Project 1 Mental Health of U.S. residents - Writeup

Yawen Deng(yd298)
Min Zheng(mz474)
Yanfei Yu(yy777)

Overview

For project 1, our group decided to analyze the mental health status of U.S. residents as well as inspecting possible underlying influence factors.

Description of Data

Our main data source is the from the 2016 survey of the Behavioral Risk Factor Surveillance System (BRFSS), which is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. The raw data and the codebook is too large to upload, so we just exclude them the zip file, but the codebook can be view from our Git repo. The survey consists of quite broad topics, but for this project we decided to focus on mental health.

Variables:
The variable we chose to represent the **mental health status** of respondents is:

Number of Days Mental Health not Good (in the past 30 days)

Label: Number of Days Mental Health Not Good
Section Name: Healthy Days — Health Related Quality of Life
Section Number: 2
Question Number: 2
Column: 93-94
Type of Variable: Num
SAS Variable Name: MENTHLTH
Question Prologue:
Question: Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

Value	Value Label	Frequency	Percentage	Weighted Percentage
1 - 30	Number of days Notes: _ _ Number of days	148,848	30.61	33.68
88	None	329,500	67.76	64.74
77	Don't know/Not sure	5,549	1.14	1.08
99	Refused	2,405	0.49	0.50
BLANK	Not asked or Missing	1	.	.

(screenshot from Codebook)

**Note: When doing data processing, we filter out observations with missing value of certain variable(e.g., filter out observation with "BLANK" MENTHLTH), or with "refused", "don't know/Not sure", "None", "NaN", etc.. Also In the grid map in "Overview" session and the heatmap, this variable(MENTHTLTH) is directly used as a continuous variable, with range [1,30]. In the "Gender" and "Habit" subsession, this variable is categorized into 5 bins, with 6 days in each bin.

To get an overview of the mental health status of the whole nation, we decided to use the variable *State FIPS Code*, which is the first variable listed in the Codebook. Next, we are interested in figuring out potential underlying cause or factors that could influence mental health status of the respondents. After inspecting the dataset and all the variables that we have, as well as the distributions of the sample, we included 5 variables in the second session (Factors Behind Mental Health), they are:

Age, *SleepTime*, *Gender*, *Current Smoke Status*, *Alcohol Consumption.*

-Age:

The original *Age* variable is in range [18, 99].

We group the original value into 13 categories, listed here.

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Age 18 to 24 Notes: 18 <= AGE <= 24	26,626	5.48	12.66
2	Age 25 to 29 Notes: 25 <= AGE <= 29	23,034	4.74	8.21
3	Age 30 to 34 Notes: 30 <= AGE <= 34	25,432	5.23	9.14
4	Age 35 to 39 Notes: 35 <= AGE <= 39	27,192	5.59	8.11
5	Age 40 to 44 Notes: 40 <= AGE <= 44	27,195	5.59	7.95
6	Age 45 to 49 Notes: 45 <= AGE <= 49	33,014	6.79	7.41
7	Age 50 to 54 Notes: 50 <= AGE <= 54	41,810	8.60	9.01
8	Age 55 to 59 Notes: 55 <= AGE <= 59	49,799	10.24	8.08
9	Age 60 to 64 Notes: 60 <= AGE <= 64	54,770	11.26	8.29
10	Age 65 to 69 Notes: 65 <= AGE <= 69	55,831	11.48	6.62
11	Age 70 to 74 Notes: 70 <= AGE <= 74	43,683	8.98	5.00
12	Age 75 to 79 Notes: 75 <= AGE <= 79	31,497	6.48	3.92
13	Age 80 or older Notes: 80 <= AGE <= 99	39,723	8.17	4.37

-SleepTime (in 1 day)

Label: Respondents Sex
Section Name: Demographics
Section Number: 8
Question Number: 1
Column: 120
Type of Variable: Num
SAS Variable Name: SEX
Question Prologue:
Question: Indicate sex of respondent.

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Male	210,606	43.31	48.66
2	Female	275,631	56.68	51.33
9	Refused	66	0.01	0.01

-Gender

Label: How Much Time Do You Sleep
Section Name: Inadequate Sleep
Section Number: 5
Question Number: 1
Column: 102-103
Type of Variable: Num
SAS Variable Name: SLEPTIM1
Question Prologue: I would like to ask you about your sleep pattern.
Question: On average, how many hours of sleep do you get in a 24-hour period?

Value	Value Label	Frequency	Percentage	Weighted Percentage
1 - 24	Number of hours [1-24]	480,577	98.82	98.97
77	Don't know/Not Sure	5,075	1.04	0.92
99	Refused	649	0.13	0.11
BLANK	Missing	2	.	.

Label: Computed Smoking Status
Section Name: Calculated Variables
Section Number: 9
Question Number: 1
Column: 2001
Type of Variable: Num
SAS Variable Name: _SMOKER3
Question Prologue:
Question: Four-level smoker status: Everyday smoker, Someday smoker, Former smoker, Non-smoker

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Current smoker - now smokes every day Notes: SMOKE100 = 1 and SMOKEDAY = 1	49,202	10.12	10.57
2	Current smoker - now smokes some days Notes: SMOKE100 = 1 and SMOKEDAY = 2	19,740	4.06	4.91
3	Former smoker Notes: SMOKE100 = 1 and SMOKEDAY = 3	134,863	27.73	23.16
4	Never smoked Notes: SMOKE100 = 2	262,670	54.01	56.24
9	Don't know/Refused/Missing Notes: SMOKE100 = 1 and SMOKEDAY = 9 or SMOKE100 = 7 or 9 or Missing	19,828	4.08	5.12

-Alcohol Consumption

Label: Heavy Alcohol Consumption Calculated Variable
Section Name: Calculated Variables
Section Number: 11
Question Number: 5
Column: 2015
Type of Variable: Num
SAS Variable Name: _RFDRHV5
Question Prologue:
Question: Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week)

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	No Notes: SEX = 1 and _DRNKWEK <= 14 or SEX = 2 and _DRNKWEK <= 7 or ALCDAY5 = 888	429,872	88.40	86.65
2	Yes Notes: SEX = 1 and _DRNKWEK > 14 or SEX = 2 and _DRNKWEK > 7	27,120	5.58	5.89
9	Don't know/Refused/Missing Notes: _DRNKWEK = 99900	29,311	6.03	7.47

-Current Smoke Status

This is variable is further grouped by us into two groups:

Current Smoker = ('_RFDRHV5' =1 AND '_RFDRHV5' =2)

Not Current Smoker = ('_RFDRHV5' =1 AND '_RFDRHV5' =2)

Data source:
[2016 survey data from BRFSS]

Mapping from Data to Visual Elements

All the data represented are categorized by the number of days of that the person feels mentally not good in 30 days. For the first two plots (overview for the U.S. and heatmap for age vs. sleep), the color-scale domain is 1-30 days, mapped to a range blue colors from light to dark. Since the maps are relatively large, they are able to convey more information.
For the third plot (for gender comparisons and smoking/drinking habit comparison), we categorized the days into 5 bins (1-6 days, 7-12 days, 13-18 days, 19-24 days and 25-30 days), since it leads to a more obvious comparison. The percentage of each category of that certain group of people is plotted using stacked bar, so that the data is presented by height of bars. Also, icons is used for this section as container, so that it may be more interesting and easier understood.

The icons are generated by group members using online tools. So that the style design is most consistent throughout the whole page.

The Story

#1 Gird Map

We choose to use a U.S. map to show the average number of days that respondents are not in good mental condition within 30 days per state, so that we can have a general overall of mental health condition in each states. We use the grid map to show the number regardless of size and population of states. Single-hue color progressions is used to map the data. The darkest hue represents the greatest number of days and the lightest shade representing the least number.

It is interesting to find that New York State have the best mental condition, which is 8 days. Wisconsin has the worst mental condition, which is 14 days. On the side, we add to facts. The average #days not in good mental status is 11 days and 32.6% people have more than 15 days they feel not mentally good.

#2 Heatmap

In the second graph we use the heat map to explore the influence of sleep hours and age together on mental health. Again single-hue color progressions is used to map the data, with the darkest hue representing the greatest number and the lightest shade representing the least number.

With the increase of sleep hours, #days not in good mental condition decreases and then increases. People who sleep 7 hours a day have the best mental health condition.It is surprising to find that mental health condition does not have a strong relation with age. Within the same sleep duration, people under 23 years old is slightly in a better mental condition.

#3 #4 Gender & Habit

In addition to exploring the influence of sleep hours and aging, we are also interested in the influence of gender and habit (smoking & heavy drinking). Again single-hue color progressions is used to map the data.

Initially we assumed that there might be some difference for male and female, but it turned out that there is no difference, with approx. 52% of people feeling mentally not good in less than 7 out of 30 days, and approx. 20% of people feeling bad in more than 24 days our of 30 days.

The usage of alcohol surprisingly does not have a strong correlation to people's mental health, However, the usage of tobacco shows a significant negative impact, as the percentage of more than 24 days group rises to 30% for current smokers.