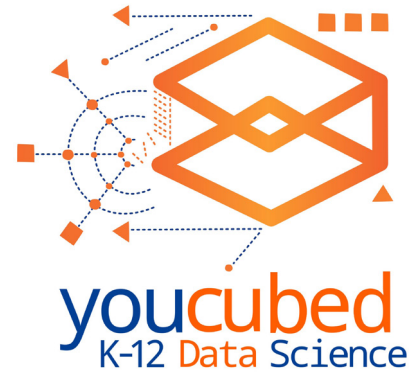# Unit 4: Understanding Variability

Introduction:

In this unit, students will explore data from the Pew Core Trends Survey, which is a national survey about internet and social media use. They will also give a version of the survey to a few of their friends and/or family and work with the two samples (national and local) to compare and contrast. The goal of this lesson is for students to get a sense of how data may vary depending on populations and sampling, as well as comparing and contrasting the data they collected to the Pew Survey. This lesson is shown in Jo Boaler's online class, taught by one of our youcubed team members: https://www.youcubed.org/21st-century-teaching-and-learning/.

## Lesson 1: Exploring Data with CODAP

Time	Activity	Materials
35 min	**Exploration:** Students explore national data from survey in CODAP	Survey data in CODAP: • https://codap.concord.org/app/static/dg/en/cert/index.html#shared=156634 • Questionnaire handout: page 6-11
20 min	**Survey creation:** • As a class, decide which questions from the survey would provide the most interesting data to collect • Make a survey using the questions • Discuss how they think the samples will be different.	• Questionnaire handout • A tool for creating and delivering a survey • As an example, Google Forms was used to make this survey for use in this lesson: https://docs.google.com/forms/d/e/1FAIpQLSfs62aJH9HgJ1ASVkJYJCMtR75tqw6V2hW3qsbu3-6FfZw-Qw/viewform
20 min	**Data collection:** • Ask students to take the survey and give it to at least 1 adult (could be teacher or parent) and 2 peers. Tell them that is a minimum, they can collect more if they want to.	• Students will need some time for their data collection. While data collection is happening class time could be devoted to further analysis of the Pew data in CODAP.

15 min	**Data talk:** • Show each graph and ask: "What do you notice?" "What do you wonder?" • What else are you curious about regarding Americans' reading habits? What data would you need to answer that question? How would you collect it?	**Graphs of Pew Research Center surveys conducted in 2019** • https://www.pewresearch.org/wp-content/uploads/2019/09/FT_19.09.25_BookReadingFormats_Print-books-more-popular-e-books-audiobooks.png?resize=420,478 • https://www.pewresearch.org/wp-content/uploads/2019/09/FT_19.09.24_BookReading_2.png?resize=310,428
25 min	**Data transfer:** • Students transfer their survey data into CODAP and examine their local survey results	
90 min	**Data study & presentations:** • In groups, explore the National and their local survey data in CODAP and make a poster sharing their findings. • Find 3 differences or similarities between the samples. How do you know and why do you think it might be the case?	• Chart paper • Colored pens or pencils • Rulers or scissors • The data and visualizations could be make in a google doc or as a slide presentation
30 min	**Poster presentations:** • You may want to ask students to present their findings or provide time for them to do a gallery walk followed by a class discussion. • Ask students to share the 3 differences or similarities they found and to justify their findings	
5 min	**Reflection/journal entry:** • What was a difference between the national data and your community's data that surprised you? What did you learn from it?	• Journal

Exploration:

This lesson begins with an opportunity for students to explore data from a national survey by the Pew Research Center. The data set is from a survey done January 3 - 10, 2018 where the focus was on home technology use and the opinion people have about using the internet. You can find the data set here. (https://codap.concord.org/app/static/dg/en/cert/index.html#shared=156634). We chose to adapt the data set for this activity by deleting some attributes (columns) we thought not useful in this initial activity. If you would like to view the entire data set and supporting documents you can find them here: https://www.pewresearch.org/internet/dataset/jan-3-10-2018-core-trends-survey/.

To start we suggest you provide students time to play with the data inside of CODAP and to explore the questionnaire document that contains the questions, answer choices and coding for the choices. This is unlike other data sets we have used in these youcubed units. For example, this file has column headers like "smart2" and "bbhome2", which are not readable without a reference. Students will need the questionnaire handout since it shows all the original questions on the survey, their corresponding attributes, and the value that was used to code the answer. For example, in the "bbhome2" column a value of "1" means "yes" and a value of "2" means "no". This document will take some time for students to understand. It is a good exercise in close reading.
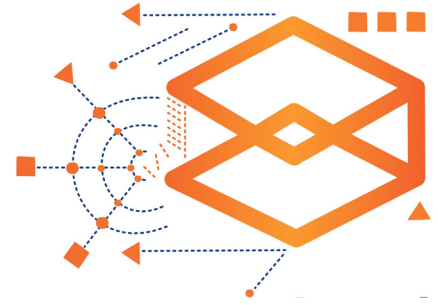
The case (row) values are all numerical with the exception of one attribute (column) which collected an open ended response about what the participant thought of the impact the internet has had on society. As students work within the data set inside of CODAP, pause and ask them to share their findings. This is a great practice to encourage further exploration.

Survey creation:

After students have had time getting familiar with the data set and the questionnaire, ask them to think about which questions they think would be most interesting to study within their own community. As a class decide on the questions that should be in their local survey and create a questionnaire. You will want to decide beforehand on a tool to create the survey or ask the class to research and decide the best survey tool for this assignment. As an example, one teacher used Google Forms for their class survey. You can view this exapmle at https://docs.google.com/forms/d/e/1FAIpQLSfs62aJH9HgJ1ASVkJYJCMtR75tqw6V2hW3qsbu3-6FfZw-Qw/viewform.

Data collection:

After the survey is created ask students to take the survey themselves and give the survey to at least 2 peers and 1 adult. Before students collect their data, discuss as a class what the differences in the data from these samples might be. One of the key ideas of this lesson is that even the same questions, when answered by different groups of people, will generate different data. This is called variability. Variability can come from many aspects, but it can be as simple as the fact that as different people we have different opinions. In fact, if variability didn't exist, we wouldn't need statistics or data science, as there would be a single answer to every question. Make conjectures on what might be some differences between the US national data and the data from their community? Do students think responses would be different across age groups?

Data talk:

This Data Talk presents students with graphs made using the data they have been exploring. Share these graphs side by side and ask students, "What do you notice?", "What do you wonder?". Ask students to think about and comment on the style of the data displays and the information provided. Ask them, "Can you think of a better way to visualize this data?".

In the trial of this lesson we showed students two graphs published by the Pew Research Center made with the same data students have been considering. This time, we showed the students both graphs side-by-side so they could discuss either of them but also the connection between them.

After our regular data talk, we took this as an opportunity to incorporate a new question-asking exercise posing the following questions:

- What else are you curious about regarding Americans' reading habits?
- What data would you need to answer that question?
- How would you collect it?

This discussion is shown in our online course.

The goal is for students to practice the data inquiry skills they will soon need to analyze and draw conclusions from the survey dataset. Rather than being handed a question to ask of the data, it is important for students to practice formulating their own questions about what interests them.
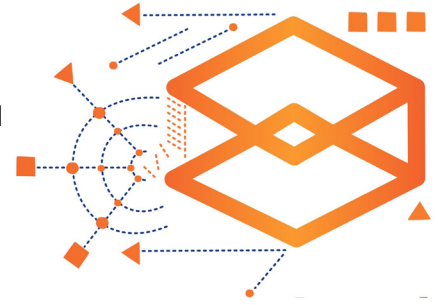
For example, some other questions that could be asked:
- What would the first graph look like if we separated the data by age group?
    - We would need to know the ages of the respondents
    - We could collect that in the same survey used to get the data above.
- Do people of different ages read different genres?
    - When people are asked how many books they read in the last year, that could be followed up by: what genres of books did you read?

Data transfer:
Share the survey data with your students. If you used an e-version of the survey, share a link to the survey results so your students can take a look at the data they and their classmates gathered from their community. If you used Google Forms you can find automatically generated visualizations under the "Responses" tab.

To put data into CODAP you must have a .csv file. When downloading or saving the class data you can choose .csv as an option.You may want to do this yourself or provide students the file so they can experience the data upload into CODAP themselves.

Give students a few minutes to look at the results. Discuss as a class how you could compare the local data that was collected to the national survey results. Ask students to think about features of their community that might be different from the national sample and how that might show itself in the survey results. Once students have some ideas of comparisons they want to make, ask them to brainstorm methods they could use to compare the datasets. Students may suggest things like making the same graphs for each dataset in 2 different CODAP windows or finding the measures of center of each dataset and comparing them. This should be an open, exploratory discussion to generate ideas.

Next, as a class, brainstorm questions students want to answer using the 2 data sets they have. As students brainstorm questions, record them on the board.

Data study:
In small groups, ask students to find 3 differences or similarities between the samples and make a poster representing their findings. For each one, ask them to answer how they know it's true and why they think it might be the case. For example, a student might find that a higher percentage of young people from the local sample uses Snapchat than people from the local adult sample. They could show how they know that's true by showing a graph or calculation. They could conjecture that this might be the case because Snapchat is targeted at young adults.

Poster presentations:
As students present their posters, focus on their justifications. How do they use the data to support their claims? What methods did they use to compare one sample to the other? Encourage students to ask these questions of each other. When sharing conjectures about why the differences or similarities might exist, allow other students to make their own and share them with the class.

Look-fors:

- **How are students tackling the problem of uploading the data into CODAP?**
  Although it will take longer, it is important for students to figure out how to upload their data into CODAP on their own without being told exactly how to do it. It will encourage students to be resourceful and it is also an opportunity to practice their growth mindsets. If they are frustrated, remind them that mistakes are key to learning and speed is not what's important. Looking up how to solve the issues that present themselves as they work with data is a skill that will be valuable regardless of the datasets or software they deal with in the future, you can emphasize this.

- **How are students using the data analysis software to analyze the data?**
  Students will have used many of the different capabilities that CODAP has to offer. They can look at a single variable and how it's distributed along with its measures of center, they can compare two variables (one on each axis or one on an axis and color-coding for the other one), or even three variables by using the two axes and the color-coding capability. This is a great time to focus on when and how they are using each of these options. If you notice any of these being under- or misused, have a conversation with your students. As novice users of data analysis software, it is easy to fall into repeating patterns and to forget the many powerful options they have.

- **How are students formulating questions to ask of the data?**
  In the poster presentation, students are asked to take each of their claims and explain how they know it is true and why they think it is the case. Make sure students are differentiating between claims that are supported by data and their own ideas or conjectures. Encourage students to think about and hypothesize why something might be the case, but remind them that those claims are different from observations made directly from data. You can ask students how they would go about testing their conjectures if they wanted to prove them with data.

- **How are students comparing the different samples?**
  As students work with the same survey questions asked to different groups, ask them how they are making sure their comparisons are valid. Make sure they are not comparing apples to oranges (for example, the mean age of one sample to the median age of another one). In order to make a valid comparison, students will need to apply the exact same analysis to each set of data they wish to compare.
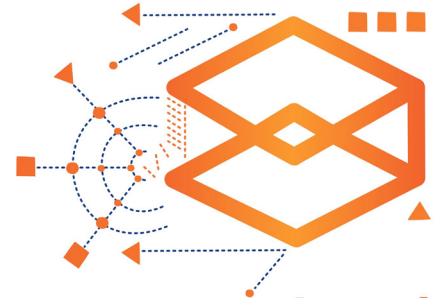
Reflection/journal entry:
What was a difference between the national data and your community's data that surprised you? What did you learn from it?

Questionnaire Handout

Pew Research / Abt Associates -- Internet Core Trends Update December 14 draft

Total n=2,000 U.S. adults age 18+
n=400 landline RDD
n=1,600 cell phone RDD
Field Dates: January 2 – January 10, 2018 (tentative)

START TIMING MODULE 1
LANDLINE INTRO:
Hello, I am _____ calling on behalf of the Pew Research Center. We are conducting a telephone opinion survey about some important issues today and we would like to include your household. May I please speak with the YOUNGEST [**RANDOMIZE**: (MALE / FEMALE)], age 18 or older, who is now at home? [**IF NO MALE/FEMALE, ASK:** May I please speak with the YOUNGEST (FEMALE / MALE), age 18 or older, who is now at home?]
GO TO MAIN INTERVIEW

CELL PHONE INTRO:
Hello, I am _____ calling on behalf of the Pew Research Center. We are conducting a telephone opinion survey about some important issues today. I know I am calling you on a cell phone. If you would like to be reimbursed for your cell phone minutes, we will pay all eligible respondents $5 for participating in this survey. This is NOT a sales call.

> [**IF R SAYS DRIVING/UNABLE TO TAKE CALL:** Thank you. We will try you another time...]
>
> **ORD AS <u>AGE INELIGIBLE</u>:** This survey is limited to adults age 18 and over. I won't take any more of your time...
> **IF S1=9, THANK AND TERMINATE – RECORD AS SCREENING REFUSAL:** This survey is limited to adults age 18 and over. I won't take any more of your time...
>
> **READ TO ALL CELL PHONE RESPONDENTS**
> **INTRODUCTION TO MAIN INTERVIEW:** If you are now driving a car or doing any activity requiring your full attention, I need to call you back later. The first question is...
>
> **INTERVIEWER: If R says it is not a good time, try to arrange a time to call back. Offer the toll-free call-in number they can use to complete the survey before ending the conversation.**

[**PROGRAMMER NOTE: PLEASE INCLUDE THE INTRODUCTION RANDOMIZATION VARIABLES IN THE ALL CONTACTS FILES. WE WOULD LIKE TO BE ABLE TO RUN RESPONSE RATES SEPARATELY FOR EAVOICEMAIL MESSAGE [LEAVE ONLY ONCE -- THE FIRST TIME A CALL GOES TO VOICEMAIL:]** I am calling on behalf of the Pew Research Center. We are conducting a national opinion survey of cell phone users. This is NOT a sales call. We will try to reach you again.

CELL PHONE SCREENING INTERVIEW:
> S1.   Are you under 18 years old, OR are you 18 or older?
> 1   Under 18
> 2   18 or older
> 9   Don't know/Refused

IF S1=2, CONTINUE WITH MAIN INTERVIEW
IF S1=1, THANK AND TERMINATE – RECCH VERSION OF THE INTRODUCTION FOR THE LANDLINE AND CELL FRAMES SEPARATELY. PLEASE RANDOMIZE INTRO LANGUAGE WITH ONE TREATMENT PER PHONE NUMBER NOT PER CALL.]

ASK ALL:
EMINUSE.   Do you use the internet or email, at least occasionally?
     1   Yes
     2   No
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

ASK ALL:
INTMOB.   Do you access the internet on a cell phone, tablet or other mobile handheld device, at least occasionally?
     1   Yes
     2   No
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

ASK ALL INTERNET USERS (EMINUSE=1 OR INTMOB=1):
INTFREQ.   About how often do you use the internet? **[READ]**
     1   Almost constantly
     2   Several times a day
     3   About once a day
     4   Several times a week, OR
     5   Less often?
     8   (VOL.) Don't know
     9   (VOL.) Refused

ASK ALL INTERNET USERS (EMINUSE=1 OR INTMOB=1):
HOME4NW.   Do you currently subscribe to internet service at HOME? {Heavily modified PIAL Trend, most recently July 2015 Tracking}
     1   Yes
     2   No
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

ASK IF HOME INTERNET SUBSCRIBER (HOME4NW=1):
BBHOME1.   Do you subscribe to dial-up internet service at home… OR do you subscribe to a higher-speed broadband service such as DSL, cable, or fiber optic service? {Heavily modified PIAL Trend, most recently July 2015 Tracking}
     1   Dial-up
     2   Higher-speed
     3   **(VOL.)** Both Slow-speed/Dial-up and Higher-speed/Broadband
     4   **(VOL.)** Access internet only using cell phone or tablet
     5   **(VOL.)** No home internet access
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

ASK IF DIAL-UP USER (BBHOME1=1):
BBHOME2.   Just to confirm, you use a dial-up connection to the internet at home, and not a higher-speed broadband connection? {Heavily modified PIAL Trend, most recently July 2015 Tracking}
     1   Yes, dial-up
     2   No, higher speed connection
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

ASK IF LANDLINE SAMPLE:
DEVICE1a.   Next, do you have a cell phone, or not? {PIAL Trend, most recently November 2016 – Info Wary}
        1   Yes
        2   No
        8   **(VOL.)** Don't know
        9   **(VOL.)** Refused

ASK IF HAVE CELL PHONE (CELL PHONE SAMPLE OR DEVICE1a=1):
SMART2.   Is your cell phone a smartphone, or not? {Modified PIAL Trend, most recently November 2016 – Info Wary}
        1   Yes, smartphone
        2   No, not a smartphone
        8   **(VOL.)** Don't know
        9   **(VOL.)**Refused

ASK ALL:
SNSINT2.   Do you ever use social media sites like Facebook, Twitter or Instagram? {Modified PIAL Trend, most recently November 2016 – Info Wary}
        1   Yes
        2   No
        8   (VOL.) Don't know
        9   (VOL.) Refused

ASK ALL:
DEVICE1.   Please tell me if you happen to have each of the following items, or not. Do you have **[INSERT ITEMS; RANDOMIZE]**? {Modified PIAL Trend, most recently November 2016 – Info Wary}
        b.   A tablet computer
        c.   A desktop or laptop computer
        d.   A game console

CATEGORIES
        1   Yes
        2   No
        8   **(VOL.)**Don't know
        9   **(VOL.)**Refused

ASK IF USE INTERNET (EMINUSE=1 OR INTMOB=1) OR OWN CELL PHONE (CELL PHONE SAMPLE OR DEVICE1a=1):
WEB1.   Please tell me if you ever use any of the following social media sites online or on your cell phone. Do you ever use… **[INSERT ITEMS; RANDOMIZE]**? {Modified PIAL Trend, most recently April 2016 - Libraries}
        a.   Twitter
        b.   Instagram
        c.   Facebook
        d.   Snapchat
        e.   YouTube
        f.   WhatsApp
        g.   Pinterest
        h.   LinkedIn

CATEGORIES
        1   Yes, do this
        2   No, do not do this
        8   **(VOL.)**Don't know
        9   **(VOL.)** Refused

ASK USERS EACH SITE IN WEB1a-WEB1e (WEB1a=1 OR WEB1b=1 OR WEB1c=1 OR WEB1d=1 OR WEB1e=1):
SNS2.   Thinking about the social media sites you use... About how often do you visit or use **[INSERT ITEMS; RANDOMIZE]? [READ CATEGORIES AS NECESSARY]** {Modified PIAL Trend, most recently April 2016 - Libraries}

        ASK a IF TWITTER USER (WEB1a=1):
        a.   Twitter
        ASK b IF INSTAGRAM USER (WEB1b=1):
        b.   Instagram
        ASK c IF FACEBOOK USER (WEB1c=1):
        c.   Facebook
        ASK d IF SNAPCHAT USER (WEB1d=1):
        d.   Snapchat
        ASK e IF YOUTUBE USER (WEB1e=1):
        e.   Youtube
CATEGORIES
1   Several times a day
2   About once a day
3   A few times a week
4   Every few weeks
5   Less often
8   **(VOL.)** Don't know
9   **(VOL.)** Refused

ASK ALL:
PIAL5.   How difficult would it be, if at all, to give up the following things in your life? If you do not use or have the item, just tell me. How hard would it be for you to give up... **[INSERT ITEM; RANDOMIZE]**? {PIAL Trend, most recently January 2-5, 2014 Omnibus}

[READ FOR FIRST ITEM: THEN AS NECESSARY: Would it be very hard, somewhat hard, not too hard, or not hard at all?]
        a.   Your television
        ASK b IF HAVE CELL PHONE (CELL PHONE SAMPLE OR DEVICE1a=1):
        b.   Your cell phone or smartphone
        ASK c IF ALL INTERNET USER (EMINUSE=1 OR INTMOB=1):
        c.   The internet
        ASK d IF SOCIAL MEDIA USER (SNSINT2=1):
        d.   Social media

CATEGORIES
1   Very hard
2   Somewhat hard
3   Not too hard
4   Not hard at all
5   **(VOL.)** Impossible
6   **(VOL.)** Do not use / Do not have
8   **(VOL.)**Don't know
9   **(VOL.)** Refused

ASK ALL:
PIAL11.   Overall, when you add up all the advantages and disadvantages of the internet, would you say the internet has mostly been **[ROTATE**: (a GOOD thing) or (a BAD thing)] for society? {January 2-5, 2014 Omnibus}
1   Good thing
2   Bad thing
3   **(VOL.)** Some of both
8   **(VOL.)** Don't know
9   **(VOL.)** Refused

ASK IF INTERNET HAS HAD GOOD OR BAD IMPACT ON SOCIETY (PIAL11=1,2):
PIAL11a.   What is the main reason you think the internet has been a [IF PIAL11=1 "good"; IF PIAL11=2 "bad"] thing for society? **[DO NOT PROBE FOR ADDITIONAL ANSWERS]**

[OPEN END]

ASK ALL INTERNET USERS (EMINUSE=1 OR INTMOB=1):
PIAL12.   How about you, personally? Overall, when you add up all the advantages and disadvantages of the internet, would you say the internet has mostly been   **[ROTATE IN SAME ORDER AS PIAL11**: (a GOOD thing) or (a BAD thing)]** for you? {January 2-5, 2014 Omnibus}
     1   Good thing
     2   Bad thing
     3   **(VOL.)** Some of both
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

ASK ALL:
BOOKS1.   During the past 12 months, about how many BOOKS did you read either all or part of the way through? Please include any print, electronic, or audiobooks you may have read or listened to. {PIAL Trend, most recently April 2016 - Libraries}
     ___   **[RECORD EXACT NUMBER 1-96]**
     0   None
     97   97 or more
     98   **(VOL.)** Don't know
     99   **(VOL.)** Refused

ASK IF READ ANY BOOKS IN PAST 12 MONTHS (BOOKS1=1-97, 98):
BOOKS2.   Thinking about all of the books you have read in the past 12 months, were any of those… **[INSERT ITEMS IN ORDER]**? Were any of those … **[INSERT NEXT ITEM]**? {Modified PIAL Trend, most recently April 2016 - Libraries}
     a.   Printed books
     b.   Audiobooks
     c.   E-bookss

CATEGORIES
     1   Yes
     2   No
     8   **(VOL.)** Don't know
     9   **(VOL.)** Refused

Demographics

[READ TO ALL:] A few last questions for statistical purposes only…

ASK ALL:
SEX.   **RECORD RESPONDENT SEX [DO NOT ASK]**
     1   Male
     2   Female

ASK ALL:
AGE.   What is your age?
     ___   years **[RECORD EXACT AGE 18-96]**
     97   97 or older
     98   Don't know
     99   Refused

ASK ALL:
MARITAL.   Are you currently married, living with a partner, divorced, separated, widowed, or have you never been married? **[IF R SAYS "SINGLE" PROBE TO DETERMINE APPROPRIATE CATEGORY]**
        1   Married
        2   Living with a partner
        3   Divorced
        4   Separated
        5   Widowed
        6   Never been married
        8   **(VOL.)** Don't know
        9   **(VOL.)** Refused

ASK ALL:
EDUC2.   What is the highest level of school you have completed or the highest degree you have received? **[DO NOT READ] [INTERVIEWER NOTE: Enter code 3- HS grad if R completed training that did NOT count toward a degree]**
        1   Less than high school (Grades 1-8 or no formal schooling)
        2   High school incomplete (Grades 9-11 or Grade 12 with NO diploma)
        3   High school graduate (Grade 12 with diploma or GED certificate)
        4   Some college, no degree (includes some community college)
        5   Two-year associate degree from a college or university
        6   Four-year college or university degree/Bachelor's degree (e.g., BS, BA, AB)
        7   Some postgraduate or professional schooling, no postgraduate degree (e.g. some graduate school)
        8   Postgraduate or professional degree, including master's, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD)
   98   Don't know
        99   Refused

[MAKE FULL NOTE AVAILABLE FOR INTERVIEWERS: Enter code 3-HS graduate if R completed vocational, business, technical, or training courses after high school that did NOT count toward an associate degree from a college, community college or university (e.g., training for a certificate or an apprenticeship)]

ASK ALL:
EMPLNW.   Are you now employed full-time, part-time, retired, or are you not employed for pay?
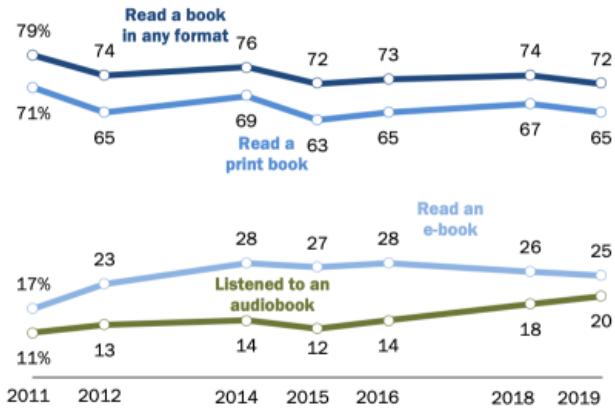   1   Employed full-time
        2   Employed part-time
        3   Retired
        4   Not employed for pay
        5   **(VOL.)** Have own business/self-employed
        6   **(VOL.)** Disabled
        7   **(VOL.)**Student
        8   **(VOL.)** Other
        98   **(VOL.)** Don't know
        99   **(VOL.)** Refused

Data Talk
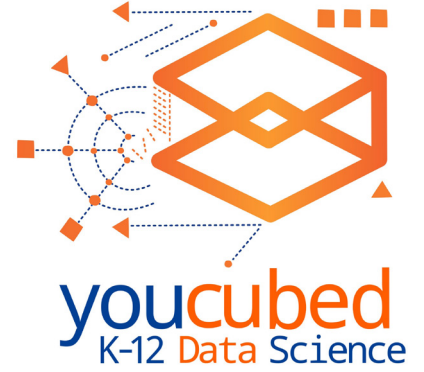
Print books continue to be more popular than e-books or audiobooks

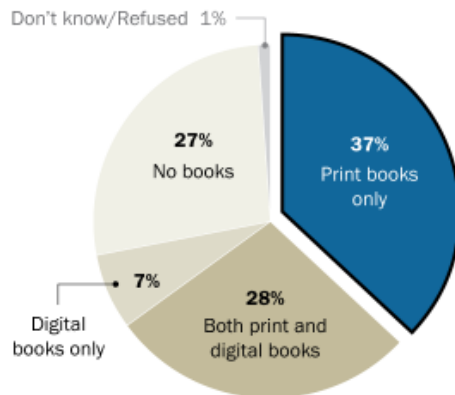*% of U.S. adults who say they have ____ in the previous 12 months*

Read a book in any format
79% | 74 | 76 | 72 | 73 | 74 | 72

Read a print book
71% | 65 | 69 | 63 | 65 | 67 | 65

Read an e-book
17% | 23 | 28 | 27 | 28 | 26 | 25

Listened to an audiobook
11% | 13 | 14 | 12 | 14 | 18 | 20

2011 | 2012 | 2014 | 2015 | 2016 | 2018 | 2019

Note: Those who gave other answers or no answer are not shown.
Source: Survey conducted Jan. 8-Feb. 7, 2019.

PEW RESEARCH CENTER

37% say they only read print books

% of U.S. adults who say they have read ____ in the previous 12 months

Don't know/Refused 1%

27% No books

37% Print books only

7% Digital books only

28% Both print and digital books

Note: The "digital books" category includes both e-books and audiobooks.
Source: Survey conducted Jan. 8-Feb. 7, 2019.

PEW RESEARCH CENTER