Multiple Linear Regression Analysis on the Number of Doctor Visits

Tong Wu

Dec 22 2020

Code and data supporting this analysis is available at: https://github.com/tongwu912/STA304finalproject

Abstract

There is an increasing trend in healthcare utilization globally, and this brings an increase in healthcare demands, especially among the aging population. The report estimates four aspects that could potentially have an impact on the annual number of doctor visits, using the longitudinal survey data collected by Medical Expenditure Panel Survey (MEPS) in the United States in 2000. Multiple linear regressions using R regarding demographic characteristics, geographical regions, ethnicity and chronic condition on the respondent's annual number of doctor visits are carried out. There is sufficient evidence to conclude that the annual number of doctor visits is correlated with respondent's demographic characteristics, some geographical regions, certain ethnicity groups and chronic condition.

Keywords

Multiple linear regression, health services utilization, longitudinal survey, annual number of doctor visits, demographic characteristics, geographical region inqeuity, racial disparities, chronic condition, private insurance

Introduction

With the advancement of technology and the improvement of living conditions, individuals' health awareness is increasing gradually. Health is not restricted to physical health, but also to mental health and to social well-being. The change in health consciousness deepened people's understanding towards health, and it led to a rising trend in the number of doctor's visits globally. Canadians on average visit the doctor seven times a year, while Americans visit the doctor four times a year on average. The mean number of doctor visits in the United States is substantially less than in other developed countries. On the other hand, Japanese visit the doctor thirteen times a year on average, which is the highest amongst all other countries (McCarthy, 2014). In addition, visits to doctor clinics in India increased to 3.2 times a year in 2018 (Mukherjee, 2019). Health care utilization is the quantification of the use of health care services. It can be reported by the visits to the physician. People use health care services for preventing and curing health problems (Carrasquillo, 2013). The aim of this report is to investigate the factors that would influence healthcare utilization measured in the annual number of doctor visits in the United States.

Healthcare utilization is an important indicator of types of care specific populations seek, and it also indicates how healthcare services may be shifting from one site to another (Bernstein et al., 2003). Health care demands increase as the world population increases. The analysis bases inference on four perspectives: individuals characteristics, geographical regions, ethnicity, chronic condition and private insurance status. On the individual level, many empirical studies report that there is a disparity in health care utilization between men and women. Women tend to have higher rates of morbidity and health care utilization than

men (Roy and Chaudhuri, 2012). Age is also a potential factor that would affect the number of physician visits. Observational data shows individuals above the age of 80 have significantly more physician visits and emergency room visits (Vegda et al., 2009). And health conditions are closely tied with age. Furthermore, socio-economic status is positively correlated with health care utilization. Individuals with high economic stability have better access to health care services. Low-income individuals are more likely to have higher medical needs, but are 6% less likely to visit a physician over a one-year period (Lemstra et al., 2009). On the geographical regions level, the accessibility to health services vary from region to region. Region-related inequity in healthcare utilization is more serious in rural areas since there are a large number of low-income groups.

On the other hand, a number of empirical studies have claimed there is a disparity in health service utilization among different ethnicity groups. Utilization across all types of healthcare services experienced underutilization compared with Whites (Gandhi et al., 2018). On the health condition level, a chronic condition was listed as the major reason for 37% of all office-based physician visits in 2016 (Ashman, 2019). Chronic condition is essential to the number of physician visits. And having private insurance would encourage individuals to visit the doctor frequently, thereby increasing the annual number of doctor visits.

The approach of the study is to carry out four regression models to accommodate four different aspects that are influencing the annual number of doctor visits. The dataset is the longitudinal survey data of medical expenses, collected by the Medical Expenditure Panel Survey (MEPS) in the United States in 2000. The dataset will be used to find the potential factors impacting the annual visits to the doctors on individuals' in four different perspectives. In the Methodology section, the source of the dataset is explained, four multiple linear regressions are performed with different sets of independent variables, and the models notations are written. Summary tables of the regression are provided in the Results section, estimated formulas of the model are written, and inferences of the model along with conclusions are interpreted in the Conclusion section. The weaknesses of the study and next steps to take in order to improve the model are stated in the Discussion section. Pie charts with respect to the distribution of the explanatory variables are illustrated in the Appendix.

Methodology

Data

The data set used in this report is the 2000 Medical Expenditure Panel Survey (MEPS) which can be downloaded on the MEPS official website under the Agency for Healthcare Research and Quality. This survey features a large-scale longitudinal survey of families and individuals, the medical providers, and employers across the United States. MEPS collects data on the specific types of health services that Americans use, the frequency they use them and the cost of these health services etc. The MEPS uses multiple data collection methods. Individuals are initially screened by telephone to confirm their mailing addresses and to establish a point-of-contact with a knowledgeable respondent. If the respondent's firm does not offer health insurance, the noninsurance questions are asked at that time, thus completing the survey by telephone. Those who do not respond to the telephone screening are mailed survey questionnaires.

The population is all civilian non-institutionalized persons in the United States. The frame population was the owners of the telephone numbers that were called, and the sampled population was the ones who answered the survey questions through the phone calls. The drawbacks of this dataset are the survey method, outdated data and unsufficient ethnicity groups. This survey initially adopted telephone screening which may introduce selection bias. The survey data was collected in 2000, which might have less reference value. And there are only 2 ethnicity groups recorded in the survey: hispanic and black. This restrict the researchers to observe the correlation amongst other ethnicity groups. To accommodate the focuses of this report, which are on the frequency of doctor visits, respondent's demograhic characteristics, geographical regions, ethnicity, chronic condition and private insurance status, the original dataset was cut down to 13 columns and 3118 observations. The response variable is docvis. The four variables that are related to respondent's demographic characteristics: edu, income, age and female. The four variables that are related to respondent's geographical region: northeast, midwest, south and msa. The two variables that are related

to respondent's ethnicity: hispanic and black. The two variables that are related to health and private insurance status: chronic and private.

Table 1: Raw Data

docvis	educ	income	age	female	northeast	midwest	south	msa	hispanic	black	chronic	private
0	12	2.405	5.0	1	0	1	0	1	0	0	0	1
0	12	7.040	2.8	1	0	1	0	1	0	0	0	1
0	12	50.000	4.9	1	0	0	1	1	0	0	0	1
6	12	20.903	4.9	1	0	1	0	1	0	1	1	1
1	12	13.633	2.7	0	0	0	0	1	0	0	0	0
6	14	1.773	4.9	1	1	0	0	1	1	0	1	1
0	16	19.005	2.7	0	0	1	0	1	0	0	0	0
0	12	10.394	2.6	0	0	0	0	1	1	0	0	0
0	12	12.000	2.6	0	0	1	0	1	0	0	0	1
1	17	30.000	2.7	1	0	0	1	0	0	1	0	1

Variables

- docvis: respondent's annual number of doctor visits
- educ: respondent's years of schooling
- income: respondent's income in $/1000
- age: respondent's age
- female: dummy variable for respondent's sex. female = 1 if the respondent is female, and 0 otherwise
- northeast: dummy variable for respondent's geographical region. northeast = 1 if the respondent lives in the northeast region, and 0 otherwise
- midwest: dummy variable for respondent's geographical region. midwest = 1 if the respondent lives in the midwest region, and 0 otherwise
- south: dummy variable for respondent's geographical region. south = 1 if the respondent lives in the southern region, and 0 otherwise
- msa: dummy variable for respondent's geographical region. msa = 1 if the respondent lives in metropolitan statistical area, and 0 otherwise
- hispanic: dummy variable for respondent's ethnicity. hispanic = 1 if the respondent is of hispanic descent, and 0 otherwise
- black: dummy variable for respondent's ethnicity. black = 1 if the respondent is of black descent, and 0 otherwise
- chronic: dummy variable for respondent's health status. chronic = 1 if the respondent has chronic condition, and 0 otherwise
- private: dummy variable for respondent's private insurance status. private = 1 if the respondent has a private insurance, and 0 otherwise

Model

Multiple linear regressions are used to measure the correlation of the annual number of doctor visits and four other aspects (demographic characteristics, geographical region, ethnicity, chronic condition status and private insurance status) in this report. The reason for choosing these aspects is because a number of empirical studies believed that there is a linear relationship between these perspectives and the number of doctor visits. Multiple linear regressions allow to predict the correlation of more than one explanatory variables. The four multiple linear regressions in this report are carried out using R. There is one response variable (docvis) and twelve independent variables. Three of the independent variables are numeric (educ, age, income), the rest are categorical (female, northeast, midwest, south, msa, hispanic, black, chronic, private).

The first model is the multiple linear regression of respondent's demographic characteristics on the frequency of the physician visits. The dependent variable is the respondents' annual number of doctor visits. The independent variables include respondent's years of education, income, age, and sex. These are the respondents' intrinsic characteristics, and they are invariant to the external environment. From the data section: educ = respondent's years of education, income = respondent's income in $/1000, age = respondent's age, female = indication of respondent's sex. It equals to 1 if the respondent is female, and 0 otherwise.

And the multiple linear regression model is:

$$docvis = \beta_0 + \beta_1 educ + \beta_2 income + \beta_3 age + \beta_4 female + \epsilon$$

where β_0 is the intercept of the model. β_1 is the parameter for the explanatory variable education, β_2 is the parameter for the explantory variable income, β_3 is the parameter for the explantory variable age, β_4 is the parameter for the explantory variable female.

The second model is the multiple linear regression of respondent's geographical region on the frequency of the physician visits. The goal of this regression model is the investigate the inequality in accessibility to health care services among different geographical locations. The dependent variable is the respondents' annual number of doctor visits. The independent variables include 4 dummy variables regarding geographical regions: northeast, midwest, south and metropolitan. From the data section: northeast is a dummy variable that equals to 1 if the respondent lives in northeast region, midwest is a dummy varianle that equals to 1 if the respondent lives in midwest region, south is a dummy variable that equals to 1 if the respondent lives in southern region, metropolitan is a dummy variable that equals to 1 if the respondent is in metropolitan statistical area.

And the multiple regression model is:

$$docvis = \beta_0 + \beta_1 northeast + \beta_2 midwest + \beta_3 south + \beta_4 msa + \epsilon$$

where β_0 is the intercept of the model. β_1 is the parameter for the dummy variable northeast, β_2 is the parameter for the dummy variable midwest, β_3 is the parameter for the dummy variable south, β_4 is the parameter for the dummy variable metropolitan.

The third model is the multiple linear regression of the respondent's ethnicity on the frequency of physician visits. The goal of this regression model is the observe the health care utilization difference among different ethnicity. The dependent variable is the respondent's annual number of doctor visits. The independent variables are the only two dummy variables that represent respondent's ethnicity in the dataset: hispanic and black. From the data section: hispanic is a dummy variable that equals to 1 if the respondent is of hispanic descent, black is a dummy varianle that equals to 1 if the respondent is of black descent.

And the multiple regression model is:

$$docvis = \beta_0 + \beta_1 hispanic + \beta_2 black + \epsilon$$

where β_0 is the intercept of this model. β_1 is the parameter for the dummy variable hispanic. β_2 is the parameter for the dummy variable black.

The last model is the multiple linear regression of the respondent's health condition and private insurance status on the frequency of physician visits. The goal of this regression model is to observe the effect of chronic condition and private insurance have on the number of physician visits. The dependent variable is the respondent's annual number of doctor visits. The independent variables are the two dummy variables in the dataset: chronic and private. From the data section: chronic is a dummy variable that equals to 1 if the respondent has chronic condition, private is a dummy variable that equals to 1 if the respondent has a private insurance. The interaction term between chronic and private is to investigate whether the frequency of physician visits would be different for respondents having both chronic condition and a private insurance.

And the multiple regression model is:

$$docvis = \beta_0 + \beta_1 chronic + \beta_2 private + \beta_3 chronic * private + \epsilon$$

where β_0 is the intercept of this model. β_1 is the parameter for the dummy variable chronic. β_2 is the parameter for the dummy variable private. β_3 is the parameter for the interaction term of chronic and private.

Results

Table 2: Coefficients of Characteristics on Number of Doctor Visits

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.2826	0.8975	-3.6574	0.0003
educ	0.1422	0.0532	2.6718	0.0076
income	0.0203	0.0063	3.2002	0.0014
age	0.8538	0.1460	5.8467	0.0000
female	1.8947	0.2989	6.3394	0.0000

Table 2 is the multiple linear regression summary table of respondent's demographic characteristics on the number of doctor visits. The estimated regression model is:

$$\hat{docvis} = -3.2826 + 0.1422educ + 0.0203income + 0.8538age + 1.8947female$$

The p-values for all explanatory variables are less than 0.05. As can be seen in the appendix figure 1, there is a positive trend between years of education and frequency of doctor visits. Additionally, figure 2 in the appendix shows the proportion of female respondents is 47.75% whereas it is 52.25% for male respondents.

Table 3: Coefficients of Regions on Number of Doctor Visits

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1857	0.4561	6.9852	0.0000
northeast	1.1253	0.4879	2.3063	0.0212
midwest	1.1729	0.4357	2.6922	0.0071
south	0.2786	0.3959	0.7038	0.4816
msa	-0.1589	0.3870	-0.4105	0.6815

Table 3 is the multiple linear regression summary table of respondent's geographical region on the number of doctor visits. The estimated regression model is:

$$\hat{docvis} = 3.1857 + 1.1253northeast + 1.1729midwest + 0.2786south - 0.1589msa$$

The p-values for explanatory variables northeast and midwest are less than 5%. As can be seen in the appendix figure 3, 15.49% of the respondents live in the northeast region, 23.51% live in the midwest region, 36.85% live in the southern region and 24.15% live in other regeions. In addition, figure 4 in the appendix shows 80.85% of the respondents live in metropolitan areas, while 19.15% live in rural areas.

Table 4: Coefficients of Ethnicity on Number of Doctor Visits

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.2887	0.1846	23.2288	0
hispanic	-1.7781	0.3639	-4.8869	0
black	-2.2703	0.4653	-4.8797	0

Table 4 is the multiple linear regression summary table of respondent's ethinicity on the number of doctor visits. The estimated regression model is:

$$\hat{docvis} = 4.2887 - 1.7781hispanic - 2.2703black$$

The p-values for all explanatory variables are less than 0.05. Furthermore, figure 5 in the appendix shows 12.19% of the respondents are of black descent, 22.61% are of hispanic descent and 65.2% are of other ethnicity background.

Table 5: Coefficients of Health and Private Insurance Status on Number of Doctor Visits

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8873	0.3932	2.2564	0.0241
chronic	2.7770	0.7782	3.5687	0.0004
private	1.5890	0.4371	3.6353	0.0003
chronic:private	2.3342	0.8508	2.7437	0.0061

Table 5 is the multiple linear regression summary table of respondent's health and private insurance status on the number of doctor visits. The estimated regression model is:

$$\hat{docvis} = 0.8873 + 2.777chronic + 1.589private + 2.3342chronic * private$$

The p-values for all explanatory variables are less than 5%. As can be seen in the appendix figure 6, 29.86% of the respondents have chronic conditions, 70.14% do not. Additionally, figure 7 in the appendix shows that 82.04% of the respondents have private insurance while 17.96% do not.

Discussion

Summary

Using the 2000 Medical Expenditure Panel Survey that is available to download on the official MEPS website under the Agency for Healthcare Research and Quality, to perform four multiple linear regressions of different aspects are carried out in the model section: respondent's characteristics, geographical region, ethnicity and private insurance status on number of doctor visits. In addition, a brief summary of the regression results and some statistics on the respondent's explanatory variables distribution are stated in the previous section. The following part will be focusing on the interpretation and conclusion of each multiple linear regression model.

Conclusions

The first regression model is respondent's demographic characteristics on the number of doctor visits. The estimated model is $\hat{docvis} = -3.2826 + 0.1422educ + 0.0203income + 0.8538age + 1.8947female$. The interpretations of the paraters are:

- holding everything else constant, for every additional year in schooling, the annual number of doctor visits increases 0.1422 units on average.
- holding everything else constant, for every \$1/1000 increase in income, the annual number of doctor visits increases 0.0203 units on average.
- holding everything else constant, for every additional increase in age, the annual number of doctor visits increases 0.8538 units on average.
- holding everything else constant, if the respondent is female, her annual number of doctor visits are on average 1.8947 units higher than male.
- The intercept corresponds to the number of doctor visits if the respondent receives 0 years of education, no income, age is 0 and is male. It is not applicable to reality, so the intercept has no indication in this context.

Since the p-values for all explanatory variables are less than 5%, there is sufficient evidence to reject the null hypothesis. In other words, years of education, income, age and sex are correlated with the frequency of doctor visits. This confirms with the empirical findings that women tend to have higher rates of morbidity and health care utilization than men (Roy and Chaudhuri, 2012). It also validates age and frequency of doctor visits move in the same direction, as individuals above the age of 80 have significantly more physician visits and emergency room visits (Vegda et al., 2009).

The second regression model is respondent's geographical region on the number of doctor visits. The estimated model is $\hat{docvis} = 3.1857 + 1.1253northeast + 1.1729midwest + 0.2786south - 0.1589msa$. The interpretations of the paraters are:

- holding everything else constant, if the respondent lives in northeast region, his number of doctor visits is on average 1.1253 units higher than those who do not live in the northeast region.
- holding everything else constant, if the respondent lives in midwest region, his number of doctor visits is on average 1.1729 units higher than those who do not live in the midwest region.
- holding everything else constant, if the respondent live in southern region, his number of doctor visits is on average 0.2786 units higher than those who do not live in the southern region.
- holding everything else constant, if the respondent is in a metropolitan statistical area, his number of doctor visits is 0.1589 units higher than those who are in rural areas.
- The intercept is the mean number of doctor visits if the respondent does not live in northeast, midwest or southern metropolitan area.

Since the p-values for northeast and midwest region are less than 5%, the null hypothesis is rejected, the estimated coefficients are statistically significant. That is, there is a linear relationship between the number of doctor visits and northeast/midwest region. However, the p-values for south and msa are larger than 5%, there is not enough evidence to conclude the effect of southern region and metropolitan area on the number of doctor visits.

The third regression model is respondent's ethinicity on the number of doctor visits. The estimated model is $\hat{docvis} = 4.2887 - 1.7781hispanic - 2.2703black$. The interpretations of the paraters are:

- holding everything else constant, if the respondent is of hispanic descent, his number of doctor visits is on average 1.7781 units lower than those who are of other ethnicity groups.
- holding everything else constant, if the respondent is of black descent, his number of doctor visits is on average 2.2703 units lower than those who are of other ethnicity groups.
- The intercept is the mean number of doctor visits if the respondent belongs to neither hispanic nor black.

Since the p-value for both explanatory variables are less than 5%, the estimated coefficients are statistically significant. The null hypothesis is rejected, there is a strong evidence to suggest that the number of doctor visits is negatively correlated with respondents of hispanic and black descent.

The last regression model is respondent's chronic condition and private insurance status on the number of doctor visits. The estimated model is $\hat{docvis} = 0.8873+2.777chronic+1.589private+2.3342chronic*private$. The interpretations of the paraters are:

- holding everything else constant, if the respondent has chronic condition, his number of doctor visits is on average 2.777 units higher than others who do not.
- holding everything else constant, if the respondent has a private insurance, his number of doctor visits is on average 1.589 units higher than others who do not.
- holding everything else constant, if the respondent has both chronic and private insurance, his number of doctor visits is on average 2.3342 units higher than others who do not have both.
- The intercept of this model corresponds to the mean number of doctor visits if the respondent have neither chronic condition nor private insurance.

Since the p-values for all explanatory variables are less than 0.05, there is sufficient evidence to reject the null hypothesis. In other words, chronic condition and private insurance status are correlated with the number of doctor visits. This matches with the empirical study which claims chronic pain increased the frequency of visits to physicians with a 95% confidence interval between 2.8 to 7.9 times (Mann et al., 2016).

In conclusion, this report identifies the annual number of doctor visits is correlated with respondent's demographic characteristics (years of education, age, income and gender), certain geographical location (northeast and midwest region of the United States), certain ethnicity (hispanic and black), chronic condition and private insurance status.

Weaknesses

There are some limitations and drawbacks of the survey and sampling method. The survey data was collected in 2000 which is 20 years ago, many things can be different in 20 years. People's values of health services might be different, thereby changing the health awareness and health care utilization, making the survey data have less indication in the current situation. The survey used telephone interview to survey respondents which may introduce selection bias. The survey only applied to individuals who were at home and answered the phone call, the characteristics of the surveyed population are different from the target population. On the other hand, the categories of ethnicity groups are insufficient. As can be seen in the appendix figure 5, 65.2% of the respondents belong to neither hispanic nor black. The survey need to take more ethnicity groups into account. Furthermore, in the regression model of geographical region on the number of doctor visits, the p-value for msa (metropolitan statistical area) is 0.6815, which is relatively large. However, a number of empirical studies have claimed the ineuity in accessibility to health services do exist among urban and rural areas. It is very likely that the null hypothesis is true, but was failed to reject it. In other words, the likelihood of making a type II error is high.

Next Steps

It would be better to use a survey data that was collected in recent years, as it would obtain a more precise estimation of the correlation on the respondent's current number of doctor visits. And recent large-scale survey allows for greater diversification in explanatory variables. Combining the four multiple regressions would be a plausible next step to take, this approach can create a relatively suited model in terms of R squared statistics, even though colinearity amongst independent variables may occur. On the other hand, to decrease the probability of making a type II error as mentioned above, one approach is to increase the statistical power of the study. Increasing the sample size can efficiently increase the statistical power. And the more data in the model, the more accurate the model is. Additionally, to make the regression model more convincing, expanding the survey range will allow to obtain a more accurate estimation.

References

1. Dataset

Lynn A. Blewett, Julia A. Rivera Drew, Risa Griffin and Kari C.W. Williams. IPUMS Health Surveys: Medical Expenditure Panel Survey, Version 1.1 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D071.V1.1

2. Software

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

3. Packages

Hadley Wickham and Evan Miller (2020). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.3.1. https://CRAN.R-project.org/package=haven

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.29.

4. Websites

Carrasquillo, Olveen. "Health Care Utilization." SpringerLink, Springer, New York, NY, 1 Jan. 1970, link.springer.com/referenceworkentry/10.1007/978-1-4419-1005-9_885.

"Characteristics of Office-Based Physician Visits, 2016." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 23 Jan. 2019, www.cdc.gov/nchs/products/databriefs/db331.htm.

Chen, Jie, et al. "Health Care Spending and Utilization by Race/Ethnicity under the Affordable Care Act's Dependent Coverage Expansion." American Journal of Public Health, American Public Health Association, July 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4455492/.

FT. Bobo, AY. Elias, et al. "The Comparison of Healthcare Utilization Inequity between URRBMI and NCMS in Rural China." International Journal for Equity in Health, BioMed Central, 1 Jan. 1970, equityhealthj.biomedcentral.com/articles/10.1186/s12939-019-0987-1.

JJ;, Gandhi K;Lim E;Davis J;Chen. "Racial Disparities in Health Service Utilization Among Medicare Fee-for-Service Beneficiaries Adjusting for Multiple Chronic Conditions." Journal of Aging and Health, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/28621152/.

McCarthy, Niall. "Americans Visit Their Doctor 4 Times A Year. People In Japan Visit 13 Times A Year [Infographic]." Forbes, Forbes Magazine, 4 Sept. 2014, www.forbes.com/sites/niallmccarthy/2014/09/04/americans-visit-their-doctor-4-times-a-year-people-in-japan-visit-13-times-a-year-infographic/?sh=fddff83e3475.

RA. Elzahaf, OA. Tashani, et al. "Frequency and Characteristics of Healthcare Visits Associated with Chronic Pain: Results from a Population-Based Canadian Study." Canadian Journal of Anesthesia/Journal Canadien D'anesthésie, Springer US, 1 Jan. 1970, link.springer.com/article/10.1007/s12630-015-0578-6.

Roy, Kakoli, and Anoshua Chaudhuri. "Gender Differences in Healthcare Utilization in Later Life." SpringerLink, Palgrave Macmillan, London, 1 Jan. 1970, link.springer.com/chapter/10.1057/9781137295408_16.

Rupali Mukherjee / TNN / Apr 4, 2019. "Visits to Doctors Rise Nearly 20% in 2018: India News - Times of India." The Times of India, TOI, timesofindia.indiatimes.com/india/visits-to-doctors-rise-nearly-20-in-2018/articleshow/68717879.cms.

Sciences, National Academies of, et al. "Factors That Affect Health-Care Utilization." Health-Care Utilization as a Proxy in Disability Determination., U.S. National Library of Medicine, 1 Mar. 2018, www.ncbi.nlm.nih.gov/books/NBK500097/.

T. Ostbye, KS. Yarnall, et al. "Trends in Health Services Utilization, Medication Use, and Health Conditions among Older Adults: a 2-Year Retrospective Chart Review in a Primary Care Practice." BMC Health Services Research, BioMed Central, 1 Jan. 1970, bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-9-217.

U;, Lemstra M;Mackenbach J;Neudorf C;Nannapaneni. "High Health Care Utilization and Costs Associated with Lower Socio-Economic Status: Results from a Linked Dataset." Canadian Journal of Public Health = Revue Canadienne De Sante Publique, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/19507718/.

Yihui Xie, Christophe Dervieux. “R Markdown Cookbook.” 10.1 The Function Knitr::Kable(), 21 Sept. 2020, bookdown.org/yihui/rmarkdown-cookbook/kable.html.

5. Book

Health Care in America: Trends in Utilization. U.S. Department of Health and Human Services, 2004.

Appendix

Figure 1 Scatterplot of Number of Doctor Visits versus Education

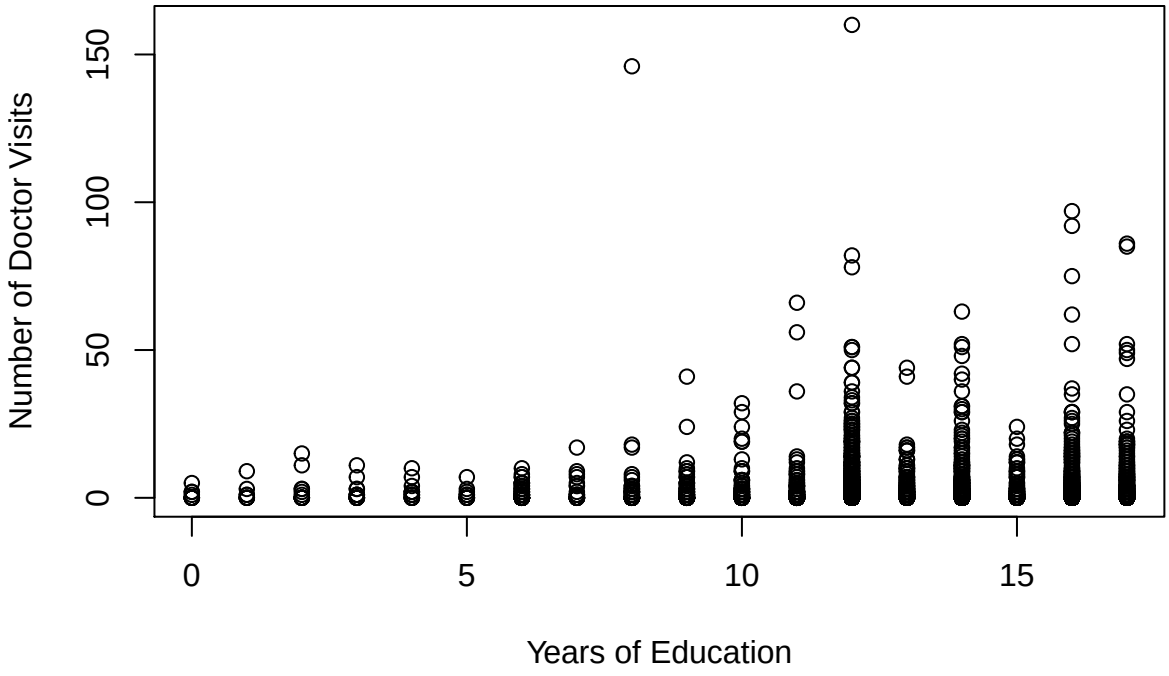

Figure 2 Pie Chart for Respondents' Gender Distribution

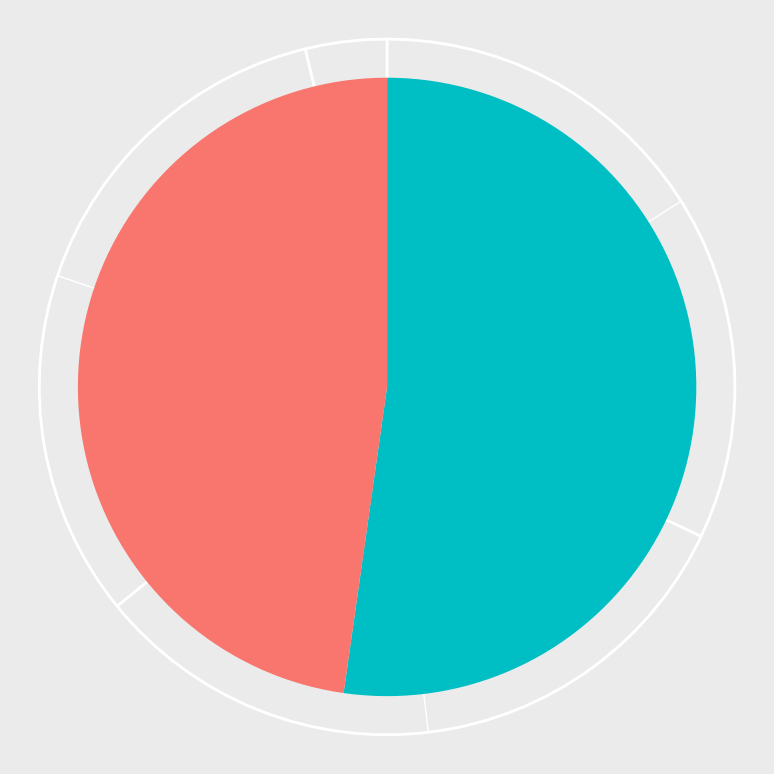

Figure 3 Pie Chart for Respondents' Geographical Region Distribution

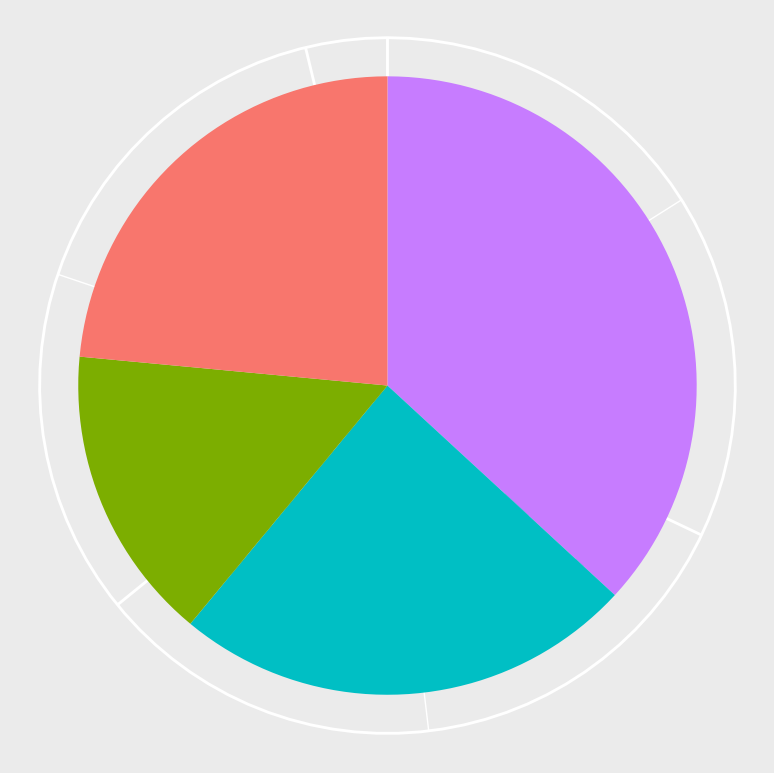

Figure 4 Pie Chart for Respondents' Metropolitan Distribution

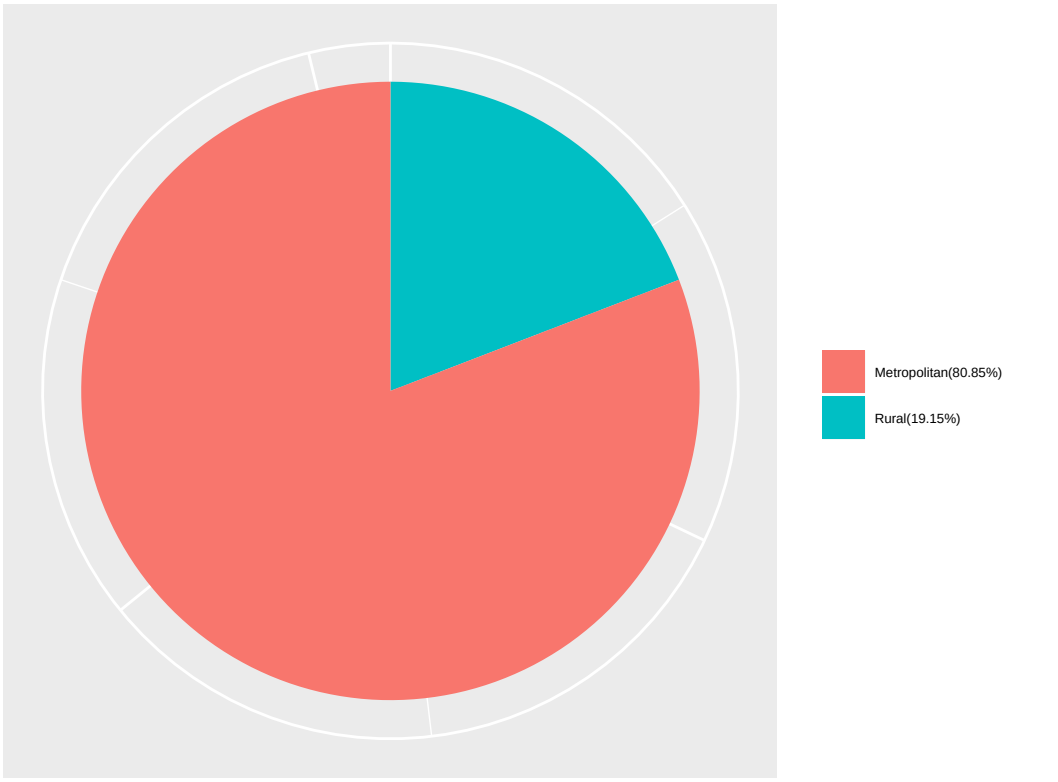

Figure 5 Pie Chart for Respondents' Ethinicity Distribution

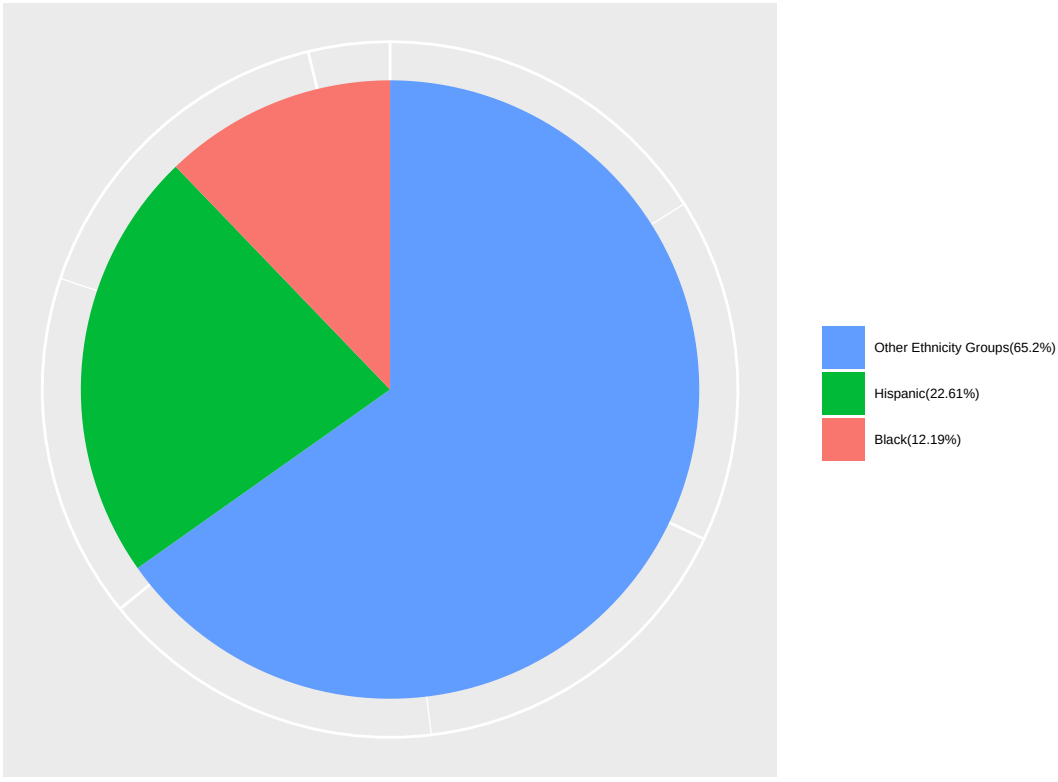

Figure 6 Pie Chart for Respondents' Chronic Condition Distribution

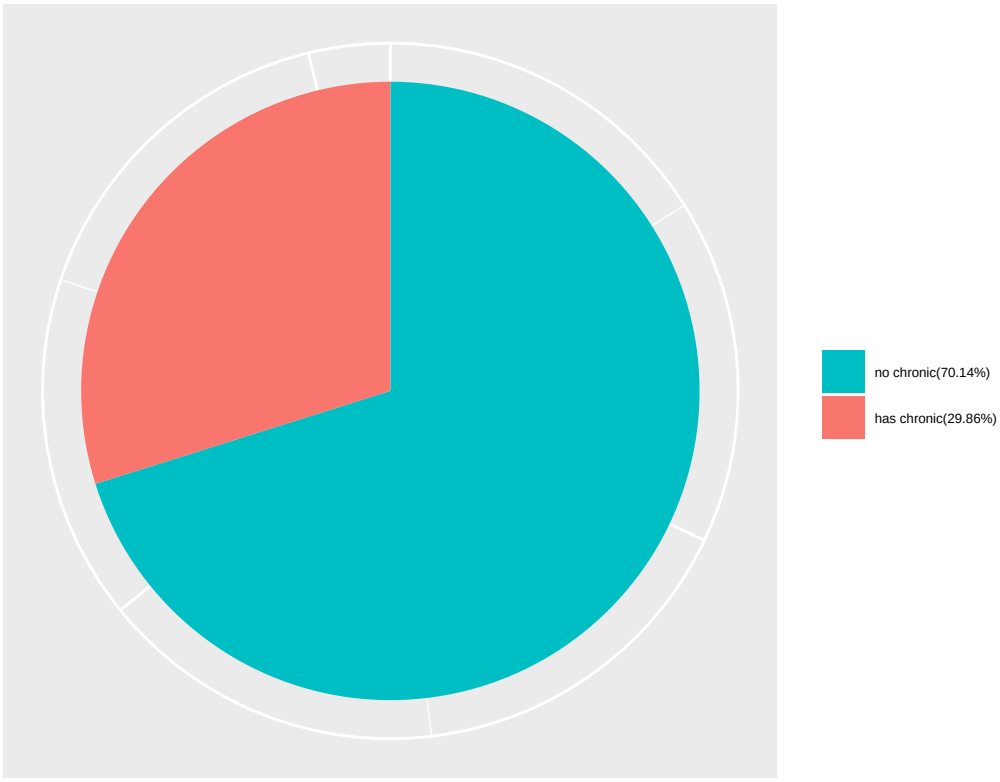

Figure 7 Pie Chart for Respondents' Private Insurance Status Distribution