**Data Supplement**

**About**

DataMine is an inaugural data mining competition hosted by the UF Jacksonville College of Medicine’s Center for Data Solutions (CDS). Participants will mine literature and open-source data to help solve a pressing clinical issue. Participants will use modern data visualization techniques to create an infographic and slide deck presentation for judging by faculty from UNF and UF College of Medicine Jacksonville.

**Project Introduction**

Diabetes is a critical public health concern in Florida, affecting 6.4% to 23.6% of adults over the age of 18.1 Patients with poor blood glucose control and diabetes management have a greater risk of acute and chronic complications, resulting in hospitalizations and significant medical expenditures.2,3 The number of emergency department (ED) visits due to diabetes has increased from 2010 to 2019 in Florida. However, it is unclear what is driving the increase in ER utilization. ED utilization among patients with diabetes is likely affected by several factors, including lack of primary care, poor adherence to care plans and lifestyle and community factors.4 This study seeks to characterize ED utilization by county using demographic, behavioral, and community indicators. Please visit [our documentation](https://github.com/doremon3210/Data-mine-competition-2021) on GitHub for detailed information, including the datasets, analysis codes, and results.

**Methods**

The counties’ characteristics were summarized using descriptive statistics. In the univariate analyses, simple regression models included each outcome (as the dependent variable) and individual predictors, once at the time (as the independent variable). Then, all the independent variables were included in two multiple regression models (one for each outcome) to account for their joint effects. The assumption of normality was assessed using skewness and kurtosis of the distribution of each outcome. The data for both outcomes, diabetes-related ED visits and deaths, respectively, showed no significant departures from normality (each skewness and kurtosis was equal to or less than 1). A predictor was considered statistically significant in predicting the outcome if p-value (p) was less than 0.05.

***Descriptive statistics:*** The rate of diabetes-related ED visits and diabetes-related deaths differ across counties in Florida. The [distribution](https://public.tableau.com/app/profile/huy.nguyen7870/viz/Distributionofdiabetes-relatedEDvisitsanddeathsinFloridacounties2019/Dashboard2#1) was mapped using Tableau (see Figure 1 for a screenshot). By hovering over each county, a pop-up box will indicate the county’s population along with the diabetes-related ED visits and deaths per 100,000. Descriptive statistics of the variables used in our analysis can also be found here. [Click here](https://github.com/doremon3210/Data-mine-competition-2021/blob/main/Results/descriptive_statistics_table.xlsx) for the descriptive statistics table

**Figure 1: Screenshot of diabetes-related ED visits and deaths across Florida counties.**

Map

Description automatically generated

***Univariate analysis:*** Univariate analyses were conducted with the following variables for diabetes-related ED visits and deaths: Black Population, Individuals 25 years and over with no high school diplomat, Renter-occupied housing units, Adults who are current smokers, Adults who are obese, Adults who are sedentary, Adults who had a medical checkup in the past year, Adults who are overweight, Median household income.

Figure 2 shows the analysis of median household income on diabetes-related ED visits. By hovering over the straight line in the middle, the analysis indicates for every one unit increase in the Median household income in county, the diabetes related ED visits decrease by 0.0056.​ The two curved lines above and below the straight line illustrates the 95% confidence interval.

Interactive versions of the graphs can be accessed [here](https://public.tableau.com/app/profile/huy.nguyen7870/viz/UnivariateanalysisEDVisitsDiabetes/Dashboard1?publish=yes) (diabetes-related ED visits) and [here](https://public.tableau.com/app/profile/huy.nguyen7870/viz/SecondaryoutcomeOverall/Dashboard1?publish=yes) (diabetes-related ED deaths) respectively. To change the variable of interest, please use the radio buttons on the right as highlighted by the red box in Figure 2 below. Sub-analyses were also performed to investigate whether there will be racial differences. Please visit the links listed in Table 1 for the sub-analyses.

**Figure 2: Screenshot of univariate analysis of diabetes-related ED visits**

Chart, scatter chart

Description automatically generated

**Table 1: Links for interactive graphs for sub-analyses**

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Race** | **Link** |
| Diabetes-related ED visits | White | [Link 1](https://public.tableau.com/app/profile/huy.nguyen7870/viz/UnivariateanalysisWhite/Dashboard3) |
| Black | [Link 2](https://public.tableau.com/app/profile/huy.nguyen7870/viz/UnivariateanalysisBlack/Dashboard4) |
| Diabetes-related deaths | White | [Link 3](https://public.tableau.com/app/profile/huy.nguyen7870/viz/SecondaryoutcomeWhite/Dashboard3?publish=yes) |
| Black | [Link 4](https://public.tableau.com/app/profile/huy.nguyen7870/viz/SecondaryoutcomeBlack/Dashboard4?publish=yes) |

***Multivariate analysis:*** Coefficients from the multiple linear regression models are utilized to present how changes in each indicator will affect diabetes-related ED visits and deaths, overall and by race respectively. Please visit the interactive graph [here](https://public.tableau.com/app/profile/shoto.fukuda/viz/Amultiplelinearregressionmodelpredictingdiabetes-relatedEDvisitsanddeathsper100000population/Dashboard2). To adjust each indicator, use the slider bar or arrow buttons on the right as highlighted by the red box in Figure 3 below. County-specific models can be found [here](https://public.tableau.com/app/profile/shoto.fukuda/viz/County-specificpredictionsfordiabetes-relatedEDvisitsanddeathsper100000population/MultiplelinearregressionDiabetes). As seen in Figure 4, the predicted rate per 100,000 is indicated by the orange dot and the blue bar denotes the actual rate per 100,000. Use the dropdown arrow highlighted by the red box in Figure 4 below to change the specific outcomes. The indicators can be adjusted using the slider bar or arrow buttons on the right, similar to Figure 3.

Please visit the Tableau codes [here](https://github.com/doremon3210/Data-mine-competition-2021/tree/main/Data%20Visualizations%20and%20Data%20Supplement/Tableau%20codes).

**Figure 3: Screenshot of multivariate analysis interactive graphs**

Chart, bar chart

Description automatically generated

**Figure 4: Screenshot of county-specific multivariate analysis graphs**

**Chart

Description automatically generated**

**References**

1. Florida Diabetes Advisory Council. 2019. Florida Diabetes Legislative Report. Tallahassee, FL
2. Washington RE (AHRQ), Andrews RM (AHRQ), Mutter RL (AHRQ). Emergency Department Visits for Adults with Diabetes, 2010. HCUP Statistical Brief #167. November 2013. Agency for Healthcare Research and Quality, Rockville, MD
3. Centers for Disease Control and Prevention. National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2011.
4. Fraze TK, Jiang HJ, Burgess J. Hospital Stays for Patients with Diabetes, 2008. HCUP Statistical Brief #93. August 2010. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb93.pdf.