age Group (10-year)) include: [0-29], [30-39], [40-49],

[50-59], [60-69], [70-79], [80-89], [90+]

Some column names were shortened for better readability. These columns include: Contributed, Race, and Age\_Group.

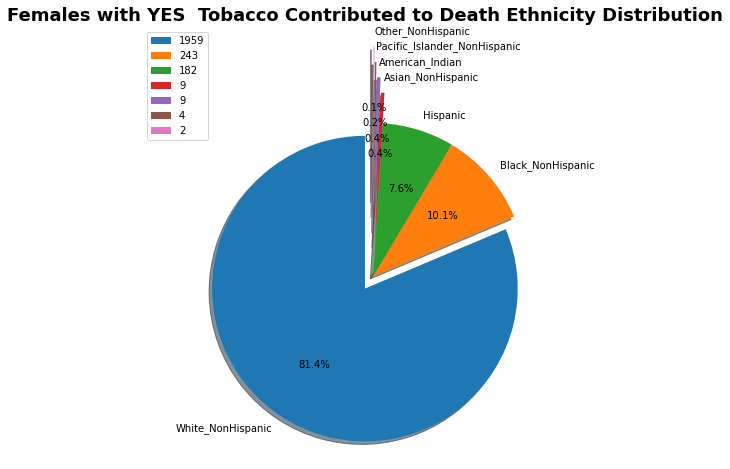


# FEMALE Analysis



There is a total of 3506 female observations. 2423 have the status of ‘YES’ for tobacco contributing to death. 1083 shows the status of ‘PROBABLY’. The percentage of the whole is shown as above. The statistical analysis of the Age of the female observations are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All Females** | **Contributed to Death - YES** | **Contributed to Death - PROBABLY** |
| **Female Observations** | 3506 | 2423 | 1083 |
| **Median Age** | 74 | 73 | 75 |
| **Minimum Age** | 2 | 2 | 2 |
| **Maximum Age** | 102 | 102 | 100 |
| **Skew** | -0.508098 | -0.448035 | -0.637626 |
| **Mode** | 82 | 71 | 80 |
| **Occurrences of the Mode** | 123 | 91 | 45 |

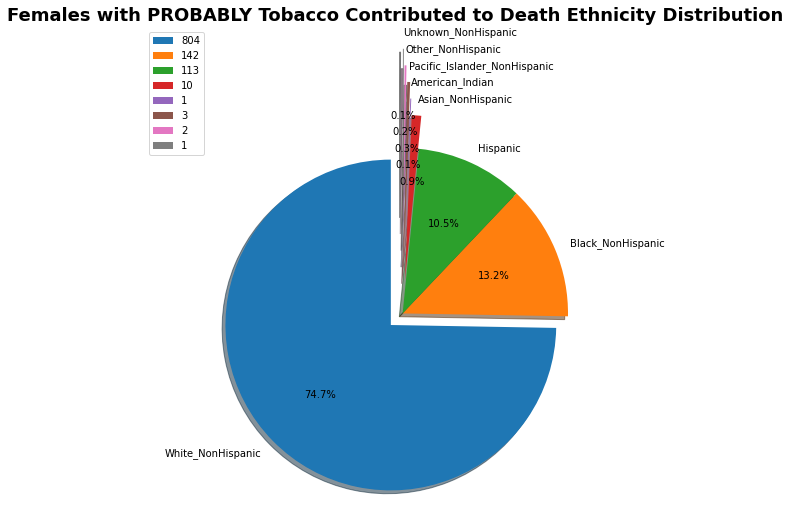


Race/Ethnicity data is missing for December YEAR 2018 due to system errors. There are 15 records that have null values for this date that are not included in this analysis.

The legend shows the number of observations that were in the data set for each ethnicity distribution.

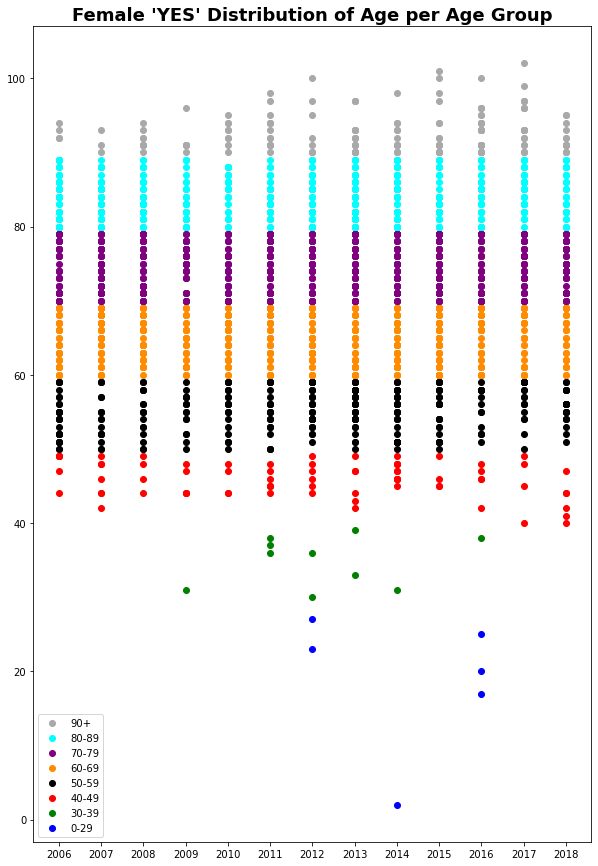
The Race/Ethnicity groups include:

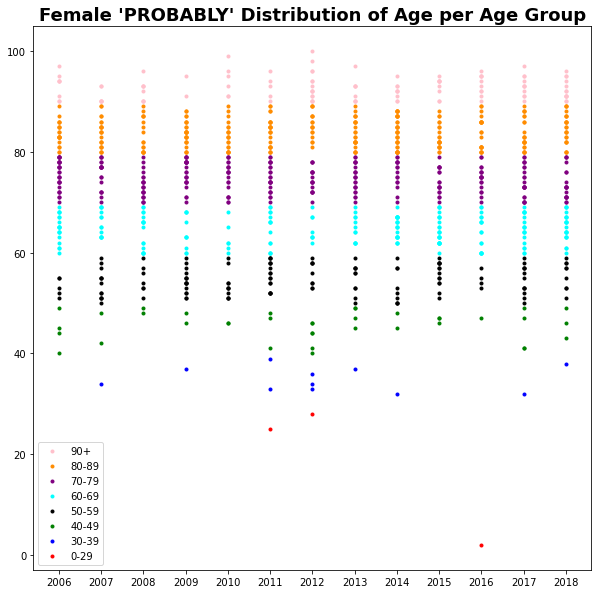
* American Indian
* Asian NonHispanic
* Black NonHispanic
* Hispanic
* Other NonHispanic
* Pacific Islander NonHispanic
* White NonHispanic



There are 7 records that have null values for this date that are not included in this analysis due to system errors from December 2018.

The following two charts show the distribution of Age in each Age Group for females:



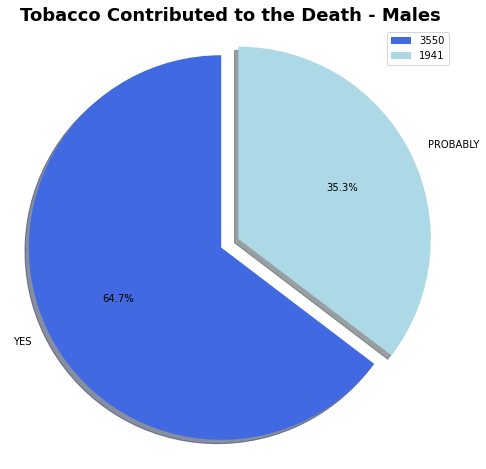


The total number of observations per year is presented in the following charts:



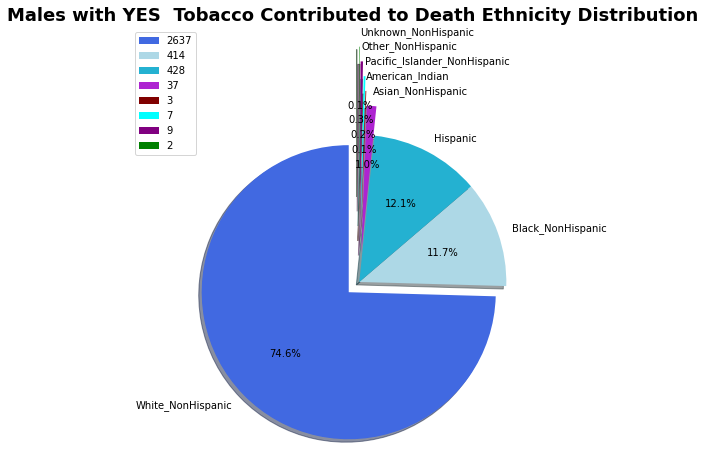


# MALE Analysis



There is a total of 5491 male observations. 3550 have the status of ‘YES’ for tobacco contributing to death. 1941 shows the status of ‘PROBABLY’. The percentage of the whole is shown as above. The statistical analysis of the Age of the male observations are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All Males** | **Contributed to Death - YES** | **Contributed to Death - PROBABLY** |
| **Male Observations** | 5491 | 3550 | 1941 |
| **Median Age** | 70 | 70.5 | 70 |
| **Minimum Age** | 0 | 0 | 24 |
| **Maximum Age** | 104 | 104 | 101 |
| **Skew** | -0.248353 | -0.274723 | -0.195193 |
| **Mode** | 73 | 73 | 74 |
| **Occurrences of the Mode** | 165 | 119 | 58 |

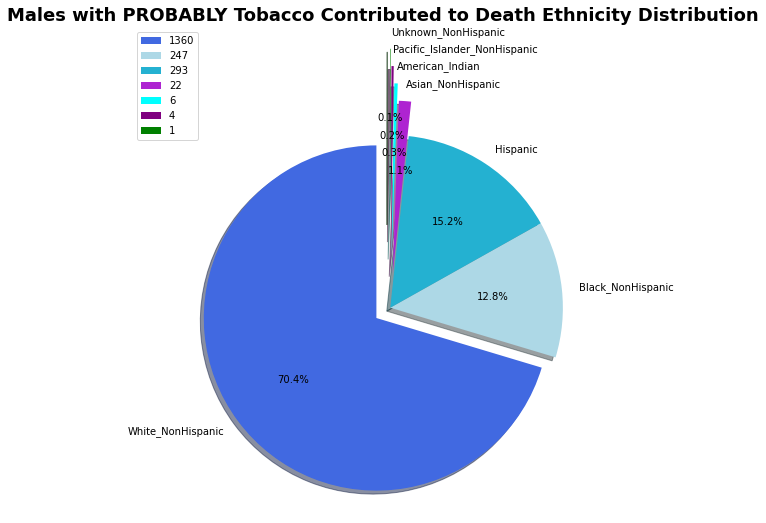


Race/Ethnicity data is missing for December YEAR 2018 due to system errors. There are 13 records that have null values for this date that are not included in this analysis.

The legend shows the number of observations that were in the data set for each ethnicity distribution. Note that there is one additional group included for males that was not included in females. That group is ‘Unknown NonHispanic’ with 2 records.

The Race/Ethnicity groups include:

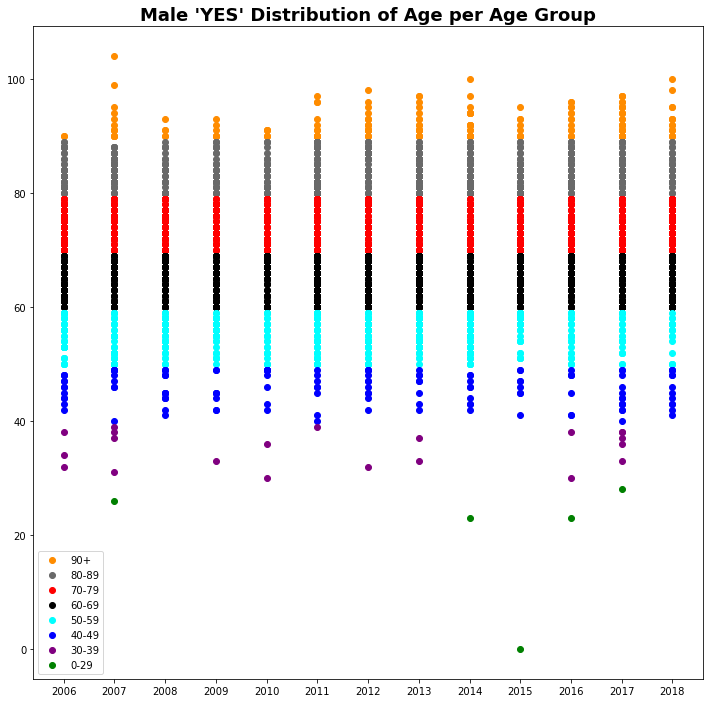
* American Indian
* Asian NonHispanic
* Black NonHispanic
* Hispanic
* Other NonHispanic
* Pacific Islander NonHispanic
* Unknown NonHispanic
* White NonHispanic

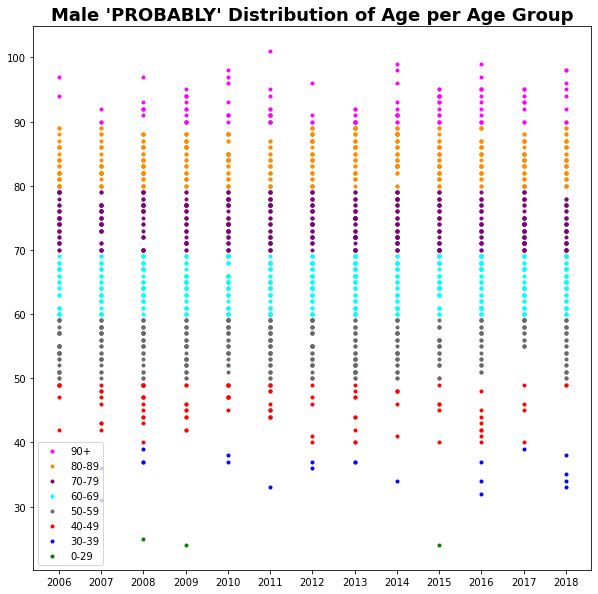


There were no records for race/ethnicity classification for ‘Other\_NonHispanic’.

There are 8 records that have null values for this date that are not included in this analysis due to system errors from December 2018.

The following two charts show the distribution of Age in each Age Group for males:





The total number of observations per year is presented in the following charts:





# Conclusion

This report is mainly rearranging the data set to get a clearer understanding of the data. Pie charts are used to show the percentage distribution of the data. They reveal which percentage contributed to the death of the observed, and which percentage was a probable cause that contributed to death. They are also used to show the ethnicity breakdown of each racial group.

Charts are used to show the statistical analysis of the ‘Age’ within the data set. The data is subdivided into all data, data containing ‘YES’ for contributing to death of the observed, and data containing ‘PROBABLY’ for contributing to death of the observed.

Finally, scatter plots are created to show the distribution of the ages within age groups for the years between 2006 and 2018.

# 