

Mini-Project: Do Fitness Habits and Perceptions of Health Change With Age?

Overview and student objectives

In-Class Activity Length: 25 minutes

Overview

In this mini-project, students will use data collected by the U.S. National Center for Health Statistics through the 2011-2012 National Health and Nutrition Examination Survey (NHANES). The data collected from the survey include demographics, various body and health measurements, and information about various lifestyle choices. The NHANES is unique in that the health measurements are collected through physical examinations. Data on all other variables are self-reported.

The primary objective of this analysis is to explore associations between age, health perception, and regular physical activity. Students will work with a dataset that includes 6,033 adults, aged 20 years and older. This mini-project will focus on the following three variables: *Age*, *HealthGen*, and *PhysActive*. Students will use various graphical displays to explore associations between these variables.

Objectives

Students will understand:

- Graphical displays can be used to communicate information about a single variable or the relationship between two or more variables.
- Creating effective graphical displays is typically an iterative process.

Students will be able to:

- Create an appropriate display to visualize the distribution of a single variable (univariate graphs) or the relationship between two or more variables (multivariable graphs).
- Write a short report to communicate statistical results to a general audience.
- Apply principles of making effective graphical displays to improve an existing graph.

Suggested resources and preparation

Materials and technology

- Computer, projector, document camera
- Student Pages for In-Class Activity
- Rubric for grading written formal papers; see the attached sample or choose another rubric from www.exemplars.com or other websites
- Excel spreadsheet: [DCMP_STAT_5D_nhanes_alldata](#)
- Excel spreadsheet: [DCMP_STAT_5D_nhanesYesNo](#)

Prerequisite assumptions

Students should be able to identify different variable types, choose an appropriate graphical display, use technology to create a graphical display, and interpret a graph in the context of the data.

Making connections

This activity connects to exploring data visually.

Background context

- The background on the full dataset is provided at <https://cran.r-project.org/web/packages/NHANES/NHANES.pdf>.
- The data used for this mini-project includes results from the 2011-2012 NHANES for 6,033 adults, aged 20 years and older. Each respondent included in this dataset reported or measured values for all variables in the analysis.
- This dataset was modified for educational purposes and should not be used for research or publications.
- See <https://www.cdc.gov/nchs/nhanes/> for more information about the survey and to obtain original data suitable for research.

Suggested instructional plan

Frame the activity (15 minutes)

Structure	Instructor Suggestions
Whole Class Discussion	<p>Questions 1–4</p> <ul style="list-style-type: none"> • The goal of these questions is to help students prepare for the mini-project, which includes more open-ended instruction than seen in previous activities and assignments. Students will use the same dataset and a similar rubric for all three projects. <p>Question 1</p> <ul style="list-style-type: none"> • The primary goal of this question is to ensure that all students are able to access the dataset and data dictionary. It also helps students review categorical vs. quantitative variables. <ul style="list-style-type: none"> ○ Part of this question could include a demonstration on how to access and download the dataset.
Whole Class Discussion	<ul style="list-style-type: none"> • Many of the quantitative variables are health measurements (<i>BPSysAve</i>, <i>TotChol</i>, <i>BMI</i>). If time permits, give students a few minutes to look up these measurements online and discuss what the units mean and what the measurements indicate about a person's general health.

Question 2

- The goal of this question is to make sure all students have identified the appropriate technology required to complete the project. Give students a few minutes to think about this question individually before discussing it as a group.
- Briefly discuss or demonstrate, if needed, how to enter data into the data analysis tool.
 - Enter Data: Select “Your Own.”
 - Copy and paste the values of the variables of interest in the observation input.

Question 3

- The primary goal of this question is to introduce students to the grading rubric and ensure that all students are aware of the expectations for the project.
- Show students where they can access the grading rubric, and then give them a few minutes to read the rubric and write their responses to Question 3.
- Discuss the responses as a class.
- This is a good opportunity to ask students if they have any questions about the grading rubric or expectations for the project.
- Give students a few minutes to read the remaining project instructions and to ask questions. If there is not enough time to do this in class, you can assign this as homework and answer questions in the next class period.

Wrap-up

Questions 5-7

- Students will complete the remaining questions outside of class or in the typical venue for completing homework.

Suggested assessment, assignments, and reflections

- Assign the mini-project at the end of the Student Pages.

Rubric for Group Project 5.D

Name: _____

Date: _____

Question 4	4	The graphical display is appropriate, and the interpretation is complete and correct. A complete description includes a discussion of shape, center, spread, and the presence of outliers.	
	3	The graphical display is appropriate, but there are some minor errors in the interpretation.	
	2	The graphical display is not appropriate, but the interpretation is consistent with the graph OR the graphical display is appropriate, but there are major errors in the interpretation.	
	1	The graphical display is not appropriate, and there are major errors in the interpretation.	
Comments:			
Question 5	4	The graphical display is appropriate, categories are presented in a logical order to improve understanding, and the interpretation is complete and correct.	
	3	The graphical display is appropriate, but there are some minor errors in the interpretation or ordering of the categories in the graph.	
	2	The graphical display is not appropriate, but the interpretation is consistent with the graph OR the graphical display is appropriate, but there are major errors in the interpretation.	
	1	The graphical display is not appropriate, and there are major errors in the interpretation.	
Comments:			
Question 6	4	The graphical display is appropriate and includes the necessary elements to compare the two variables. The interpretation is correct and uses details from the graph to support observations.	
	3	The graphical display is appropriate, but there are some minor errors in the interpretation.	
	2	The graphical display is not appropriate, but the interpretation is consistent with the graph OR the graphical display is appropriate, but there are major errors in the interpretation.	
	1	The graphical display is not appropriate, and there are major errors in the interpretation.	
Comments:			



Question 7	4	Issues are correctly identified in the original graph. The improved graph is appropriate and addresses the concerns identified in the original graph, and all three variables of interest are included and compared. If more than one graph is used, the graphs are displayed so appropriate comparisons can be made (e.g., side-by-side, same scale). The interpretations are correct.
	3	The improved graph is appropriate, but there are minor errors in the issues identified in the original graph or in the interpretations.
	2	The improved graph is not appropriate but correctly identifies issues with the original graph, and the interpretations are consistent with the graphs OR the graphical display is appropriate, but there are major errors in the interpretations.
	1	The improved graph is not appropriate, and there are major errors in the issues identified in the original graph and the interpretations.
Comments:		
Presentation	4	The graphical displays have informative titles and axis labels. The paper is neatly prepared with interpretations written in full sentences, as well as correct spelling and grammar.
	3	Most of the graphical displays have informative titles and axis labels. The paper is neatly prepared with interpretations written in full sentences and a few minor spelling or grammar errors.
	2	Fewer than half of the graphical displays have informative titles and axis labels. The paper is not neatly prepared, interpretations are generally not written in full sentences, or it contains several spelling or grammar errors.
	1	No graphical displays have informative titles and axis labels. The paper is not neatly prepared, few to no interpretations are written in full sentences, and it contains several spelling or grammar errors.
Comments:		
Total Points:		



Mini-Project: Do Fitness Habits and Perceptions of Health Change With Age?

In this mini-project, you will use data collected by the U.S. National Center for Health Statistics through the 2011-2012 National Health and Nutrition Examination Survey (NHANES). The NHANES has been conducted every two years since the early 1960s. The data collected from the survey include demographics, various body and health measurements, and information about various lifestyle choices. The NHANES is unique in that the health measurements are collected through physical examinations. Data on all other variables are self-reported.



Credit: iStock/Nattakorn Maneerat

Before you begin the analysis, we will discuss the data and expectations for this project.

Objectives for the mini-project

You will understand:

- Graphical displays can be used to communicate information about a single variable or the relationship between two or more variables.
- Creating effective graphical displays is typically an iterative process.

You will be able to:

- Create an appropriate display to visualize the distribution of a single variable (univariate graphs) or the relationship between two or more variables (multivariable graphs).
- Write a short report to communicate statistical results to a general audience.
- Apply principles of making effective graphical displays to improve an existing graph.

- 1) Locate the NHANES dataset ([DCMP_STAT_5D_nhanes_alldata](#)) containing the following variables:

Age: Age in years at screening of the study participant (Note: Subjects 80 years or older were recorded as 80)

HealthGen: Self-reported rating of the study participant's health in general (Excellent, Vgood, Good, Fair, Poor)

PhysActive: Whether the study participant reported performing moderate or vigorous sports, fitness, or recreational activities (Yes, No)

AttendCollege: Whether the study participant attended or completed college at the time of the study (Yes, No)

BadPhysHlthDay: Whether the study participant self-reported having at least one bad physical health day in the past 30 days (Yes = at least 1 day; No = 0 days)

BadMentHlthDay: Whether the study participant self-reported having at least one bad mental health day in the past 30 days (Yes = at least 1 day; No = 0 days)

BMI: Body mass index (weight/height² in kg/m²)

SleepTrouble: Participant told a doctor or other health professional that they had trouble sleeping (Yes, No)

SleepHrsNight: Self-reported number of sleep hours the study participant usually got on weekdays or workdays

TotChol: Total HDL cholesterol in mmol/L

BPSysAve: Average of three systolic blood pressure readings in mm/Hg

Part A: Identify two quantitative variables in the data. What possible values do these variables take in the dataset?

Part B: Identify two categorical variables in the data. List the possible categories these variables can take.

- 2) In this project, you will create univariate, bivariate, and multivariate graphical displays for quantitative and categorical variables. Identify one or two tools from *DCMP Data Analysis Tools* at <https://www.utdanacenter.org/our-work/higher-education/curricular-resources-higher-education/dcmp-data-analysis-tools> (or other tools) that you can use to create these graphs. Include in the answer the name of the tool and the type of graph you can create.

It is recommended that you use the DCMP Data Analysis tools to create the graphs.

Bar graphs and histograms: https://dcmpdatatools.utdanacenter.org/eda_quantitative/

Bar graphs and pie charts: https://dcmpdatatools.utdanacenter.org/eda_categorical/

- 3) One goal of this project is to continue developing your statistical writing skills. This includes writing your interpretations in a way that can be clearly understood by a general audience and presenting your results in a report suitable for an academic or professional setting.

Part A: Describe what is meant by “writing in a way that could be clearly understood by a general audience.”

Part B: Refer to the project rubric. Based on the rubric, what are qualities of a report that are suitable for an academic or professional setting?

The primary objective of this analysis is to explore associations between age, health perception, and regular physical activity. The data you'll use to investigate these associations include responses from 6,033 adults, aged 20 years and older.

This analysis focuses on the following variables:

Age: Age in years at screening of the study participant (Note: Subjects 80 years or older were recorded as 80)

HealthGen: Self-reported rating of the study participant's health in general (Excellent, Vgood, Good, Fair, Poor)

PhysActive: Whether the study participant reported performing moderate or vigorous sports, fitness, or recreational activities (Yes, No)

4) Let's begin by looking at the ages of the survey respondents.

Part A: Use technology to create a histogram of *Age*.

Part B: Describe the distribution. Include the shape, center, spread, and the presence of outliers in your description, using appropriate summary statistics as needed.

Part C: The distribution shows a small peak around 80 years old. (Note: Changing the binwidth of the histogram makes the peak more or less noticeable.) Briefly explain why there may be a peak around this value.

5) Next let's examine how the respondents generally perceived their health. Create an appropriate graphical display of *HealthGen*. Then use the graph to describe two different observations about the respondents' general perceptions of their health.

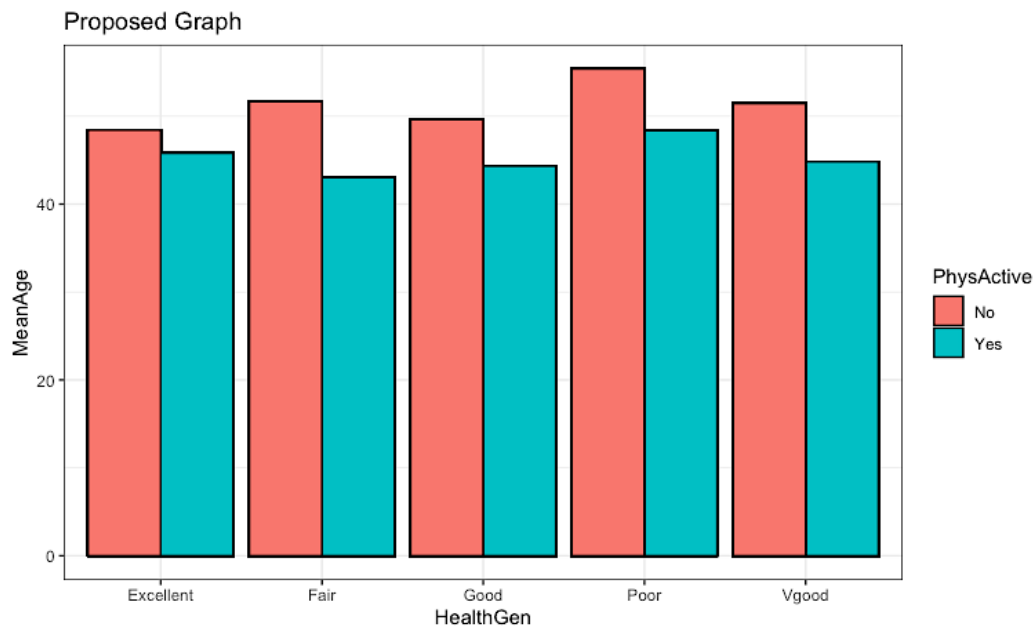
6) Now let's examine whether there is an association between health perception and whether a person performs regular physical activity. Create a graphical display to visualize the distribution of *HealthGen* for each category of *PhysActive*.

Does there appear to be a difference in general health perception between people who perform regular physical activity and those who do not? Write two observations from the graphical display to support your response.

7) A popular health and wellness website is writing an article that explores the following questions:

- (1) Does having a habit of regular physical activity change with age?
- (2) Does the association between age and health perception differ based on performing regular physical activity?

They would like to include a graphical display in the article to help readers visualize conclusions in the article about the associations between the variables. The authors propose the following graph, but the website editor is concerned it may be confusing to readers. She has asked for your help to improve the graph and write an interpretation of the graph that will be included in the article.



Part A: Describe two ways this graphical display may be confusing and/or makes it difficult to explore the questions of interest for the article.

Part B: Use technology to make an improved graphical display that can be effectively used to answer the two questions of interest. You may use multiple graphs, if needed.

In statistics, we often need to manipulate the structure of the dataset in order to create the visualizations. Locate the **reorganized** dataset ([DCMP_STAT_5D_nhanesYesNo](#)) that includes only the variables for this question that aid in the creation of your visual display.

Part C: Use the graph(s) to answer the two questions of interest for the article. Write a short paragraph (three to six sentences) that includes your answers to the questions and observations from the graph(s) that support your response. The response should be written in a way that is clearly understood by a general audience.