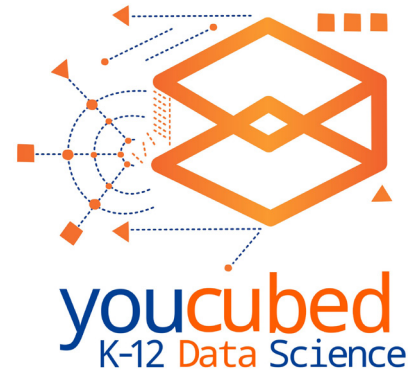# Introduction to Data Science Units

Welcome to our data science introductory units – intended for students from grade 6 up to grade 10 (or higher). One of the lovely qualities of data science is that it is a naturally low floor and high ceiling content area, all students can benefit from looking at a set of data or a data visualization and analyzing it. This will help them learn important content from the mathematics in their grade level and help them develop data literacy, so they can become better sense-makers in the world.

You may be wondering: What is data science? If so, I recommend my new online class, that can be taken free, or for a certificate. It will not only teach you about data science, but give you lots of teaching ideas, and is the teaching companion to the units on the site. You can access the course from here: www.youcubed.org/21st-century-teaching-and-learning/. Inside the course you will see some of these data units taught in a school. If you are concerned that the data science and statistics that you need to teach students is not content that you are confident with, do not worry as the online course is designed to help you learn the new content. The online class also includes 2 animated films, to teach students about measures of center (mean, mode and median) and measures of spread (standard deviation).

You may also be wondering: What is the difference between statistics and data science? My answer to that is the 2 fields share some methods but they are completely different subjects. To me statistics fits within the 19th century way of thinking about mathematics – as a list of methods. Students in statistics classes are usually taught to learn each method and to use the method to answer a question that has one correct answer. Data science does not start with methods, it starts with a data-set, that is sufficiently complex that any learning from the data will requires some organization and analysis. At the start of data science students are encouraged to come up with questions and to be curious. They move from this phase of curiosity and question asking, to pattern seeking. As they explore the data-set looking for patterns, they will need some useful mathematical tools. After they have explored the data, and developed some useful insights they will make a report, that includes data visualizations and writing. A professional data scientist would also spend a lot of time cleaning the data, preparing it to be ready for data exploration. As students learn data science in our units they are given relatively straight-forward data sets – our new high school course for older students will engage the students in data cleaning and preparation.

This short video (less than 2 minutes) is one I have prepared for you, to introduce the idea of data science. I thought it might be useful to show to parents, school boards, administrators and students: https://www.youcubed.org/resources/data-science-initiative-video/.

An important principle of K-12 data science is playful exploration and pattern-seeking. Another principle is personal meaning and relevance. Students should be encouraged to ask questions that are interesting to them, and different students in a class will probably be exploring alternative questions, which will make for engaging discussions at the end of the units. The world is now filled with data and we hope teachers will have many opportunities to bring in data visualizations and data sets from students' local regions at some times, to encourage students to learn that they can ask and answer questions that can help their local communities.

Many of our lessons in the units start with what we are calling a Data Talk. For those of you who know the important pedagogical routine called a Number Talk, you will see parallels. In a data talk we show students a data visualization and ask them what they see, and what they are curious about. One of the best sites for sources of data talks is the New York Times: What's Going on in this Graph? Students are prompted with the questions: what do you notice? What do you wonder? https://www.nytimes.com/column/whats-going-on-in-this-graph. We like those questions. Data talks provide a space for students to practice considering and interpreting a variety of data and data representations in a low-stake, exploratory environment.
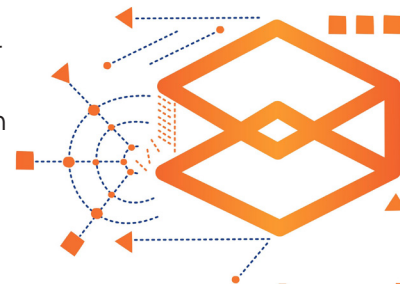
We have provided some ideas for data talks but feel free to choose a data visualization of your choosing. An ideal data visualization is one that is interesting or relevant to the students, but also that displays data in a way that is new to students or that might have some quirks or features that make the visualization harder to read (and often more interesting!).

Students are shown the visualization, given some time to process it, and discuss what they notice. These can be observations about how the visual is structured, a question the data is answering, or a lingering curiosity that is raised by the data or that the visual doesn't address.
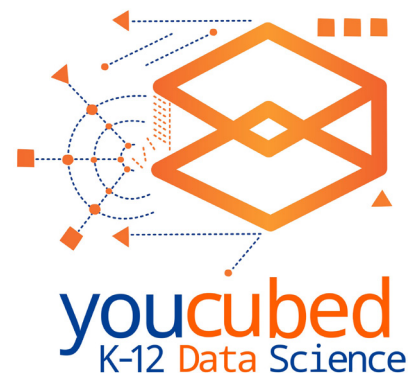
The different units in these lessons are somewhat interchangeable. We recommend that students work in journals. Our favorite journals (that can be used for all of mathematics tasks) are those with triangular or square dotty paper, they, in my view, encourage mathematical freedom. We do not know how long your lessons are but we have added approximate times. The openness of data exploration means you can probably run the lessons in less or more time. Sometimes data explorations will need to go over different lessons.

Most of the lessons involve a free data analysis tool called the Common Online Data Analysis Platform (CODAP). This tool has been developed by the Concord Consortium and we think it will provide rich learning experiences for students as it allows students to explore data with different data visualization and measurement tools. It is not hard to learn, we recommend giving students time with CODAP to explore its different features. If you have taken the online course you will have been introduced to CODAP there. Alternatively CODAP provide some training lessons. CODAP runs on Apple and PC products, and on different devices (laptop, ipad, etc).

If you would like to watch leaders from across the world discuss the importance of this approach in mathematics, this Stanford press release includes a short video with a few leaders underlining the importance of this data science K-12 work: https://ed.stanford.edu/news/bringing-math-class-data-age.

As you work on data science explorations with your students, enjoy the freedom that you and they can experience as you all take a pattern seeking approach to mathematics. Join the activities as an explorer looking for hidden treasures in the data – the unnoticed patterns and connections. You don't need to be an expert and have all the answers, instead explore with the students, and see what you can all find out together, as you take on the role of data explorers.

Unit 1:
Data is Everywhere.

Lesson 1: Dear Data

Time	Activity	Materials
10 min	**Mindset messages (optional):** • Choose and show a mindset video to encourage struggle and embrace mistakes.	www.youcubed.org
20 min	**Data is everywhere:** • Share the following: We live in a really interesting time - In 2018 it was found that: 90% of the world's data was created in the last 2 years. • Ask students to think about that statistic - is it surprising? What do they think are examples of the data that is being talked about? Why is there so much more data now than 2 years ago? • Give out the Data Handout or show it on a slide. • Ask students to discuss in groups • Discuss the group findings as a class. Make sure to share that this is called "big data" and learning to make sense of this data is an important part of being a 21st century citizen.	• Data Handout
10 min	**Dear Data Introduction** Introduce the Dear Data Project: • Dear Data was a yearlong data drawing project by Giorgia Lupi & Stefanie Posavec. The 2 women lived on opposite sides of the Atlantic, one in New York and one in London. • Every week they decided to capture data about their lives and send it to each other. One put her data visualizations in a blue US mailbox and one in a red UK post box. • At the end of the year they made a book and a website. http://www.dear-data.com/theproject • The different data visualizations are creative and helpful for students to see, to expand their ideas about what data can be, and also to show them how 3 variables can be captured in one data visualization.	• Interesting optional youtube video: https://www.youtube.com/watch?v=iqaVe-1MCTlA

20 min	**Dear Data Project:** • Group students in 4's if possible. Use 1 handout for each group. Cut the handout so 1 pair shares Georgia's data and the other pair shares Stevanie's data. Instead of handouts, students can look online at digital images. • Ask pairs of students to spend time getting to know what is represented on the postcard. • Ask both pairs to explain their data visualization to the other pair.	• Dear Data Hand-out, 1 per group of 4, cut so each pair in the group has one data postcard to review. Another option would be to share the images online.
10 min	**Class discussion:** • Conduct a class discussion about the ways the 2 women represented data. • Make sure students realize Georgia and Stefani collected data on 1 topic that varied in 3 or more ways. They used different colors, shapes and sizes to show the different ways their data varied.	• Dear Data handout for projection

30 min	**Dear Data Project:**
	• In groups of 4 ask students to decide as a group something they will record and make into a data visualization.
	  • Decide as a group what you will record
	  • Think of 1 recordable event- (eg who do I see when I look up every second)
	  • Make sure it varies in at least 2 ways (eg child/adult).
	
	• This could be an opportunity for students to reflect on the CoVid19 time they have come through. You may want your students to capture a CoVid day with different events through a typical day.
	
	• Ask the pairs or groups to make a data visualization and share with the class.
		• Chart paper
		• Colored pens or pencils
		• Rulers or scissors
		• The data and visualizations could be created in a google doc or slide presentation

5 min	**Reflection/journal entry:** • What are you thinking about data now? What did you learn today that surprised you?	• Journal
Assignment	**Data diary:** Ask students to collect their own data over a certain number of hours (at least 12 may be good) - and to write down every time they provide someone with digital personal data about themselves (see below for more details). This happens whenever they accept terms for an app and use the app, or when they search on the internet, or with Siri or Alexa etc. Come to the next lesson with a record of all the data they have "given" away in the designated time period.	• Data collection handout

Mindset messages:

Youcubed.org has many videos (https://www.youcubed.org/resource/mindset-boosting-videos/) and posters (https://www.youcubed.org/resource/posters/) that help students know they can learn anything and that mistakes and struggle should be valued. These messages are especially important for the beginning of the year as well as later in the year as a reminder.
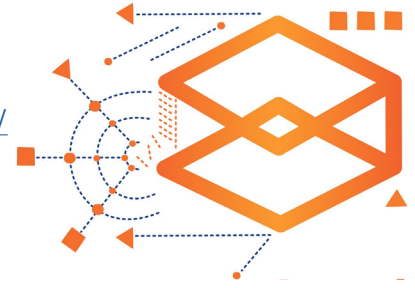
Data is everywhere:

The lesson starts by sharing a really interesting statistic: 90% of the world's data was created in the past 2 years. Students also learn about other "big data" through the handout. This part of the lesson is intended to help students know that they are living in a data filled world, and it is important for them to learn to make sense of data. Explain to them that this will make them more powerful as citizens in the world, and less vulnerable to people who misrepresent the truth.

Dear Data Project:

In the next part of the lesson students are introduced to the "Dear Data" project to help expand their understanding of data and show that data can be so many things, beyond just numbers. It can be words, descriptions, preferences, timestamps, emotions. Data is generated every day, all around us, all over the world, and that's why data science is increasingly relevant and important to understand.
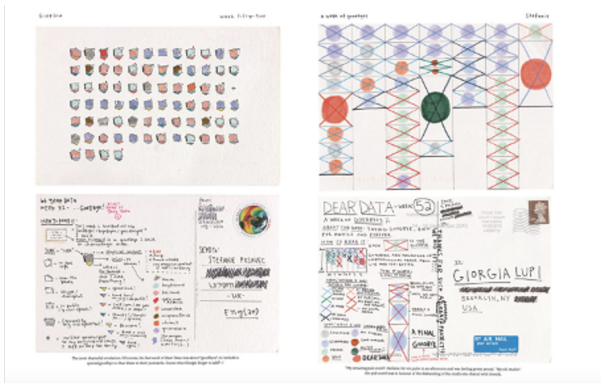
Look fors:

- **Are students expressing curiosity and wonder about data?**
- **Are students starting to notice and chatter about other types of data and interesting ways to visualize data?**
- **Are students wondering about data and how different displays can be misleading or inaccurate?**

Reflection/journal entry:

What are you thinking about data now? What did you learn today that surprised you?

Data diary:

At the end of lesson 1 introduce students to the idea of the "data diary". Agree on a set period of time between lessons - preferably at least 12 hours - when students will write down every time they provide someone with digital personal data about themselves. You may want to ask the students to collectively brainstorm things they could record. The goal of having a 12-hour period is to cover the different facets of students' lives rather than just their time at home or in school. However, feel free to adjust so it does not become a burden for your students.

Lesson 2: What is Data? What is my Data?
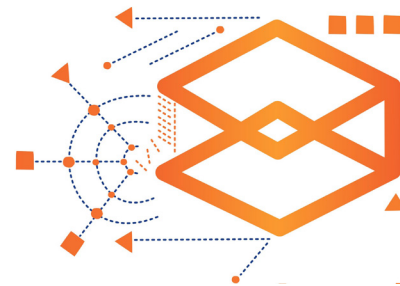
Introduction:

In this lesson, students will be invited to understand the ways data is being collected in the modern world and to broaden their idea of what data is, by collecting data on themselves through keeping a "data diary". Students will work together to explore different ways of representing their data to tell a meaningful, interesting story. This lesson can provide an opportunity to discuss the importance of data privacy and safe internet usage.

| Time	Activity	Materials
10 min	**Data talk:** • What do you notice? • What do you wonder? • What is going on in this graph?	• Change in the working-age population (https://www.nytimes.com/2019/09/19/learning/whats-going-on-in-this-graph-sept-25-2019.html)
10 min	**My data discussion:** • Ask students to discuss the most surprising or interesting row of their Data Diary in groups of 3 or 4.	• Students should each have a page of data they collected
20 min	**My data visualization:** • Ask students to work in pairs or in groups using the data they collected on the data they "give away" – students can either make their own data visualization or make one for each pair or group	• Chart paper • Colored pens or pencils • Rulers or scissors • The data and visualizations could be done in a google doc or slide presentation
5 min	**Reflection/journal entry:** • What have you learned about data visualizations?	• Journal

Data talk:
The driving principle behind selecting visuals for this week was introducing students to a data representation they may not have seen before in order to spark discussion and inspire them for the second part of their Data Diary Activity. Refer to the Data Talks Section below for more information on our selections and for inspiration for your own.

My data discussion:
Students should have completed their Data Diaries. Ask students to discuss the most surprising or interesting row of their Data Diary in groups of 3 or 4.

My data visualization:
In groups, students should take the data they collected from their Data Diaries and create a poster or shared google doc to represent some aspect of the data they all collected. Their first step should be selecting something that interests them about their data. They might want to draw from what they noticed in their discussion at the beginning of the class. Remind students that the poster should represent the thinking of the entire group, and everyone should be included in the poster-making process.
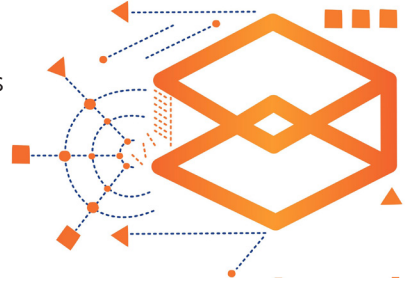
If your teaching is online, you could ask students to create images and add them to a google doc that everyone can view.

This activity is open and gives space for students to explore different ways data can be represented and the reasoning behind choosing certain options. Students discuss and problem-solve with their groups to come up with a representation for their data.

Look fors:
- **Are students thinking about data in creative ways?** Are they only using typical types of displays like line plots and bar graphs?
- **Are they considering quantitative and categorical types of data?**
- **Are students making sense of the color coding they see in other data visualizations?** Are they using color coding in their own data visualizations?
- **Are students including informational text to help the reader understand their data visualizations?**

Reflection/journal entry:
What have you learned about data visualizations?

## Lesson 3: Communicating About Data & Data Visualization
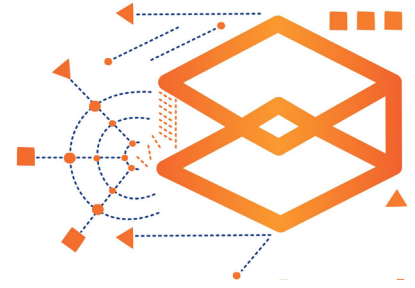
Introduction:

In this lesson, we begin by sharing a picture book we created to engage students in thoughts and conversation around the data they share. After this important data discussion students return to the posters they created the previous day. Their goal is to create a new and improved creative representation of the data.

Time	Activity	Materials
5 min	**A Picture Book Introduction to Data Science:** • Give out picture book - talk in groups about anything of interest to them	https://www.youcubed.org/re-sources/an-introduction-to-da-ta-science-in-pictures/
5 min	**Introduce "creative poster" activity:** • Students will create new posters in the same groups as yesterday, this time choosing what they represent and working out a creative /interesting way of representing the information - examples are given below.	• This could be the making of a shared google doc with all the images together.
20 min	**Groups work on their "creative posters":** Ask students to consider: • What are different ways data can be represented? • What story is your representation telling? Is it interesting? • What information needs to be provided on the display so a reader understands your data story?	• Poster paper • Markers • Meter sticks/rulers
10 min	**Gallery walk:** • Display the posters and ask students to walk around and pause to read and understand each poster. • What information does this data display show?	• This activity could be sharing of a google doc instead of posters.
10 min	**Class discussion:** • Choose one or more posters for a class discussion • You might consider some of the following questions: • What story does this visualization share? • Is the data representation easy to follow? • Do you need any more information to understand the data story?	
5 min	**Reflection/journal entry:** • What have you learned about data visualizations?	• Journal

Picture Book:
Provide students a copy of the picture book about data science. After they
have had some time to look through the book on their own, ask them to
discuss what they learned, or found interesting about the ideas, in small
groups. Some students may be surprised at the information in the picture
book and they may have a lot of questions. We believe this is a very important
and meaningful conversation and one that you may want to continue as you
go through these data science lessons. If you are wanting more information so
you can lead the discussion you will find many articles available about this topic. Add links
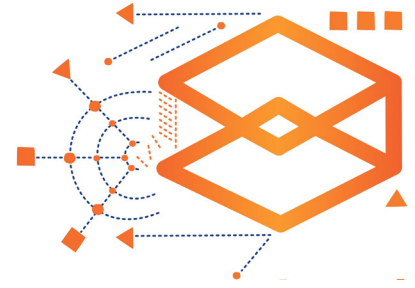
Introduce creative poster activity:
Ask the students to consider the posters they made in the last lesson and to think about the ways they
represented visuals. Students will return to their groups and think about the most creative way they can share
some or all of their group's data. You may want to remind students of the trans-Atlantic data sharing in the
"dear data" project (http://www.dear-data.com/theproject)

Students might add to their original representation or represent something entirely different. In the piloting
of these lessons a group went from representing their screen time data in a bar graph to using color-coded
rings.

Another group initially created a bar graph displaying the group's average screen time each day. They went
back and added lines to the bar graph displaying each individual group member's screen time and color-coded
the line that marked the average.
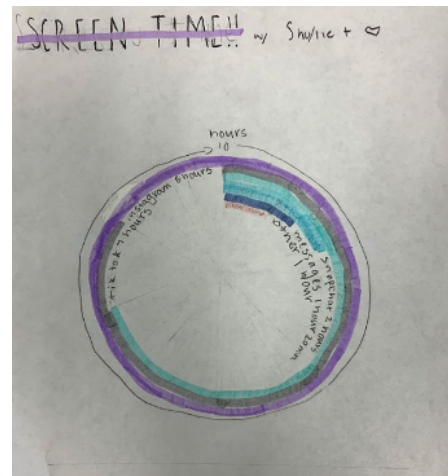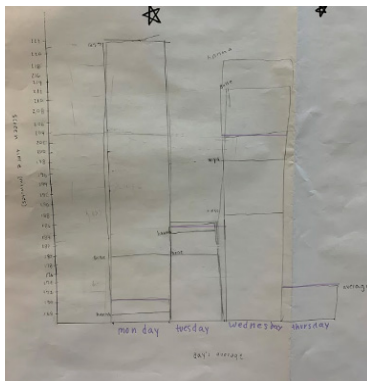
If students are struggling to get started, you might encourage them to look at their data, find something
not represented in their original visualization, and then find a way to share it (as opposed to picking a
representation, and then trying to fit the data to it).

Finally, ask students to share their representations on wall space inside the classroom - or in a google doc.
You may want to ask students to put their two posters side by side so everyone can see how their display
changed for the second version. Ask students to move around the room, spending time at each visualization.
Ask students to notice and wonder about each display. You can invite volunteers from the class to share their
ideas and start a discussion. You can also model what it looks like to say "I don't know!" and get curious about
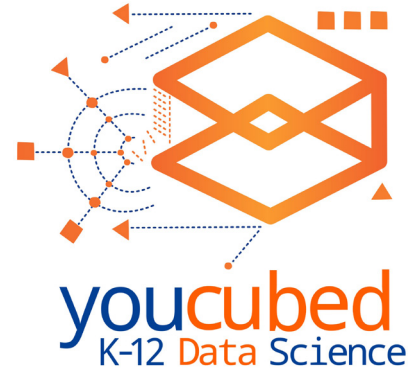finding an answer.

9

Look-fors:

- **Are students drawing on examples from their everyday lives?** A goal of this lesson is to encourage students to see that data science is relevant to their lives. In the "Data is everywhere" discussion and in introducing the Data Diary, you might ask them what things they enjoy doing online (listening to music, playing video games, watching TV, etc), and then discuss what is being captured during that activity. For example, if students bring up watching videos on YouTube, some examples of data that could be captured are video preferences or time spent watching each video.
- **How are students representing their data?** When students are first asked to represent their data, they will use data representations that they are familiar with (ex: bar graphs). It is fine for students to start wherever they are comfortable. This is also a good chance for you to see which data displays they are comfortable with.
- **Are students realizing the way data is visualized impacts the story it tells?** There are countless ways to visualize data beyond the few that are typically covered (bar chart, line graph, pie chart). Later on, as they work to generate new, creative ways of representing their data, you may want to prompt them with questions like:
    - What story is your representation telling? Is it interesting?
    - How should data from multiple sources be combined and reorganized?
    - What information do you need to include with your visual so the reader can understand it?
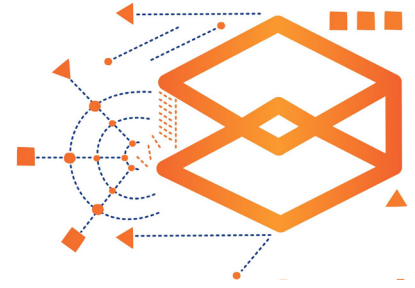
Reflection:

What did you learn from other data stories?

Data Handout

In 2018 90% of the world's data had been created in the past 2 years

Source: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#d6aa12960ba9

Some statistics for you to ponder.

From the Internet:

- We conduct more than half of our web searches from a mobile phone now.
- More than 3.7 billion humans use the internet (that's a growth rate of 7.5 percent over 2016).
- On average, Google now processes more than 40,000 searches EVERY second (3.5 billion searches per day)!
- While 77% of searches are conducted on Google, it would be remiss not to remember other search engines are also contributing to our daily data generation. Worldwide there are 5 billion searches a day.

From Social Media:

Every Minute of the day...

- Snapchat users share 527,760 photos
- More than 120 professionals join LinkedIn
- Users watch 4,146,600 YouTube videos
- 456,000 tweets are sent on Twitter
- Instagram users post 46,740 photos

Facebook:

- 1.5 billion people are active on Facebook daily
- Europe has more than 307 million people on Facebook
- There are five new Facebook profiles created every second!
- More than 300 million photos get uploaded per day
- Every minute there are 510,000 comments posted and 293,000 statuses updated

Instagram, owned by Facebook:

- There are 600 million Instagrammers; 400 million who are active every day
- Each day 95 million photos and videos are shared on Instagram
- 100 million people use the Instagram "stories" feature daily

Communication: volume we send out every minute...

- We send 16 million text messages
- There are 990,000 Tinder swipes
- 56 million emails are sent; worldwide it is expected that there will be 2.9 billion email users by 2019
- 15,000 GIFs are sent via Facebook messenger
- Every minute there are 103,447,520 spam emails sent
- There are 154,200 calls on Skype

Digital Photos:

- People will take 1.2 trillion photos by the end of 2017
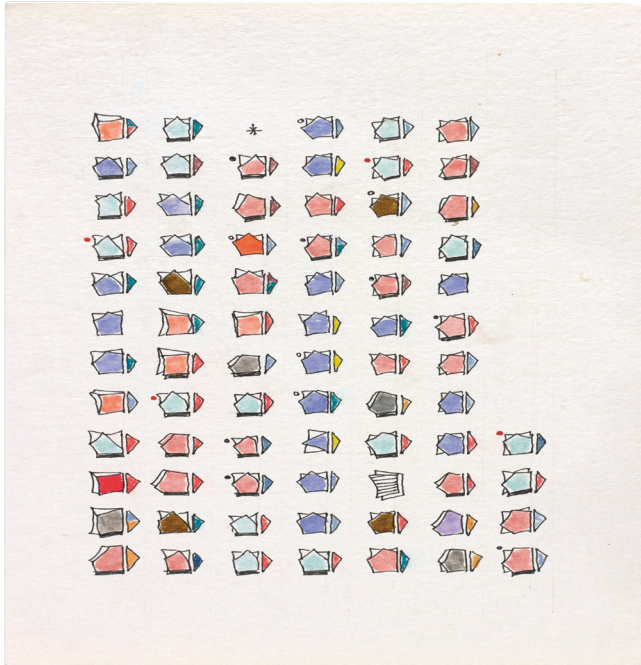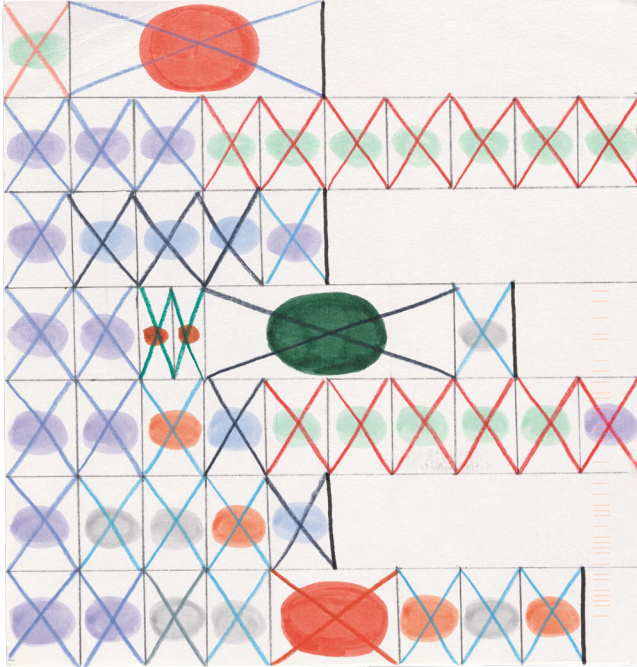- There will be 4.7 trillion photos stored

Services: generated every minute...

- The Weather Channel receives 18,055,556 forecast requests
- Venmo processes $51,892 peer-to-peer transactions
- Spotify adds 13 new songs
- Uber riders take 45,788 trips!
- There are 600 new page edits to Wikipedia

Voice Searches:

- There are 33 million voice-first devices in circulation
- 8 million people use voice control each month
- Voice search queries in Google for 2016 were up 35 times over 2008

Data Diary Handout

Data Diary

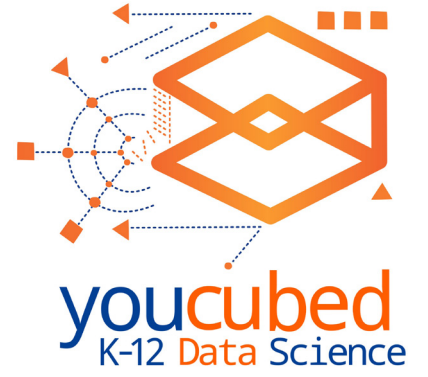| Time	Activity	Type of Data Collected From You

Inspired by IDS from UCLA

Data Talk

Change in the working-age (25 to 54 years old) population from 2007 to 2017