AP COMPUTER SCIENCE PRINCIPLES
Civic Knowledge & Action: Voter Registration

TEACHER RESOURCE

Overview

Computing enables and empowers new methods of information processing that have led to monumental change across disciplines, from art, to business, to science, to government. Managing and interpreting available data to answer questions, tell a story, and solve problems is foundational to our society and economy. Computers consume data, transform data, and produce new data, allowing users to create new information and knowledge to solve problems through the interpretation of those data.

In this lesson, students will use programs to process real voter registration and turnout data to discover information and create new knowledge. Students will be guided to identify questions and use a program to process voter registration and turnout data to answer these questions, tell a story, or solve a problem.

Connections to the AP Computer Science Principles Course Framework

Part 1: Identifying Questions to Answer with Data (25 minutes)

- Students will identify a list of non-trivial questions they can answer through examining census and voter turnout data.

Enduring Understanding	Skill	Learning Objective	Topic
DAT-2	5.B	DAT-2.A	2.3

Part 2: Selecting Data Sets to Tell a Story (45 minutes)

- Students will identify the data they will need to investigate to answer 2–3 questions and tell a story or solve a problem.

Enduring Understanding	Skill	Learning Objective	Topic
DAT-2	5.B	DAT-2.A	2.3
DAT-2	5.D	DAT-2.C	2.3

Part 3: Using Spreadsheet Software to Extract Information from Data (45 minutes)

- Students will transform, filter, and combine data.
- Students will create visualizations from data.

Enduring Understanding	Skill	Learning Objective	Topic
DAT-2	2.B	DAT-2.D	2.4

Part 4: Tell Your Data story (45 minutes)

- Students will extract information from data to answer their questions and tell a story with their data.

Enduring Understanding	Skill	Learning Objective	Topic
DAT-2	2.B	DAT-2.D, DAT-2.E	2.4

Before You Start

This lesson provides students the opportunity to extract information from real-world data related to voter registration and turnout. It covers Topics 2.3 and 2.4 from the AP Computer Science Principles Course and Exam Description. This lesson provides scaffolding as an introduction to the learning objectives or as a culminating project for these topics. To fully cover Topic 2.3, teachers will need to supplement this lesson to cover Learning Objective DAT-2.B.

ACTIVITY

Part 1 Identifying Questions to Answer with Data

Students will identify a list of non-trivial questions they can answer through examining census and voter turnout data.

TIME:
25 minutes

- Take a few minutes to brainstorm as a group about what questions or concerns students have about the upcoming local or federal election(s).

- Let students know that some of their questions can be answered by looking at publicly available census and voter registration and turnout data.

- Distribute the **Identifying Questions about Voting in the US** Student Handout.

- Students will need to develop non-trivial questions that will provide a focus for their investigation and should reflect the rationale and overall direction. For example, a student might propose the question, "In my state, how many registered voters are there between the ages of 18 and 25?" While this question can certainly be easily answered using the data, it doesn't allow us to tell a story or solve a problem related to voting. A better question that provides a direction for their investigation would be, "In my state, what demographic groups are underrepresented as registered voters and would be best to target in a voter registration campaign?"

- Review the summary of available census and voter registration data provided in the **Identifying Questions about Voting in the US Student Handout**.

- Have students work in groups of 3–4 to brainstorm questions about voting and election outcomes. They should use the available data guide their question development. Each group should generate 5 questions they think could be answered using the data provided.

- Have students write a first draft of the questions in the first column, and then as needed, revise questions to provide a focus, rationale, and overall direction for their investigation.

Virtual Teaching Tip
Some video conferencing platforms include a "breakout room" feature that can be used to create multiple video meeting spaces within a single video call. If enabled, use this feature to place students into collaborative groups throughout this activity. Groups can be determined randomly or manually. Provide students the opportunity to interact with the entire class as well as their collaborative group by moving students between the main video call and their breakout room throughout the activity.

Instructional Strategy: Discussion Group
This instructional strategy allows students to gain new understanding of or insight into a problem or solution by listening to multiple perspectives. In this activity, students are gaining insight into the questions students with different perspectives have about voting.

- Once each group has developed their list of questions, have them share their questions with the class to develop a master list of potential questions. As necessary and as a whole class, work to further revise questions to ensure they are non-trivial and meaningful to understanding voter registration issues.

Before Moving On

Be sure that students can modify questions to be non-trivial which allow them to tell a meaningful story or solve a problem with the data. Students may fall back on what is easy, such as figuring out the state with the most voting districts, but this one piece of information does not tell a meaningful story about the data.

Part 2 Selecting Data Sets to Tell a Story

Students will identify the data they will need to investigate to answer 2–3 questions and tell a story or solve a problem.

TIME:

45 minutes

- Provide students with **Student Handout 2: Investigating Data Sets** Student Handout. This handout contains the following list of resources that contain data files related to the census and voter registration. Most links provide Excel files, however the 2016 presidential election web site requires additional steps to obtain the data in an Excel file.
    - For each district in a state, the number of votes cast in the congressional election: **https://www2.census.gov/programs-surveys/demo/tables/voting/table01.xlsx**
    - For each district in a state, the number of residents who are of voting age, and the voting rate: **https://www2.census.gov/programs-surveys/demo/tables/voting/table02a.xlsx**
    - For each district in a state, the estimated number of voters by age, gender, and poverty level: **https://www2.census.gov/programs-surveys/demo/tables/voting/table02b.xlsx**
    - For each district in a state, the education level of voters: **https://www2.census.gov/programs-surveys/demo/tables/voting/table02c.xlsx**
    - For each county in a state during the 2016 presidential election, the party affiliation, education level, wage earnings, ethnicity, and age of voters as well as the relative health, crime rate, and weather of districts: **https://public.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county/table/?disjunctive.state**

- Have students explore several of the available data sources. The first four data files open directly in Excel. Students will open each set and provide a summary of what data is being provided. As students examine the data, they should begin to think about ways the data can help in answering the previously developed questions.

- The data provided in these tables have already been cleaned. Review with students what it means for data to be clean and how the way in which data are collected might influence the amount of data cleaning that would need to happen before the data would be usable in an investigation. For example, if a form had an

open field for a user to identify the state they live in, the user could enter "New York", "NY", "ny", or "N.Y.".

- Use the *Think-Pair-Share* instructional strategy and have students focus on one of the data sets and identify ways the data in the table represent clean data. Provide students with one minute to think individually about the ways the data are clean then two minutes to talk with another student, and then share as a whole group.

- Brainstorm with students ways we can make the collection of data more uniform. Lead a discussion about forms students have had to fill out that contain their address and how these forms have managed uniform data collection by using drop-down menus for the state.

- Assist students with downloading the 2016 presidential data. Have students go to the 2016 Presidential election data site. To download the data table on this site, students should first select **Export**.

- Students will have file format options to select for the data. Since these lessons are focused around using spreadsheet software, have students select the **Excel** versions.

Instructional Strategy: Think-Pair-Share

In this strategy, students think through a problem alone, pair with a partner to share ideas, and then share results with the class. This strategy enables the development of initial ideas that are then tested with a partner in preparation for revising and then sharing the ideas with a larger group.

- Students can also pre-filter this data by state and county by clicking on the filter on the left side of the screen. This will give them an option of downloading a smaller Excel file.

- Have students look at the 2016 presidential election data and determine if they think the data are clean and explain their reasoning.

- While the data in the 2016 presidential election file have already been cleaned and validated, this is a large data set that can make processing the data more difficult and sometimes not possible on a single computer. Lead a discussion with students on how the size of the data set might affect our ability to investigate the data. Explain to students some of the challenges that exist based on the size of the data. Be sure to also address the challenge of when we have insufficient data as well as large data sets, such as the 2016 presidential election.

- Using the *Think-Pair-Share* instructional strategy, have students consider ways in which they can reduce the amount of data in the 2016 presidential election file to make it more useful. Some ways of reducing, such as deleting rows, could make the data incomplete and could lead to inaccurate conclusions.

- For this data file to be useful for students, they will need to first identify what data values need to be investigated and then copy these columns into another sheet. Another option would be to remove any column(s) they will not be used in their analysis.

- In the 2016 presidential election file, voting information for the state of Alaska is not available. Have students scroll through the data set and see if they can find the state without any information in the *Votes* column. Lead a discussion about how missing voting information for Alaska can lead to inaccurate conclusion during our analysis and how this might impact Alaska.

- Introduce students to how biases can be introduced into the data. Bias can be created by the type or source of data being collected. Bias is not eliminated by simply collecting more data.

- Have a group discussion about ways that our analysis could be biased by making assumptions about the data that we choose to analyze. For example, the analysis can be biased if we make conclusions about the entire country based on patterns or demographics within a particular state such as high poverty rates, racial homogeneity, or education level.

- Provide students with **Student Handout 3: Selecting Questions and Data Sets** Student Handout. Depending on the amount of class time available, students can complete this handout outside of class.

- In their groups and after having time to further explore the data available, have students select 2–3 questions from the collaborative class list of questions. Their questions should be related and meaningful to them so they can be used as the basis to tell a story or solve a problem with the data. The questions the group selects do not need to be ones that the group originally developed.

- From the list of Excel files, have students select both the data files and columns in the data files they need to answer their questions and tell a meaningful data story or solve a problem.

- For some questions that the students pose, they may need to investigate additional data sets outside of the ones provided in this lesson. This may require additional time to locate the data needed.

Activity Variation

The 2016 presidential election data file is being hosted on a website with an interactive data analysis tool that will allow users to filter data and create graphs. This website contains data sets on other topics that students might find interesting as well. Instruction for using this data tool is not supported in these lesson plans, so its use will need to be explored on your own.

Before Moving On

Ensure that the focus questions each group has selected are related so that students are able to tell one cohesive story during their project rather than just presenting 2–3 facts about voting.

Part 3 Using Spreadsheet Software to Extract Information from Data

Students will transform, filter, and combine data sets. Students will create visualizations from data.

TIME:
45 minutes

When modeling how to format and process data and create charts, the level of detail provided to students should be tailored to their experience level. An appendix is provided that walks through each of the steps described, if additional guidance is needed.

- Have students open the data file they will be using.

  – Depending on your students' experience with Excel and your own comfort level, you can decide on a standard data set the entire class will use or have students use their own data set. If students have little to no experience using Excel, it might be worth having them all use the same data set so that you can ensure they are following along and are comfortable with each step before asking them to work on their own data set. However, if students are

Virtual Teaching Tip

This activity is designed to be completed using Excel. However, teachers who have experience with analyzing data and creating non-textual representations using Google Sheets can modify this activity to support their students' data exploration using Google Sheets instead.

proficient in Excel, allowing them to use their own data set while walking through the process will allow them to become more familiar with the data and start their investigation sooner.

- Have students turn the data (excluding header rows) into a table. This will allow for easier advanced filtering. See "Turning Data Into a Table" in the Appendix.

- After students have the data in table format, have them select a column and filter the data by state.

- Once a filter has been applied, have students practice using formulas to perform at least two calculations, such as finding the sum, average, or count of filtered data. See "Basic Filtering and Formulas" in the Appendix.

- Once students are comfortable applying basic filters, have them practice filtering a column based on a combination of filters, or using one of the built-in filtering features. Have students apply a formula to a column where not all cells are in contiguous rows to ensure that they know how to select only the cells visible. See "Advanced Filtering" in the Appendix.

- Whether students are using their chosen data set or a standard data set, have them think about what type of filtering or formulas might be needed to answer the questions they selected during the last lesson, and ensure that they are comfortable using them before moving on.

- Before having students create charts of their data, it is important that the data are formatted appropriately. This will allow students time to focus on formatting the chart instead of continually moving data around. See "Formatting Data" in the Appendix.

- Once the data have been properly formatted, have students create several different charts of the data. Pair students and have each student share with their partner two charts that they created and ask the partner which one they feel is most readable and why. A chart is considered readable if the trend or characteristic students are trying to convey is readily apparent. Trends and characteristics that are noticeable in one type of chart might be different in another chart. See "Creating a Chart from Data" in the Appendix.

- Have each student pick their favorite chart and properly format it, including adding a title, formatting the axes, and adding at least one additional chart element such as data labels. See "Formatting a Chart" in the Appendix.

Virtural Teaching Tip

Have students share a picture of their chart and a brief 1–2 sentence summary of what is being shown in the chart on a virtual wall such as **Padlet www.padlet.com**. This way students will have more examples to pull from when choosing the chart(s) they will use in their final story.

Before Moving On

Investigate each student's formatted chart as well as the filters and formulas used in the creation of the chart to ensure that every student can create a chart from filtered and processed data. If students need additional help or guidance on creating or formatting the chart, provide additional feedback and support before they begin Part 4.

Part 4 Tell Your Data Story

Students will extract information from data to answer their questions and tell a story with their data.

TIME:
45 minutes

- In groups, have students apply computational tools and techniques to answer their questions identified in Part 2 (e.g., by finding patterns in the data, by transforming data, or by finding connections between the data and other sources of knowledge).

- After processing the data, group students and have them apply computational tools and techniques to create non-textual representations (i.e., diagrams, charts, graphs, or other visualizations) that communicate the results of their investigation.

- Provide students with **Student Handout 4: Telling a Story with Data**. Students will complete the first part of this handout in groups will be completed in groups. Students will complete the second part individually.

- When completing the activities in the handout, students will use what they learned in their investigation to synthesize answers to their questions and use the data to tell a story. A *data story* builds a narrative around the data and its visualizations. This narrative could be used to educate someone on an issue, sway opinions, or present a solution to a problem.

- Have students present their data story. This can be accomplished in several ways depending on your time constraints. For example, students might:
  - create a short video and share on your LMS
  - present to family members
  - post their questions and findings to a class discussion board
  - present to another class or after-school club
  - create an elevator speech to encourage people to register to vote and include relevant data and their findings.

- (Optional) This project may inspire some students to get more involved in the voter registration process by helping their classmates and others in their community to register to vote. If students want to be more involved, they can register with DoSomething: **https://www.dosomething.org/ckap2** to run a voter registration drive.

✎ **Teacher Notes**

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Identifying Questions about Voting in the U.S.

To have a meaningful data investigation, the questions we pose need to be non-trivial. These questions should reflect the rationale and overall direction for your investigation.

Below is an example question that is trivial because it doesn't really provide a rationale or direction for why we are conducting our investigation. Revise this question to be more meaningful.

Question	Revised Question
In my state, how many registered voters are there between the ages of 18 and 25?	

We can use publicly available census and voter turnout data to better understand the current trends in voting to answer our questions and solve problems. Below is some of the data that is available:

- For each district in a state, the number of votes cast in the congressional election, the number of residents who are of voting age, and the voting rate.
- For each district in a state, the estimated number of voters by age, gender, and poverty level.
- For each district in a state, the education level of voters.
- For each county in a state during the 2016 presidential election, the party affiliation, education level, wage earnings, ethnicity, and age of voters.
- For each county in a state during the 2016 presidential election, the relative health, crime rate, and weather of districts.

In small groups, develop a list of 5 questions. Record your questions in the first column in the table on the next page. After you have done first round of brainstorming questions, look at each question and revise it to ensure that you have included a rationale and direction to focus your investigation. Put your revised questions in the second column of the table.

Question	Revised Question
1.	
2.	
3.	
4.	
5.	

Investigating Data Sets

In the table below, there are data sources related to the U.S. census, voter registration, and voter turnout. Open each data source and provide a summary of what data are available. Add your own notes on how you would use the data to answer one of the questions and tell a story or solve a problem.

Data Source	Summary of Data Available	Notes
https://www2.census.gov/programs-surveys/demo/tables/voting/table01.xlsx		
https://www2.census.gov/programs-surveys/demo/tables/voting/table02a.xlsx		
https://www2.census.gov/programs-surveys/demo/tables/voting/table02b.xlsx		
https://www2.census.gov/programs-surveys/demo/tables/voting/table02c.xlsx		

These data tables have already been cleaned and are ready to use. *Cleaning data* is a process that makes the data uniform without changing their meaning (e.g., replacing all equivalent abbreviations, spellings, and capitalizations with the same word).

1. Select one of the data sources from the table. Identify ways the data represent clean data.

2. What are some techniques that can be used to ensure that data are more uniform during the collection process to limit the amount of data cleaning that is necessary before the data are usable?

The 2016 presidential election data are available at the link below. The data can be explored using the interactive platform provided as well as downloaded for more extensive investigations.
https://public.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county/table/?disjunctive.state

Download and open the 2016 presidential election data in Excel.

3. Are the data provided clean? Explain your reasoning.

This file is much larger than the first few data files we investigated.

4. How does the size of the data set affect our ability to investigate and extract information from the data?

5. In what ways can we reduce the amount of data in the 2016 presidential election file to make it more useful?

6. In what ways might we reduce the amount of data in the 2016 presidential election file and lead to inaccurate conclusions during our investigation?

7. Another challenge associated with processing data is when the data are incomplete. For the 2016 presidential election file, scroll down the *Votes* column. For what state are there missing data in the Votes column?

Bias is a prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair. When working with data, bias can mean a disproportionate weight in favor of or against an idea or thing.

Problems of bias are often created by the type or source of data being collected. Bias is not eliminated by simply collecting more data.

8. Identify one assumption that would cause our analysis to be biased.

Selecting Questions and Data Sets

Now that you are more familiar with the data available for your investigation, select 2–3 questions from the collaborative class list of questions. These questions should be related and meaningful to you so that they can be used as the basis to tell a story or solve a problem through your data investigation. The questions you select do not have to have been developed by your group.

In the chart below:

- List the questions in the left column.
- List the data set you will use in your investigation in the middle column.
- List the data columns that will be important for your investigation in the right column.

Question	Data Sets	Column(s) in the Data Set
1.		
2.		
3.		

Telling a Story with Data

As a group, respond directly to each of the following prompts. This report will provide the details of your investigation and should include non-textual representations that communicate the results of your investigation. These representations could include charts, tables, graphs, visualizations, or other appropriate resources that augment or clarify your questions, analysis, and answers.

1. Describe the area of focus your group chose to explore and explain why you chose this area.

   ....

   ....

2. Identify the 2–3 questions that are related to your area of focus. Clearly present the answer to each of the questions you investigated, supporting each answer with your data investigation and appropriate non-textual representations. Include a one-sentence description of each non-textual attachment.

   ....

   ....

On your own, respond directly to each of the following prompts.

3. Write a paragraph that goes beyond the answer to your questions and demonstrates your discovery of information or creation of new knowledge to tell a data story. A *data story* builds a narrative around the data and their visualizations. This narrative could be used to educate someone on an issue, sway opinions, or present a solution to a problem.

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

4. Describe the specific techniques (filtering, visualizing, analyzing, etc.) your group used. Provide a detailed description of how your group processed the information in the data set(s) to conduct the investigation and how this process enabled you to meet your objective of discovering new information or creating new knowledge. This description should be detailed enough that someone could recreate it. Some things to consider in your response:

    a. Did the data require any filtering? If so, how did you filter the data?

    b. Did you transform the data to determine patterns? If so, how?

    c. What type of non-textual representation(s) did you use? How did you create them?

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Identifying Questions about Voting in the U.S.

To have a meaningful data investigation, the questions we pose need to be non-trivial. These questions should reflect the rationale and overall direction for your investigation.

Below is an example question that is trivial because it doesn't really provide a rationale or direction for why we are conducting our investigation. Revise this question to be more meaningful.

Question	Revised Question
In my state, how many registered voters are there between the ages of 18 and 25?	(Sample Student Response) *In my state, what demographic groups are underrepresented as registered voters and would be best to target in a voter registration campaign?*

We can use publicly available census and voter turnout data to better understand the current trends in voting to answer our questions and solve problems. Below is some of the data that is available:

- For each district in a state, the number of votes cast in the congressional election, the number of residents who are of voting age, and the voting rate
- For each district in a state, the estimated number of voters by age, gender, and poverty level.
- For each district in a state, the education level of voters
- For each county in a state during the 2016 Presidential election, the party affiliation, education level, wage earnings, ethnicity, and age of voters
- For each county in a state during the 2016 Presidential election, the relative health, crime rate, and weather of districts

In small groups, develop a list of 5 questions. Record your questions in the first column of the table on the next page. After you have done this first round of brainstorming questions, look at each question and revise it to ensure that you have included a rationale and direction to focus your investigation. Put your revised questions in the second column of the table.

Question	Revised Question
1. Answers will vary.	Answers will vary.
2. Answers will vary.	Answers will vary.
3. Answers will vary.	Answers will vary.
4. Answers will vary.	Answers will vary.
5. Answers will vary.	Answers will vary.

Investigating Data Sets

In the table below, there are data sources related to the U.S. census, voter registration, and voter turnout. Open each data source and provide a summary of what data are available. Add your own notes on how you would use the data to answer one of the questions and tell a story or solve a problem.

Data Source	Summary of Data Available	Notes
https://www2.census.gov/programs-surveys/demo/tables/voting/table01.xlsx	For each district in a state, the number of votes cast in the congressional election.	Answers will vary.
https://www2.census.gov/programs-surveys/demo/tables/voting/table02a.xlsx	For each district in a state, the number of residents who are of voting age, and the voting rate.	Answers will vary.
https://www2.census.gov/programs-surveys/demo/tables/voting/table02b.xlsx	For each district in a state, the estimated number of voters by gender and poverty level.	Answers will vary.
https://www2.census.gov/programs-surveys/demo/tables/voting/table02c.xlsx	For each district in a state, the education level of voters.	Answers will vary.

These data tables have already been cleaned and are ready to use. *Cleaning data* is a process that makes the data uniform without changing their meaning (e.g., replacing all equivalent abbreviations, spellings, and capitalizations with the same word).

1. Select one of the data sources from the table. Identify ways that data represent clean data.

All state abbreviations are formatted the same with both letters capitalized. Data that represent a count are all integers and data that represent percentages are all represented as decimal numbers.

2. What are some techniques that can be used to ensure that data are more uniform during the collection process to limit the amount of data cleaning that is necessary before the data are usable?

When collecting data, if possible, use pre-populated drop-down menus or check boxes rather than having open fields for users to enter their own information.

The 2016 presidential election data are available at the link below. The data can be explored using the interactive platform provided as well as downloaded for more extensive investigations.
https://public.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county/table/?disjunctive.state

Download and open the 2016 presidential election data in Excel.

3. Are the data provided clean? Explain your reasoning.

Yes. The abbreviations and data formats are all the same.

This file is much larger than the first few data files we investigated.

4. How does the size of the data set affect our ability to investigate and extract information from the data?

If the data set is too small, a complete and accurate answer will be difficult to find. It is also more likely that the answer is biased based on what data are available if it doesn't give a complete picture. If the data set is too large, processing the data on a single computer may be difficult or impossible.

5. In what ways can we reduce the amount of data in the 2016 presidential election file to make it more useful?

We can remove the columns that are not necessary for our investigation. This can be accomplished by either deleting unnecessary columns or copying the necessary columns into a new sheet.

6. In what ways might we reduce the amount of data in the 2016 presidential election file and lead to inaccurate conclusions during our investigation?

Removing rows and essentially removing districts for a state would cause inaccurate conclusions to be drawn by only focusing on a few select districts rather than looking at the state or country as a whole. For some states or counties where the predominant number of votes is for the opposing candidate, you may be led to a different conclusion as to the result of the election by only looking at these states or counties.

7. Another challenge associated with processing data is when the data are incomplete. For the 2016 presidential election file, scroll down the Votes column. For what state are there missing data in the Votes column?

Alaska

Bias is a prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair. When working with data, bias can mean a disproportionate weight in favor of or against an idea or thing.

Problems of bias are often created by the type or source of data being collected. Bias is not eliminated by simply collecting more data.

8. Identify one assumption that would cause our analysis to be biased.

The analysis can be biased if we make conclusions about the entire country based on patterns or demographics within a particular state, such as high poverty rates, racial homogeneity, or education level.

Selecting Questions and Data Sets

Now that you are more familiar with the data available for your investigation, select 2–3 questions from the collaborative class list of questions. These questions should be related and meaningful to you so that they can be used as the basis to tell a story or solve a problem through your data investigation. The questions you select do not have to have been developed by your group.

In the chart below:
  – List the questions in the left column.
  – List the data set you will use in your investigation in the middle column.
  – List the data columns that will be important for your investigation in the right column.

Question	Data Set	Column(s) in the Data Set
1. Answers will vary. For example: In my state, what demographic groups are underrepresented as registered voters and would be best to target in a voter registration campaign?	Answers will vary. For example: **https://www2.census.gov/programs-surveys/demo/tables/voting/table02a.xlsx** **https://public.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county/table/?disjunctive.state**	Answers will vary. For example: Use the age estimate and percent of total in table 02; The Median age of voters in the 2016 Presidential election data.
2. Answers will vary.	Answers will vary.	Answers will vary.
3. Answers will vary.	Answers will vary.	Answers will vary.

Civic Knowledge & Action: Voter Registration

Telling a Story with Data: Teacher Scoring Guidelines

Telling a Store with Data: Scoring Guidelines and Notes			
Row / Learning Objective	Scoring Criteria	Decision Rules	Score & Notes
Row 1 DAT-2.A	The written response: ■ clearly describes the area of focus related to the data available ■ explains why this area of focus was chosen	**Do NOT award a point if any of the following is true:** ■ There is not a clear connection between the area of focus and the data available. ■ No explanation about why the area of focus was chosen is provided.	Student response earns this point: ❑Yes ❑No Comments:
Row 2 DAT-2.D	The written response: ■ identifies 2–3 questions and corresponding answers ■ provides at least one appropriate non-textual representation for each question ■ provides a one-sentence description for each non-textual representation	**Do NOT award a point if any of the following is true:** ■ Fewer than 2 questions are presented. ■ The answer presented is not supported by the data investigation. ■ Fewer than 2 non-textual representations are included. ■ Some non-textual representations are missing sentence descriptions.	Student response earns this point: ❑Yes ❑No Comments:
Row 3 DAT-2.D	The written response: ■ demonstrates the discovery of information or creation of new knowledge ■ builds a narrative to educate, sway opinions, or present a solution to a problem	**Do NOT award a point if any of the following is true:** ■ Restates the answers to the original questions without formulating a narrative for a data story.	Student response earns this point: ❑Yes ❑No Comments:
Row 4 DAT-2.C, DAT-2.D, DAT-2.E	The written response: ■ describes in detailed steps how the data were processed in enough detail that someone else could recreate it	**Do NOT award a point if any of the following is true:** ■ The description of the process does not match the results presented.	Student response earns this point: ❑Yes ❑No Comments:

GETTING STARTED

For the purposes of this walkthrough we will be using **Table 2C. Characteristics (Educational Attainment) of the Citizen Voting-Age Population for Congressional Districts: 2018**. The process will be similar if you use another Excel file from the provided data sets, however the "USA 2016 Presidential Election Data by County" would require a slightly different process given the size and amount of data included in the exported file.

This appendix is meant to provide teachers (and students) with a basic introduction to Excel and manipulating data within an Excel file. The directions given reflect Excel 2016, 2019, and 365, with some slight variations depending on the specific version being used. If using an earlier version of the software, some features might not be available or may be accessed in a different manner than described. This appendix does not cover advanced topics such as pivot tables or lookups.

Teacher Resource

The file Table2c.xlsx includes several sheets where these steps have already been completed. Look for the callouts to specific tabs throughout the appendix.

TURNING DATA INTO A TABLE

This allows for advanced filtering and additional data.

- Highlight the data without including headers.
    - In this file, the data start at row 6.
- On the **Insert** tab, click **Table.**
    - Ensure that the **My table has headers** checkbox is **unchecked**.
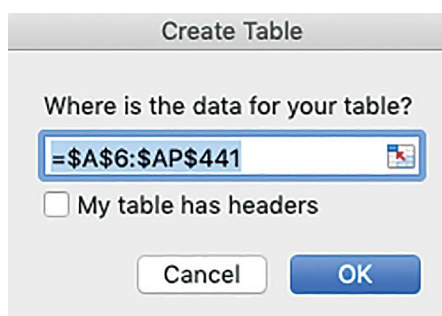
    - Rows will be changed to white and blue, with the data now starting on row 7. Row 6 is dark blue and is now an additional row of column headers. If using the provided Table2c.xlsx file, click on the **Tab_1** sheet located on the bottom left-hand side of the Excel window.

	Line #	State abbreviation	State name	Congressional district	Citizen voting-age population		Less than 9th grade		
					Estimate	Margin of error (MOE)	Estimate	Margin of error (MOE)	Percent of total
6	Column	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
7	1	AL	Alabama	1	544,464	3,424	13,409	1,845	2.5
8	2	AL	Alabama	2	516,295	5,674	18,295	2,059	3.5
9	3	AL	Alabama	3	543,854	4,099	19,578	2,504	3.6
10	4	AL	Alabama	4	515,701	4,678	24,346	2,997	4.7
11	5	AL	Alabama	5	551,968	2,121	14,908	2,282	2.7
12	6	AL	Alabama	6	535,753	7,635	11,300	2,173	2.1
13	7	AL	Alabama	7	504,177	9,052	16,616	1,923	3.3
14	8	AK	Alaska	At-large	532,244	2,539	10,640	1,693	2.0
15	9	AZ	Arizona	1	568,661	9,203	21,704	2,404	3.8
16	10	AZ	Arizona	2	550,044	9,439	11,867	2,153	2.2
17	11	AZ	Arizona	3	500,787	12,277	30,121	3,164	6.0
18	12	AZ	Arizona	4	633,528	10,386	17,595	2,673	2.8
19	13	AZ	Arizona	5	587,681	10,348	8,798	1,427	1.5
20	14	AZ	Arizona	6	577,899	11,084	8,834	2,083	1.5
21	15	AZ	Arizona	7	454,957	12,175	34,537	3,445	7.6

Table 2C. Characteristics (Educational Attainment) of the Citizen Voting-Age Population for Congressio...

All of the census files are separated by congressional district. If you want to compare states to other states, or a state to federal/national data, you will first need to filter and process the data. Most of the census data files include margins of error for each category. While you may want to have a conversation about this with your students, for the sake of time and processing data we suggest removing these columns and looking at the *Estimate* and *Percent* of total columns for each category during the modeling for Part 3. If using the provided Table2c.xlsx file, click on the **Tab_2** sheet located on the bottom left-hand side of the Excel window.

BASIC FILTERING AND FORMULAS

To get state-level totals, it is necessary to filter the data by the selected state. To do this, follow these steps:

- Click the [▼] for the **column 2, State abbreviation** or **column 3, State name** header.
- Uncheck the **(Select All)** checkbox and then select the state abbreviation or name that you would like to total. The screenshot on the next page shows the filter for only the congressional districts in AL (Alabama).
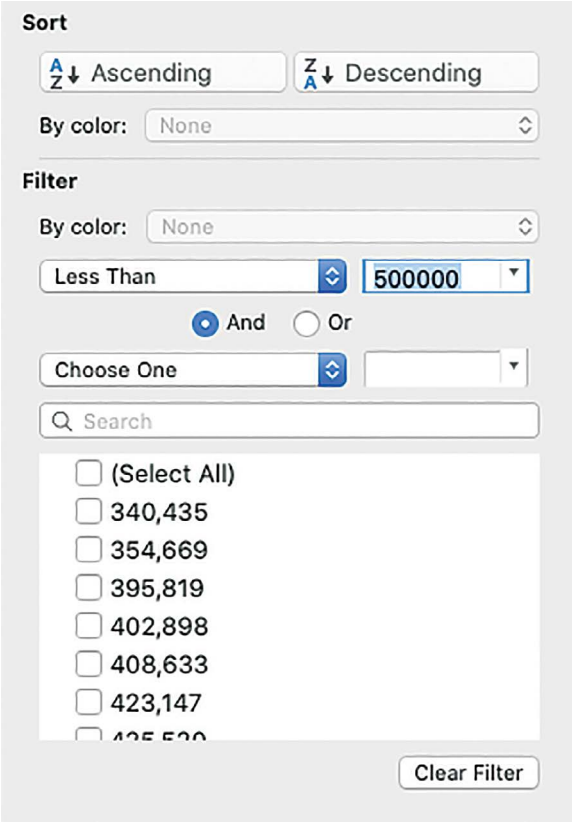
- After filtering, we will add sum totals to a new row (such as row 450). Starting in column 5 of row 450, we are going to sum up each column using the SUM function.

  - Type **=SUM** in the cell, OR click **Fx** in the formula bar to bring up a list of formulas and select **SUM** from the list.

  - Highlight the cells in **column 5** that contain the data you are planning to sum, specifically **E7** through **E13** in this example. This will result in the following formula in the cell: **=SUM(E7:E13)**, which displays as the given sum, in this case 3,712,212.

- Next we are going to fill the remaining cells of the **State total** row with the sum and average (or percent) as described.

  - For percent of total columns (such as column 9) you can use the AVERAGE function over the given cells as opposed to SUM, or you can compute the percentage by taking the **estimated value** (for example, **column 7**) and dividing it by the **state total of citizens of voting age** (**column 5**) and multiplying by 100. (You might also want to adjust the decimal location so that it shows additional information to the right of the decimal. This is especially helpful for very small percentages.) If using the provided Table2c.xlsx file, click on the **Tab_3** sheet located on the bottom left-hand side of the Excel window.

ADVANCED FILTERING

One advantage of using a table is that it makes advanced filtering easy. For instance, we can easily look at multiple states, say Washington, Oregon, and California, and then filter out any congressional district with a citizen voting-age population of 500,000 or more.

- To do this, follow the same step as above but instead of selecting a single state, select **all of the states** that you want to include.

- Once you have only the congressional districts for the states that you are interested in, click the ▼ on **Column 5**, the **Citizen voting-age population estimate**.

- Instead of selecting a checkbox or a selection of checkboxes, you can filter based on the values in the cells of that column. In this case, we want to filter, or select, values **Less Than 500,000**.

- The entire list of built-in filters is below:
  - Equals
  - Does Not Equal
  - Greater Than
  - Greater Than or Equal To
  - Less Than
  - Less Than or Equal To
  - Between
  - Top 10 (which can be Top N—items or percent)
  - Bottom 10 (which can be Bottom N—items or percent)
  - Above Average
  - Below Average
- In addition to these built-in filters, there is also the ability to use multiple filters using **And/Or**. In this way we can choose to keep only those districts with a population **Less Than 500,000** or **Greater Than 600,000**.

- When using filtering, especially advanced filtering, and then attempting to use values in a column within a formula, it is important to pay attention to the row numbers on the left-hand side. Simply selecting the first visible cell in a column and dragging to the last cell will include the cells for all rows between where you start and end in the formula, and not limit it to only those that are visible. If using the provided Table2c.xlsx file, click on the **Tab_4** sheet located on the bottom left-hand side of the Excel window.
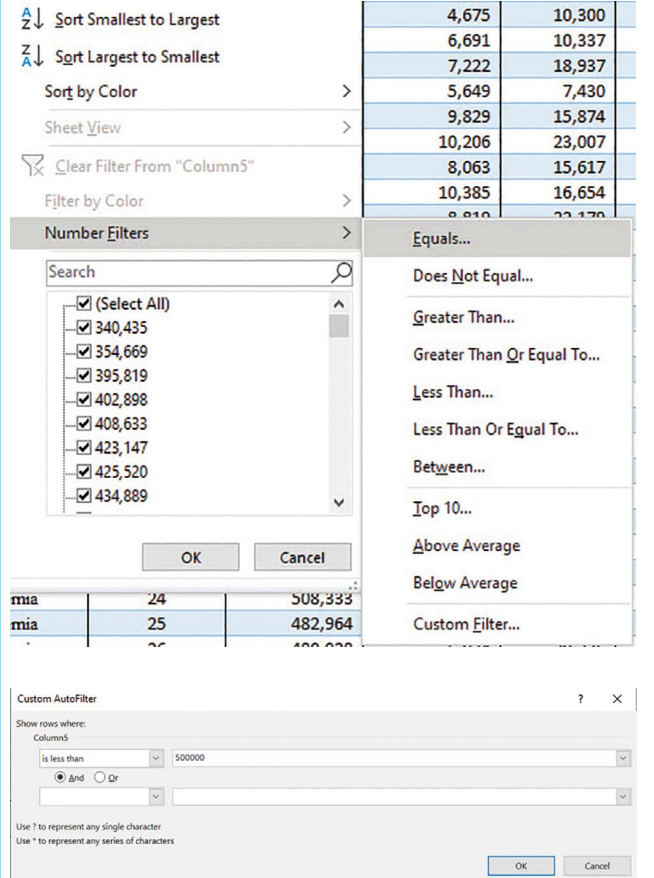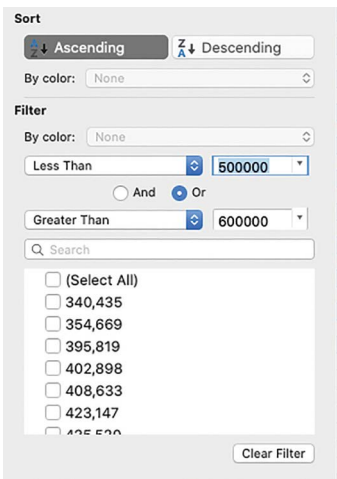- To limit the formula to only those cells that are visible, follow these steps:
  - Start to **enter the formula** that you want to use in a given cell.
  - When you begin selecting values to use in the formula, start with the first cell in the column. Select the cell and drag/highlight until the last contiguous row. In this example, that means highlighting the cells in **rows 35-37** as shown in the following screenshot.

AVERAGE ↕ ✕ ✓ *fx* =SUM(F35:F37|

SUM(number1, [number2], ...)

	A	B	C	D	E	F	G	H	I	J
3					Citizen voting-age population					
4	Line #	State abbreviation	State name	Congression al district		Less than 9th grade		9th to 12 Grade, no diploma		High scho
5					Estimate	Estimate	Percent of total	Estimate	Percent of total	Estimate
6	Column1 ▼	Column2 ▼	Columr ▼	Column ▼	Column5 ▼	Column7 ▼	Column9 ▼	Column11 ▼	Column13 ▼	Column15 ▼
35	29	CA	California	8	477,512	16,654	3.5	45,292	9.5	154,981
36	30	CA	California	9	482,967	22,179	4.6	38,221	7.9	147,403
37	31	CA	California	10	493,409	25,293	5.1	43,288	8.8	154,120
43	37	CA	California	16	423,147	33,543	7.9	50,462	11.9	124,002
44	38	CA	California	17	438,374	13,228	3.0	19,579	4.5	69,782
45	39	CA	California	18	491,580	6,902	1.4	12,128	2.5	57,910

- While holding **CTRL** (or **Command** on a Mac), select the next cell in the column and then drag and highlight until the last contiguous row. In this example that would be **rows 43-50**.

AVERAGE ↕ ✕ ✓ *fx* =SUM(F35:F37,F43:F50|

SUM(number1, [number2], [number3], ...)

	A	B	C	D	E	F	G	H	I	J
3					Citizen voting-age population					
4	Line #	State abbreviation	State name	Congression al district		Less than 9th grade		9th to 12 Grade, no diploma		High scho
5					Estimate	Estimate	Percent of total	Estimate	Percent of total	Estimate
6	Column1 ▼	Column2 ▼	Columr ▼	Column ▼	Column5 ▼	Column7 ▼	Column9 ▼	Column11 ▼	Column13 ▼	Column15 ▼
35	29	CA	California	8	477,512	16,654	3.5	45,292	9.5	154,981
36	30	CA	California	9	482,967	22,179	4.6	38,221	7.9	147,403
37	31	CA	California	10	493,409	25,293	5.1	43,288	8.8	154,120
43	37	CA	California	16	423,147	33,543	7.9	50,462	11.9	124,002
44	38	CA	California	17	438,374	13,228	3.0	19,579	4.5	69,782
45	39	CA	California	18	491,580	6,902	1.4	12,128	2.5	57,910
46	40	CA	California	19	482,363	28,909	6.0	31,052	6.4	89,585
47	41	CA	California	20	434,889	19,397	4.5	26,214	6.0	91,612
48	42	CA	California	21	354,669	30,382	8.6	49,482	14.0	114,452
49	43	CA	California	22	483,064	20,958	4.3	33,175	6.9	113,960
50	44	CA	California	23	490,446	19,789	4.0	51,116	10.4	136,106
52	46	CA	California	25	482,964	19,148	4.0	40,232	8.3	113,657
53	47	CA	California	26	480,928	20,235	4.2	24,094	5.0	101,186
54	48	CA	California	27	471,963	29,513	6.3	23,456	5.0	82,256

- Continue selecting cells in contiguous rows until all have been selected.

| AVERAGE | ✕ ✓ fx | =SUM(F35:F37,F43:F50,F52:F54,F56,F58:F59,F61:F62,F64:F68,F70:F71,F73:F74,F78) |

SUM(number1, [number2], [number3], [number4], [number5], [number6], [number7], [number8], [number9], [number10], [number11], …)

Line #	State abbreviation	State name	Congressional district	population	Less than 9th grade		9th to 12 Grade, no diploma		High scho
				Estimate	Estimate	Percent of total	Estimate	Percent of total	Estimate
Column1	Column2	Column3	Column4	Column5	Column7	Column9	Column11	Column13	Column15
29	CA	California	8	477,512	16,654	3.5	45,292	9.5	154,981
30	CA	California	9	482,967	22,179	4.6	38,221	7.9	147,403
31	CA	California	10	493,409	25,293	5.1	43,288	8.8	154,120
37	CA	California	16	423,147	33,543	7.9	50,462	11.9	124,002
38	CA	California	17	438,374	13,228	3.0	19,579	4.5	69,782
39	CA	California	18	491,580	6,902	1.4	12,128	2.5	57,910
40	CA	California	19	482,363	28,909	6.0	31,052	6.4	89,585
41	CA	California	20	434,889	19,397	4.5	26,214	6.0	91,612
42	CA	California	21	354,669	30,382	8.6	49,482	14.0	114,452
43	CA	California	22	483,064	20,958	4.3	33,175	6.9	113,960
44	CA	California	23	490,446	19,789	4.0	51,116	10.4	136,106
46	CA	California	25	482,964	19,148	4.0	40,232	8.3	113,657
47	CA	California	26	480,928	20,235	4.2	24,094	5.0	101,186
48	CA	California	27	471,963	29,513	6.3	23,456	5.0	82,256
50	CA	California	29	402,898	36,869	9.2	38,058	9.4	101,848
52	CA	California	31	482,081	18,582	3.9	45,590	9.5	133,160
53	CA	California	32	449,452	34,351	7.6	34,156	7.6	124,345
55	CA	California	34	425,520	43,810	10.3	37,570	8.8	89,614
56	CA	California	35	450,565	32,600	7.2	46,007	10.2	126,690
58	CA	California	37	468,463	25,297	5.4	30,477	6.5	78,851
59	CA	California	38	481,296	30,376	6.3	35,106	7.3	136,716
60	CA	California	39	486,713	13,837	2.8	22,452	4.6	91,136
61	CA	California	40	340,435	43,948	12.9	40,368	11.9	98,979
62	CA	California	41	478,613	29,445	6.2	43,330	9.1	152,838
64	CA	California	43	454,185	24,838	5.5	37,427	8.2	101,686
65	CA	California	44	395,819	36,450	9.2	48,320	12.2	118,351
67	CA	California	46	408,633	35,485	8.7	36,525	8.9	106,524
68	CA	California	47	491,712	30,115	6.1	29,772	6.1	91,523
72	CA	California	51	445,034	38,878	8.7	50,724	11.4	123,014

USING UNFILTERED COLUMN DATA

When computing national totals which process the unfiltered columns within the table, you can use cell ranges as we did with individual states, or you can refer to an entire column within the table using **Table[Column#]**. For the data that you are looking to compare between state and national levels, calculate the national **SUM** or **AVERAGE** of those columns. For example, to obtain the national total of citizens of voting age, you would use **=SUM(Table1[Column5])**. To obtain the percentage of voting-age citizens that have graduated high school, you would use **=AVERAGE(Table1[Column17])**. For additional help with structured references with Excel tables, click here: **https://support. microsoft.com/en-us/office/using-structured-references-with-excel-tables-f5ed2452-2337-4f71-bed3-c8ae6d2b276e.**

If comparing states with significantly different populations (such as comparing Wyoming with Texas) or comparing state(s) data to national data, it is sometimes better to look at the percentage of the population in each category as opposed to the overall number. Because of this, we will be looking at the percentage of the population in each of the following educational attainment categories:

- Less than 9th grade
- 9th to 12th grade, no diploma
- High school graduate
- Some college, no degree

- Associate's degree
- Bachelor's degree
- Graduate degree

We will use the values from Alabama and New York and compare them to the national numbers. While it is possible to calculate the national values for all columns, because we are only going to be looking at the percent in each category it is not necessary to calculate the estimated population in each category.

FORMATTING DATA

To make chart creation easier, it helps to lay out your data together, either in row or column format. This is also a good time to review the data to ensure that there were no copy/paste errors and that it all looks accurate. You can click on a cell and ensure that the formulas are correct or highlight the appropriate cells and compare the value in the table with the value in the status bar. Excel defaults to showing Average, Count, and Sum when multiple cells containing numerical values are selected.

	Less than 9th grade	9th to 12 Grade, no diploma	High school graduate	Some college, no degree	Associate's degree	Bachelor's degree	Graduate degree
AL	3.2	9.8	31.5	23.9	8.4	15.0	8.3
NY	4.1	6.7	25.9	18.7	8.8	20.8	15.1
National	3.2	6.8	27.8	23.3	8.5	19.2	11.2

Education	AL	NY	National
Less than 9th grade	3.20	4.1	3.2
9th to 12 Grade, no diploma	9.8	6.7	6.8
High school graduate	31.5	25.9	27.8
Some college, no degree	23.9	18.7	23.3
Associate's degree	8.4	8.8	8.5
Bachelor's degree	15	20.8	19.2
Graduate degree	8.3	15.1	11.2

CREATING A CHART FROM DATA

At this point some students may feel comfortable looking at and interpreting the data in their current, table-based format. It is possible to identify trends or highlight similarities and differences between the states, as well as how those states compare to overall national data. It is also possible that students will want to look at a non-textual representation of these data in a chart or graph.

To create a chart, highlight the **table containing the data** that you want to chart, click on **Insert**, and then select either **Recommended Charts** or a specific type of chart from the options given. If using the provided Table2c.xlsx file, click on the **Tab_5** sheet located on the bottom left-hand side of the Excel window.
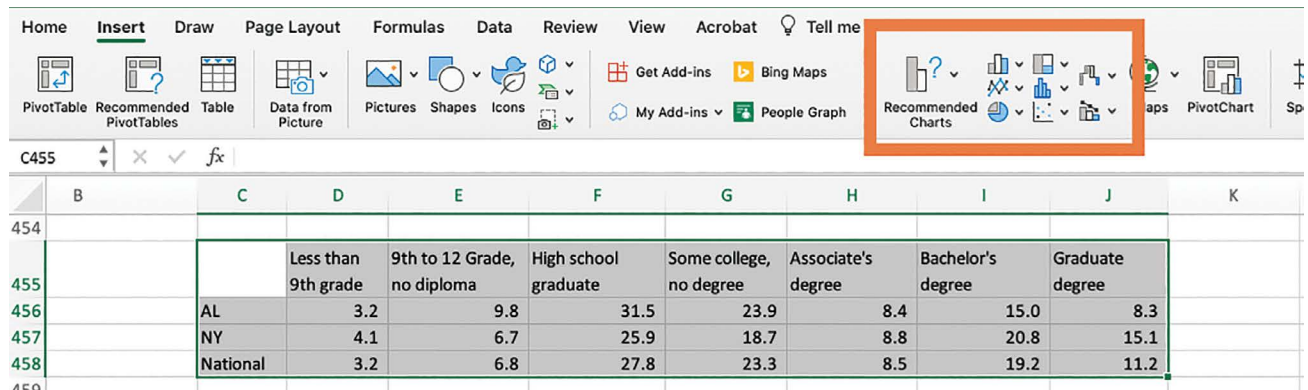
There are several charts that Excel can create, and what makes the most sense depends on the data and the audience. Here are four different charts created from the data:

When creating the stacked bar/column charts, it was necessary to click on the **Switch Row/Column** button so that the categories within each state were more easily comparable. Otherwise the national and state values for each category were stacked on top of each other. This can be helpful when looking at percentages and comparing several states side by side.

There are several other charts and graphs that can be helpful when looking at data, and students should explore different ways to display their data.

FORMATTING A CHART

Chart Title

Once a chart has been created, it can still be modified. Right-clicking on an existing chart provides multiple ways to edit the chart, including changing the chart type and selecting data (useful if you forget a row or add data that you want included). You can modify the default chart title by clicking on it and changing it to something more descriptive. For the purpose of this chart something like "Educational Attainment in Voting-Age Citizens" might be appropriate.

Formatting Axes

Another element that students will likely want to edit is the axis. Excel will choose a default maximum and minimum based on the values in the cells being used. For example, the stacked column and bar charts go to 120 as the maximum value, however the maximum sum of all the values is only 100.

By changing the maximum to 100, the empty space at the top of the chart is removed, making it easier to see the similarities and differences that exist. There might be times when the minimum also needs to be modified. Although this chart is only using numeric values on the vertical axis, it is possible that students will be charting data that is numeric on both axes, such as a scatter plot, and in this situation, they should ensure that both axes are formatted appropriately. Another possibility that they may need to consider depending on their graph is the use of a logarithmic scale.

Data Labels

Data labels can also be added to the chart if the values are unclear. If the chart will be side by side with the table, it is likely unnecessary. However, if the chart will be provided on its own, data labels can help highlight trends, similarities, and differences. To add data labels, right-click on a data series and choose **Add Data Labels**.
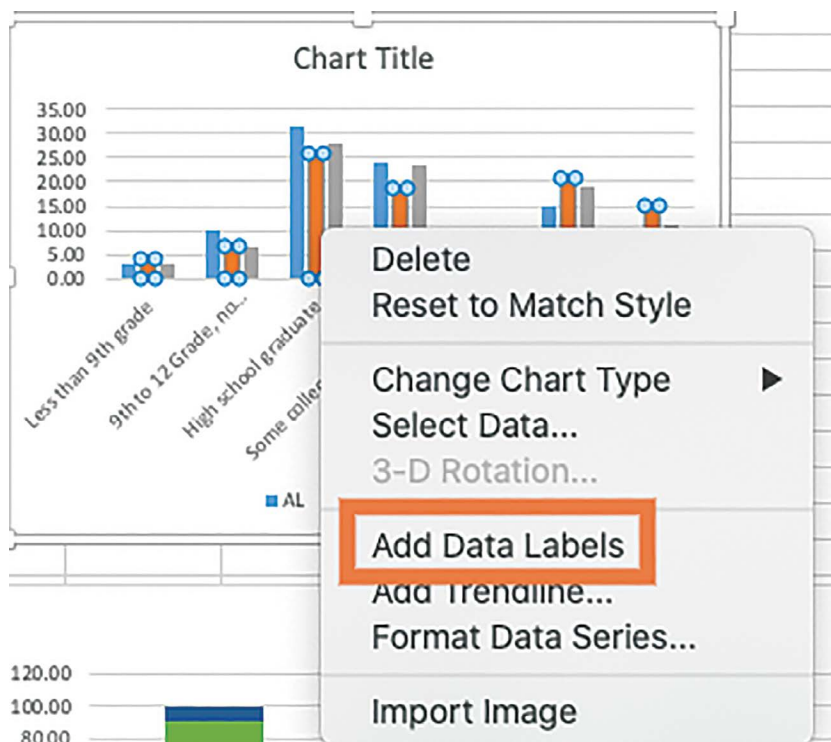
Data labels can be further formatted after they have been added by right-clicking on the data series and selecting **Format Data Labels**. This includes where the labels are placed, the font used, and for numeric labels, number formatting. The values selected in the given screenshot are meant to be illustrative of the options available to students, and they should experiment with the formatting options they have available to them.

Changing Colors and Additional Chart Elements

Charts are often created with a default color palette. This can be changed by selecting the chart, and then selecting the **Chart Design** tab. This tab provides the ability to add different chart elements, change the default layout of the chart (such as where the key and data labels are placed), and change the color palette used. As with other chart options, students should explore what options make their charts more readable and visually appealing. If using the provided Table2c.xlsx file, click on the **Tab_Charts** sheet located on the bottom left-hand side of the Excel window.