Analysis Diabetes Disease Diagnosed for US Population

Team 4: Big Bang Data Team

Members: Manuel Borrego

Orlando García

Jose H. Sierra

Gustavo Maldonado

**Introduction:**

These days everyone has a relative or a close friend who has been diagnosed with diabetes. We can see that is getting more frequent every day. In this matter we decided to **Analyze Diabetes Disease Diagnosed for US Population**.

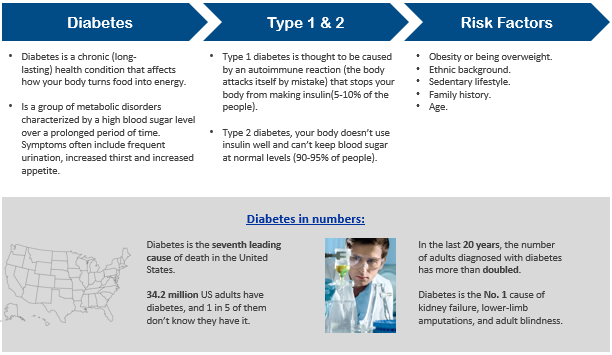
As a team we define that Our Project main objective is to discover the incidents of social demographics in diagnosed diabetes for population in US. We will analyze the relationship between gender, race-ethnicity, ages and educational levels through the years, as the data admits.

We approach our project in a way to be able to answer the following Questions:

* What is the proportion by ethnicity with diagnosed diabetes?
* How is the disease behavior in range of ages to define if there is a correlation between them?
* What´s the gender distribution for diagnosed diabetes?
* How are the impacts in the educational and income levels in relationship with diagnosed diabetes?

In order to answer these previous questions, we found that the best place to find our data Based on population in the US is the CDC, Center for Disease Control and Prevention.

**Diabetes Background:**



**Processing the data:**

The Diabetes National Data CSV was first read, filtered and cleaned. Main processing was done using the following:

* CSV reader.
* Declare main variables.
* Cut & Bins to define social demographic variable.
* Loc & dropna according to the variables.
* Get pivot table to summarize data.
* Plot the results (1,000s & percentage change).

Data filtering was done according to many measurements in the education level, age, race and gender:

* Diagnosed Diabetes.
* Percentage change & 1,000 inhabitants.
* Crude data.
* Adult aged 18+ Years.

**Work Description:**

In this work we passed through a lot. First, we did the guidelines that include the title, the team members, the project description, the research questions to answer and the dataset to be used. We divided the work by Variable, and each member pushed his work constantly in git branches.

After resolving some issues about names of files and modifications in common files, we did the merge operation, letting in the main branch all our teamwork, with data, and resources.

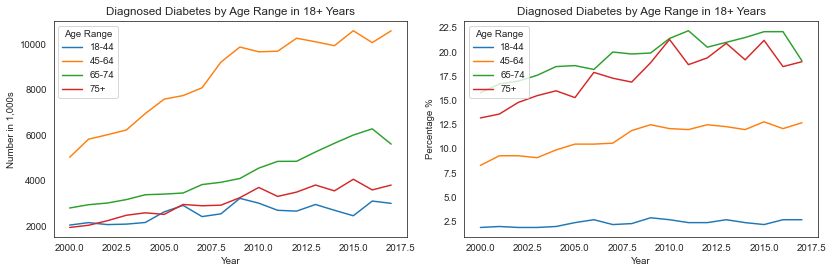
Data was cleaned and cleared, all the NAN rows where dropped, and did two pivot tables across the years from 2000 to 2017. First Pivot Table was the number of diagnosed diabetes cases for each year. Second Pivot Table was about the increase in percentage from year to year in diabetes diagnosed cases.

**Age:**

The measurements were made in four classifications:

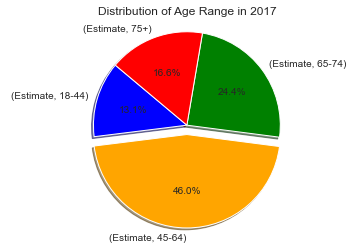
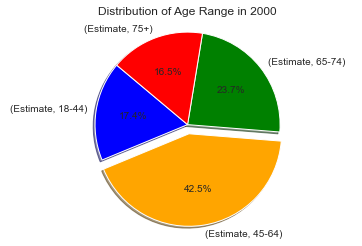
1. 18 – 44 years old.
2. 45 – 64 years old.
3. 65 – 74 years old.
4. +75 years old.

**Trend Analysis by 1,000s & percentage change.**



1. In total numbers, for High School levels, Basic Education cases don´t grow as much as Superior levels of education.
2. In percentages, from year to year the more advanced percentage growth is for Basic Education, but approximately they all keep growing the same.

**Distribution 2000 & 2017.**



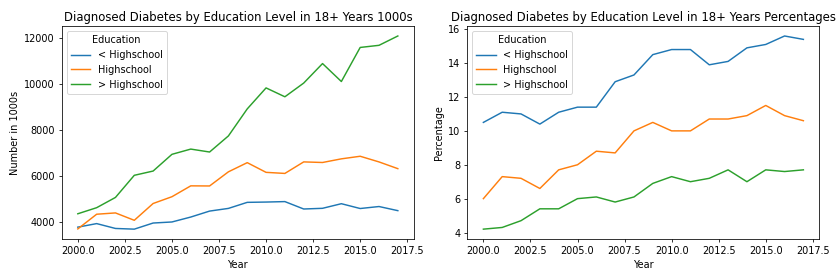
1. In 2000, 45-64 Age Range group shows a higher population diagnosed and in 2017 with more than 40%.
2. 18-44 Age Range group drops more than 4% compared with 2000 which increases in 45-64 group.

**Education:**

The measurements were made in three classifications:

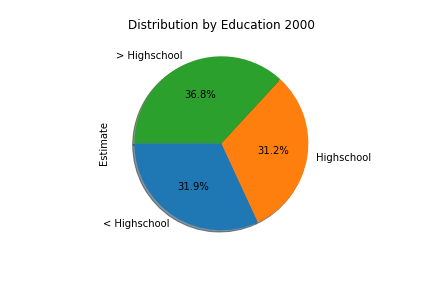
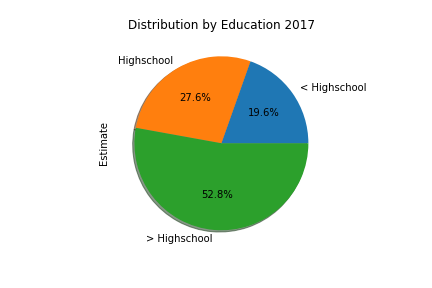
1. Less than Highschool coursed (< Highschool)
2. Highschool coursed (Highschool)
3. Studies beyond Highschool (> Highschool).

**Trend Analysis by 1,000s & percentage change.**



1. In total numbers, for High School levels, Basic Education cases don´t grow as much as Superior levels of education.
2. In percentages, from year to year the more advanced percentage growth is for Basic Education, but approximately they all keep growing the same.

**Distribution 2000 & 2017.**



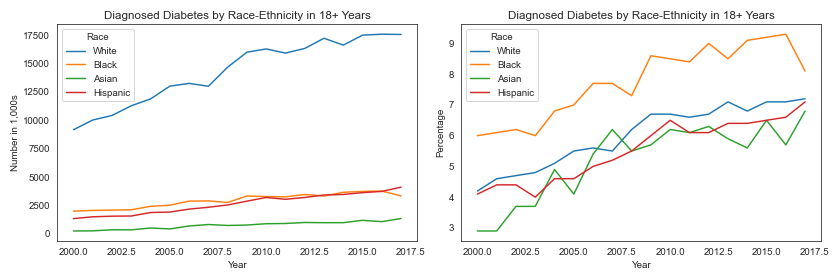
1. In the year of 2017, superior to high school levels have more than half of the cases of Diabetes.

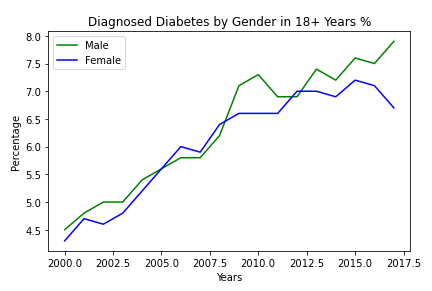
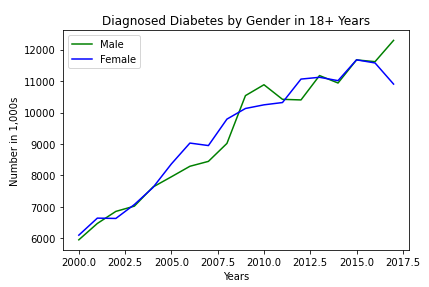
**Race:**

The measurements were made in four classifications:

1. White
2. Black
3. Asian
4. Hispanic

**Trend Analysis by 1,000s & percentage change.**





1. In total numbers, the White race has the higher level of diagnosed diabetes through the years.
2. In percentages, is interesting to observe how Black race has the higher increase proportions vs other races.

**Center for Disease Control recommendations by race.**

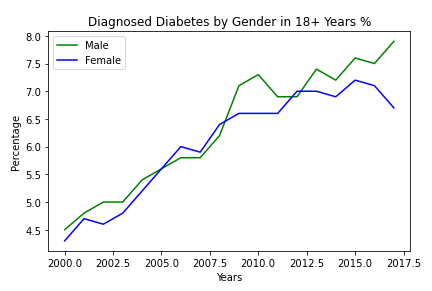
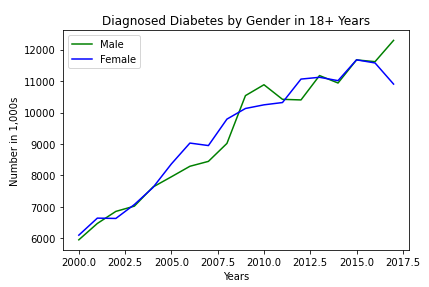
The race variable has been a key factor in a cosmopolitan country as US, that is how they tend to develop special programs to keep control and reduce the impact. Some examples:

**Gender:**

The measurements were made in two classifications:

1. Male
2. Female

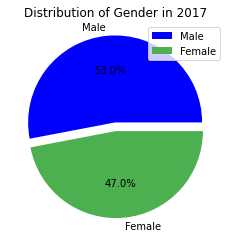
**Trend Analysis by 1,000s & percentage change.**



1. Since the year 2000 there has been an increase of 3.4% diagnosed diabetes cases representing 6,343 new cases for Male and 2.4% which are 4,805 new cases for Female.

**Distribution in 2017.**

1. In the year of 2017, the Diabetes population by Gender was 53% for Male and 47% for Female.



**Conclusions:**

* The probability to get diagnosed diabetes increases from 45 years old. It´s important to avoid risks factors and improve general health in previous years.
* The white ethnicity has the higher level of diabetes. The increase by race has been a key factor to develop programs to reduce the incidents in Asian & Hispanic people, which have a predisposition by their genetics & feeding habits.
* Higher education level represents the large group with diagnosed diabetes. As the person works more time in the office, habits derivate often to sedentarism and not good health lifestyle balance (fast food – not exercise).
* Even though the increase in all education groups is considerable, the higher education level remains as the largest group.
* The risk of developing diabetes is higher within the Male population than the Female population in the U.S. There are different reasons why man are more susceptible than woman to the develop diabetes, per example: indolence and obesity.