

NHANES dataset report

2025-09-27

Contents

NHANES dataset	1
Dataset exploration	1
Summary statistics of each wave	1
Information on gender of each wave	2
Information on ethnicity of each wave	2
Summary statistics of systolic blood pressure by wave	3
References	4

NHANES dataset

NHANES datasets used is a combined dataset between NHANES survey results that includes health exams and lab tests of all ages for systolic and diastolic blood pressure measurements and demographics data including age, gender, ethnicity, income, education and marital status. (Health Statistics 2015. 2016. 2017)

The measurements were measured in four different times across different waves (2015, 2016, and 2017). For current analysis, the following variables will be examined for initial exploration process: gender, ethnicity, and systolic blood pressure.

Noted that for ethnicity, ethnicity_2 variable will be included as the only ethnicity variable for the dataset as the variable includes more detailed classification.

Dataset exploration

The following variables and summary statistics are conducted.

Summary statistics of each wave

In total, there are three waves 2015, 2016, and 2017; with sample size differed between the waves 10,175, 9971, 9254 respectively based on the table below. Noted that the percentage of the total sample is the highest for 2015, and decrease in 2016 and 2017 with 2017 has the lowest sample size.

Table 1: Summary on sample size and proportion across waves

Wave	Sample size	Percentage % total sample
2015	10175	34.60884
2016	9971	33.91497
2017	9254	31.47619

Information on gender of each wave

In addition to analysis on the sample size, we conducted analysis on the proportion of male and female observations in the dataset. From the table below, the number of observations of male slightly decrease from 2015 to 2017; however, the percentage of male from each wave shows a different trend where male proportion decrease in 2016 and increase by 0.18% to 49.24% in 2017.

Table 2: Sample size and percentage by gender across waves

Wave	Gender	Sample size	Percentage of wave
2015	Female	5172	50.83047
2015	Male	5003	49.16953
2016	Female	5079	50.93772
2016	Male	4892	49.06228
2017	Female	4697	50.75643
2017	Male	4557	49.24357

Information on ethnicity of each wave

Ethnicities available in the dataset include White, Black, Mexican, Asian, Other hispanic, and Other. The table below shows the sample size of each ethnicity by wave and their associated proportions. Based on the data below, we can see that White samples across three years have the highest number of observations; noted that the proportion decreased in 2016 and increased again in 2017. The table below also shows a consistent trend in the number of samples from different ethnicities where the sample size is in the following order: White, Black, Mexican, Asian and Other Hispanic switched ranking for 2016, and Other ethnicity.

Table 3: Sample size and percentage of ethnicity across waves

Wave	Ethnicity	Sample size	Percentage of wave
2015			
2015	White	3674	36.108108
2015	Black	2267	22.280098
2015	Mexican	1730	17.002457
2015	Asian	1074	10.555283
2015	Other Hispanic	960	9.434889
2015	Other	470	4.619165
2016			
2016	White	3066	30.749173
2016	Black	2129	21.351921
2016	Mexican	1921	19.265871
2016	Other Hispanic	1308	13.118042
2016	Asian	1042	10.450306
2016	Other	505	5.064688
2017			
2017	White	3150	34.039334
2017	Black	2115	22.854982
2017	Mexican	1367	14.771990
2017	Asian	1168	12.621569
2017	Other Hispanic	820	8.861033
2017	Other	634	6.851091

Summary statistics of systolic blood pressure by wave

In order to analyse the average systolic blood pressure, we need to calculate the average from the four measurements of systolic blood pressure for each participant. The calculation is based on the average of available systolic blood pressure data, the average is missing if all measurements are missing. Please refer to the code below for the calculation of average systolic blood pressure.

```

nhanes_avg_bp <- nhanes_data %>%
  mutate(
    avg_sbp = rowMeans(
      select(., systolic_bp_1, systolic_bp_2, systolic_bp_3, systolic_bp_4),
      na.rm = TRUE
    ),
    avg_sbp = if_else(
      is.na(systolic_bp_1) & is.na(systolic_bp_2) & is.na(systolic_bp_3) & is.na(systolic_bp_4),
      NA_real_,
      avg_sbp
    )
  ) %>%
  select(patient_id, wave_id, avg_sbp)

```

Subsequently, a summary statistics table of average systolic blood pressure by wave as follows:

Table 4: Summary statistics of average systolic blood pressure by wave

Wave	Minimum avg systolic bp	Maximum avg systolic bp	Median avg systolic bp	Mean avg systolic bp	Standard deviation avg systolic bp
2015	64.67	228.67	115.33	118.32	18.06
2016	74.00	231.33	117.33	120.40	18.40
2017	72.67	234.00	118.00	121.66	20.28

In conclusion, based on the table above of the average systolic blood pressure, we observed there is an increase in the median, mean and standard deviation of average systolic blood pressure from 2015 to 2017. In addition, we noted that 2015 has the biggest sample size, and the sample size decreased in 2016 and 2017.

References

Health Statistics, National Center for. 2015. 2016. 2017. “National Health and Nutrition Examination Survey 2015-2016-2017.” <https://wwwn.cdc.gov/nchs/nhanes/Default.aspx>.