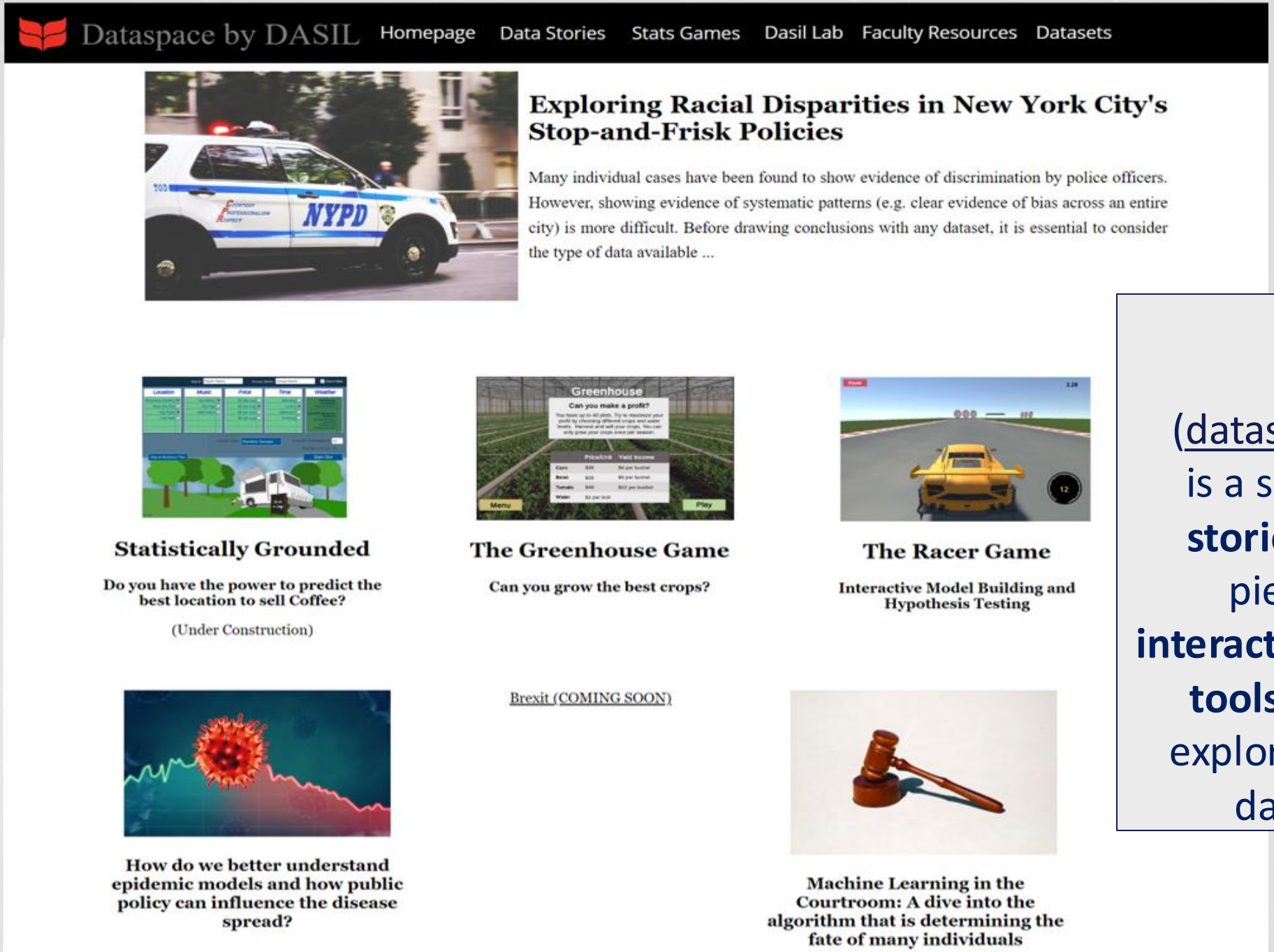
LET'S TALK ABOUT DATA

DATASPACE: USING DATA & STORIES TO BRING DISCIPLINES TOGETHER

Shonda Kuiper
Grinnell College
https://dataspace.sites.grinnell.edu



Dataspace (dataspace.sites.grinnell.edu) is a site for **interdisciplinary stories** and **data journalism** pieces that incorporate **interactive apps** and **pedagogical tools** that allow students to explore, simulate, and modify data from publications

Why use Dataspace?

When teaching statistics or data science:

There is the math part (Concrete Ideas, Follow the Rules)

– Give them the formula, they can get the answer with some training

Then there is the hard part (Requires More Abstraction, Guidelines)

– Putting it all together

Dick De Veaux: Math is Music Math is Music –Stats is Literature

It is relatively easy to teach students a set of rules and calculations

- Students dislike uncertainty
- Students … want to look smart even if it means not learning a thing in the process. —Dweck

It is challenging to coach students to think creatively and critically

- Give space to explore questions and unique approaches
- Articulate uncertainty in context when making decisions (drawing appropriate conclusions)
- Communicate these decisions with their peers

Growth Mindset-Carol Dweck, a professor of developmental psychology at Stanford University and author of "Mindset: The New Psychology of Success."
http://slavieboy.blogspot.com/2011/06/mindsets-and-carol-dweck-live-from.html
http://www.nytimes.com/2014/07/25/opinion/dont-teach-math-coach-it.html?_r=0

Adapting our teaching style to fit the current generation of learners will deeply engage more students.

The strongest students can adapt to any style.

Gen Z students want the content to be meaningful

- Contribute to design/data collection
- Answer is not known prior to the study.
- Authenticity
- Social Purpose
- Instantaneous feedback
- Want to change the world
- Individuality
- Clear examples of the final product
- Not digital experts, but digital dependent

Goals for Dataspace Stories

- Incorporating **data stories** into courses to emphasize **meaningful applications** and conceptual understanding.

- Create materials that are designed to **ease the workload of faculty**

- Provide engaging class activities that could be used on the **first day of class for any data-based undergraduate courses**.

Structure of Dataspace

- **Introductory Article:** We will start with an easy-to-read article

- **Interactive Apps to Investigate Claims**: Throughout the online article, we provide interactive data visualizations, data tables, and/or statistical models to explore claims made in the article.

- **Additional Questions to Investigate**: We will provide additional lists of questions for readers to explore by modifying each of the data visualizations, data tables, or models.

Introductory Article

Homepage Data Stories Stats Games Dasil Lab Faculty Resources Datasets

Exploring Racial Disparities in New York City's Stop-and-Frisk Policies

By Shonda Kuiper. Contributors: Yusen He, Allie Jones, Shreyas Agrawal '24, Bowen Mince '22, Wagih Henawi '22, Adam Solar '22, Ying Long '17, Krit Petrachaianan '17, Zachary Segall '18

Picture by Villa, Giandrea. (2019)

Part 1: Introduction

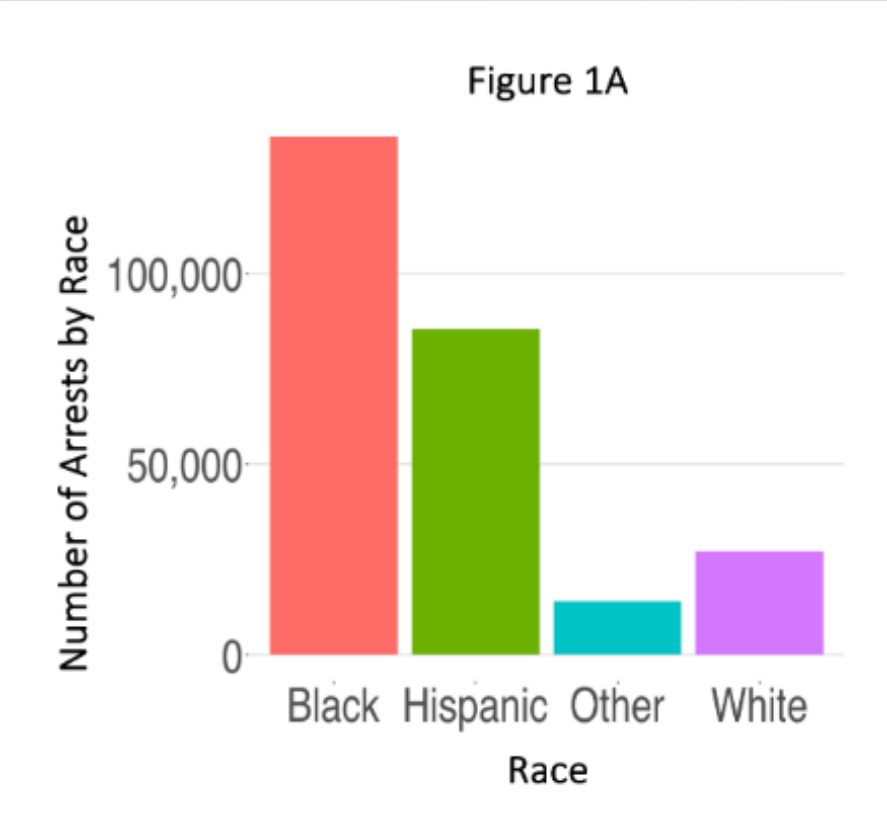
Ronnie is headed home after a day of school. He's got a spring in his step as he makes his way through his Brooklyn neighborhood. As he reaches for his keys, he gets a sinking feeling, realizing he left them on the kitchen table. He shuffles past the bushes and cups his hands to look into the first-floor window checking to see if his younger sister is home yet. As he's fumbling with the front door, two men come up and gruffly ask him what he's doing.

CLAIM 1: There is evidence of racial discrimination in the NYPD stops and arrests.

- Every year, the New York City Police Department (NYPD) stops individuals for suspected criminal involvement.
- "This is a proven law enforcement tactic to fight and deter crime, one that is authorized by criminal procedure law (Long 2009)."
- In recent years, the NYPD had been accused of being racially discriminatory in their stops and arrests.

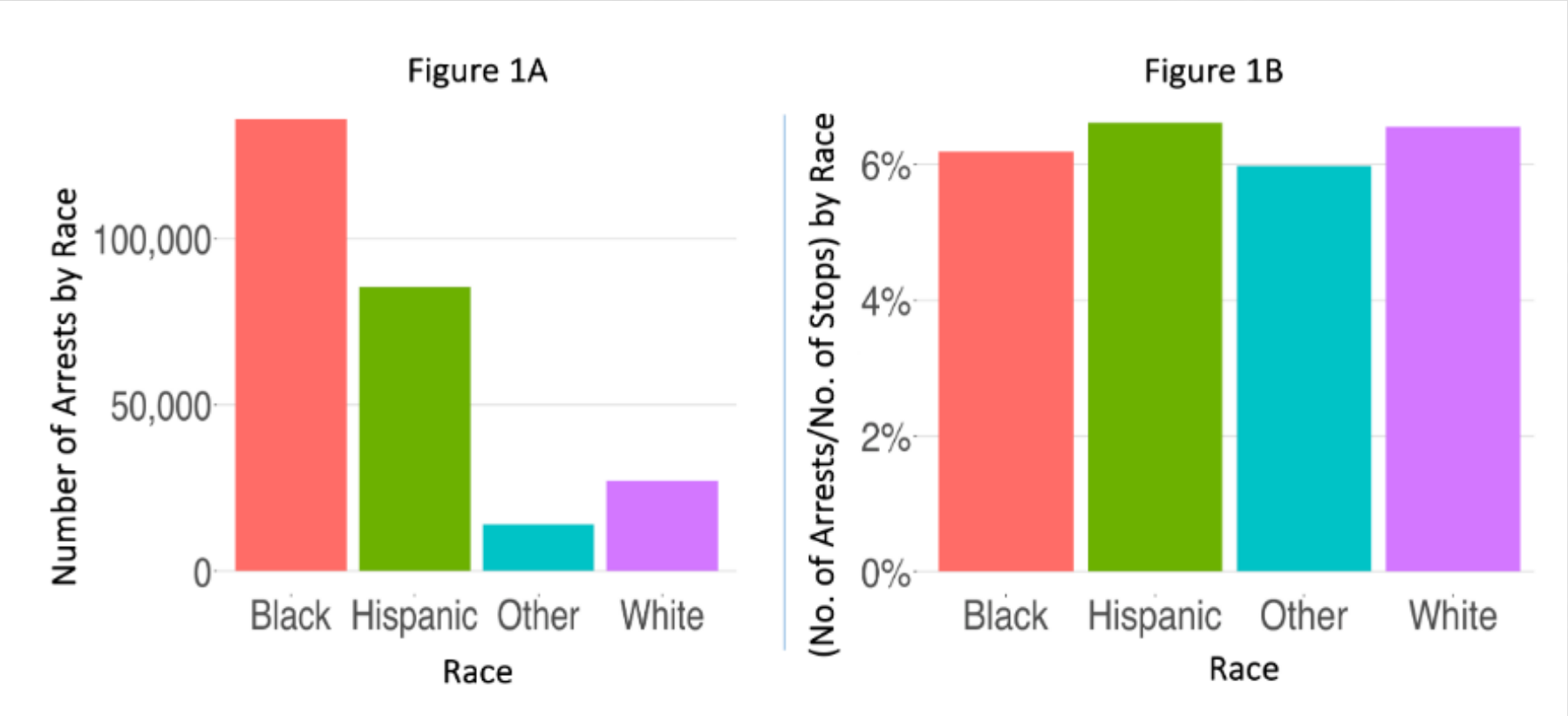
https://dataspace.sites.grinnell.edu/index.html

Introductory Article



- Figure 1A shows the total number of arrests broken down by the race of the suspect.

Introductory Article



- Figure 1A shows the total number of arrests broken down by the race of the suspect.
- Figure 1B shows the percentage of arrests (number of people arrested/number of people stopped) for each race.

Introductory Article

Homepage Data Stories Stats Games Dasil Lab Faculty Resources Datasets

1A: What Does the Data Say?

The data we are working with comes from the New York City Police Department.[3] Every time a person is stopped or questioned by the NYPD, an officer is expected to fill out a police report that records the details of the interaction. Since 2002, there were more than five million stop-and-frisk reports.[4] However, large amounts of data don't necessarily lead to clear conclusions. Let's take a look at just how easy it is to draw incorrect conclusions from large and messy data sets.

The graphs in Figure 1 below visualize the same data from these reports, but in two different ways.

Key Idea:

Percentages can be mathematically accurate, but lead to very different conclusions.

Figure 1A

Figure 1B

Figure 1A shows the ***total number*** of arrests broken down by race of the suspect. This graph clearly tells the story that people of color are much more likely to be arrested than white people.
Figure 1B displays the ***percentage of people stopped***, who were arrested for each race. When we focus only on people who have already been stopped, there is no longer a clear racial disparity in the percentage of those people who are then arrested; (number of people arrested)/(number of people stopped) for each race is similar.

By considering Figures 1A and 1B, we see an example of how two people could reasonably tell two different stories depending on the way they summarize the data.

KEY IDEA: Percentages can be mathematically accurate, but lead to very different conclusions.

- Figure 1A shows that well over 50% of all arrests made in New York involve a person of color.
- Figure 1B shows that just over 6% of people are arrested after they are stopped.
- Whenever we are shown a percentage, we should always ask the question, "Percentage of what?"

Interactive Data Visualizations

ataspace by DASIL | Homepage | Data Stories | Stats Games | Dasil Lab | Faculty Resources | Datasets

1B: Explore the data:

It's one thing to study a graph, but you can really understand the nuance and complexity of the data when you manipulate it yourself! See if you can use the NYPD Bar Chart App to recreate Figure 1A and Figure 1B on your own. Then modify the graph to answer the questions below.

To make a graph that looks like Figure 1A, select:

- Y-axis Variable: ***Arrested***
- X-axis Variable: ***Race***
- Y-axis Measurement: ***Counts***
- Choose years: ***(2006-2018)***
- Facet By: ***None***
- Color By: ***Race***

Instructors Note: Go to faculty resources to access student data

NYPD Bar Charts

Axes | Filters

Y-axis Variable: Arrested

X-axis Variable: Race

Y-axis Measurement

- Counts
- Percentage Of Stops
- Percentage of Arrests

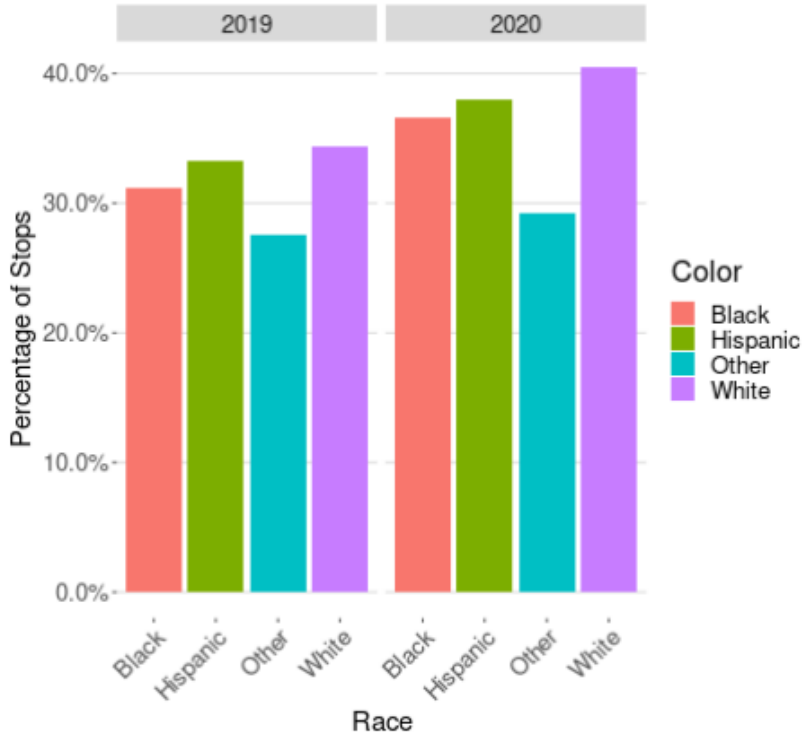
Choose the Years: 2005 — 2019 — 2020

2005 2007 2009 2011 2013 2015 2017 2019 2020

Facet By: Year

Color By: Race

Download



NYPD Bar Charts

Axes | Filters

Y-axis Variable

Stopped

X-axis Variable

Race

Y-axis Measurement

(•) Counts ( ) Percentage Of Stops
( ) Percentage of Arrests

Choose the Years

2005 — 2020

2005 2007 2009 2011 2013 2015 2017 2019 2020

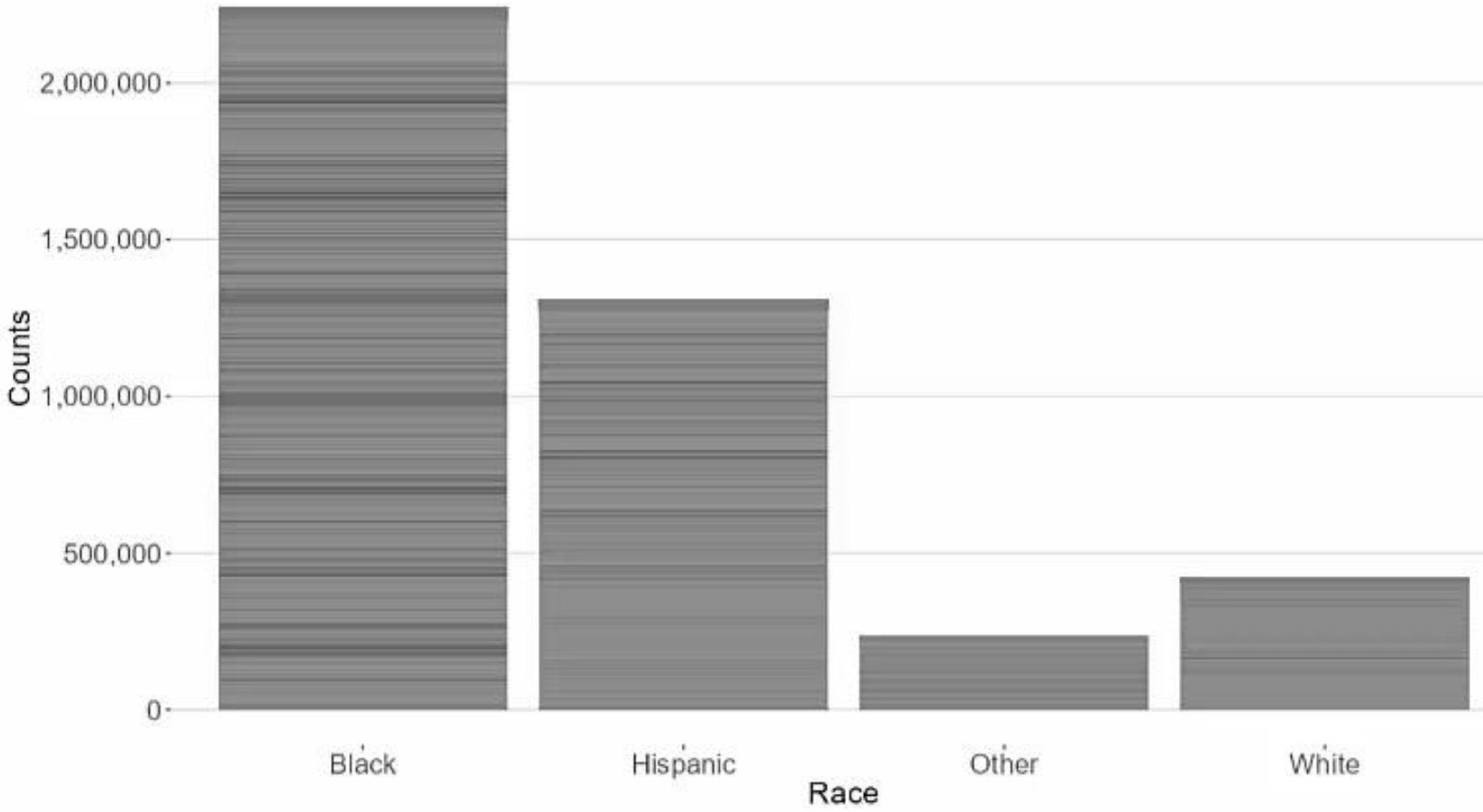
Facet By

None

Color By

None

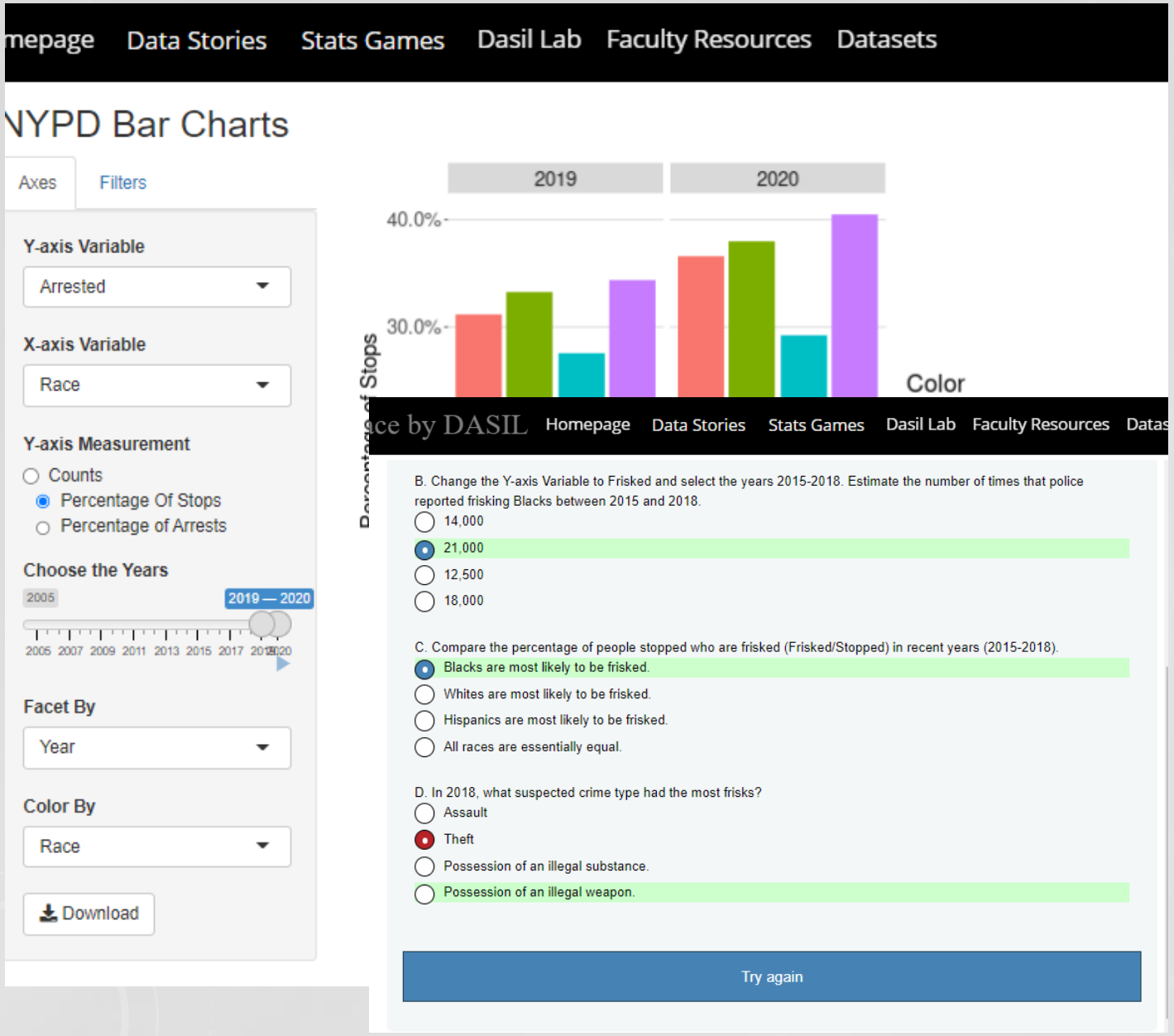
Download



Additional Questions to Consider

- Has the pattern in the percentage of arrests changed over the past ten years?
- Has the amount or type of force used in a stop changed over time?
- Are there disparities in the stop or arrest data related to gender?
- What patterns occur when the data is restricted to a particular type of force, such as restricting the data to only stops where firearms were used?
- What percentage of arrests involved cases where the police drew a weapon (handgun, Taser, pepper spray or baton)?
- Are there any relationships between the types of force used and the suspected crime type? For example, are firearms used more often when the suspected crime is a felony instead of a misdemeanor?

Interactive Data Visualizations



Instant feedback helps students feel confident they are understanding

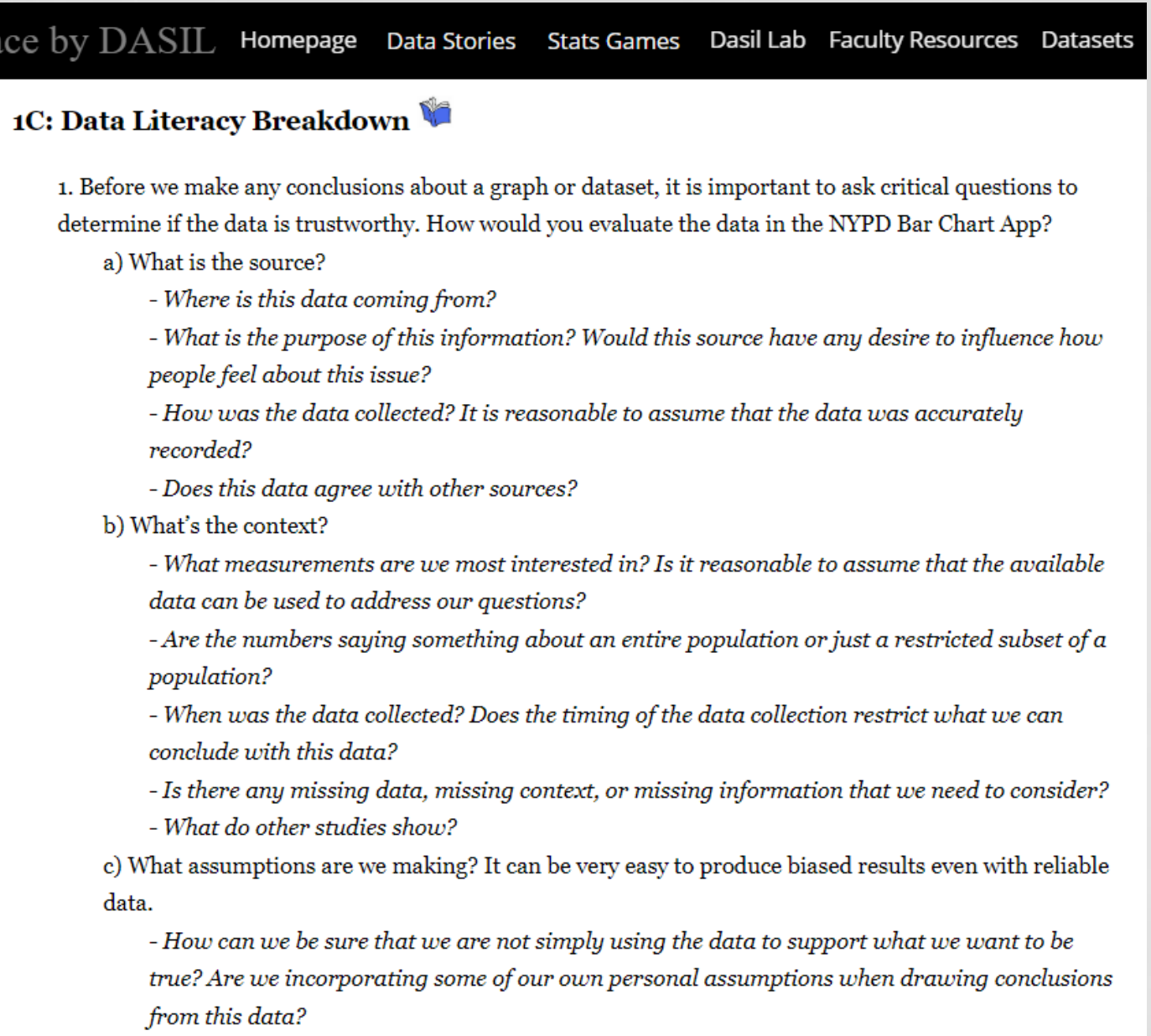
- the app
- counts and percentages

Instructors can automatically download GRADED student answers.

Discussion: optional additional articles

- Data can be easily manipulated to support a particular preconceived notion.
- Summarizing a complex dataset with only one graph (or one hypothesis test) can easily misrepresent the true patterns within the data.
- Objectively look at the entire dataset before drawing conclusions.

Ask the right questions before drawing conclusions



ce by DASIL Homepage Data Stories Stats Games Dasil Lab Faculty Resources Datasets

1C: Data Literacy Breakdown

1. Before we make any conclusions about a graph or dataset, it is important to ask critical questions to determine if the data is trustworthy. How would you evaluate the data in the NYPD Bar Chart App?

a) What is the source?

- Where is this data coming from?

- What is the purpose of this information? Would this source have any desire to influence how people feel about this issue?

- How was the data collected? It is reasonable to assume that the data was accurately recorded?

- Does this data agree with other sources?

b) What's the context?

- What measurements are we most interested in? Is it reasonable to assume that the available data can be used to address our questions?

- Are the numbers saying something about an entire population or just a restricted subset of a population?

- When was the data collected? Does the timing of the data collection restrict what we can conclude with this data?

- Is there any missing data, missing context, or missing information that we need to consider?

- What do other studies show?

c) What assumptions are we making? It can be very easy to produce biased results even with reliable data.

- How can we be sure that we are not simply using the data to support what we want to be true? Are we incorporating some of our own personal assumptions when drawing conclusions from this data?

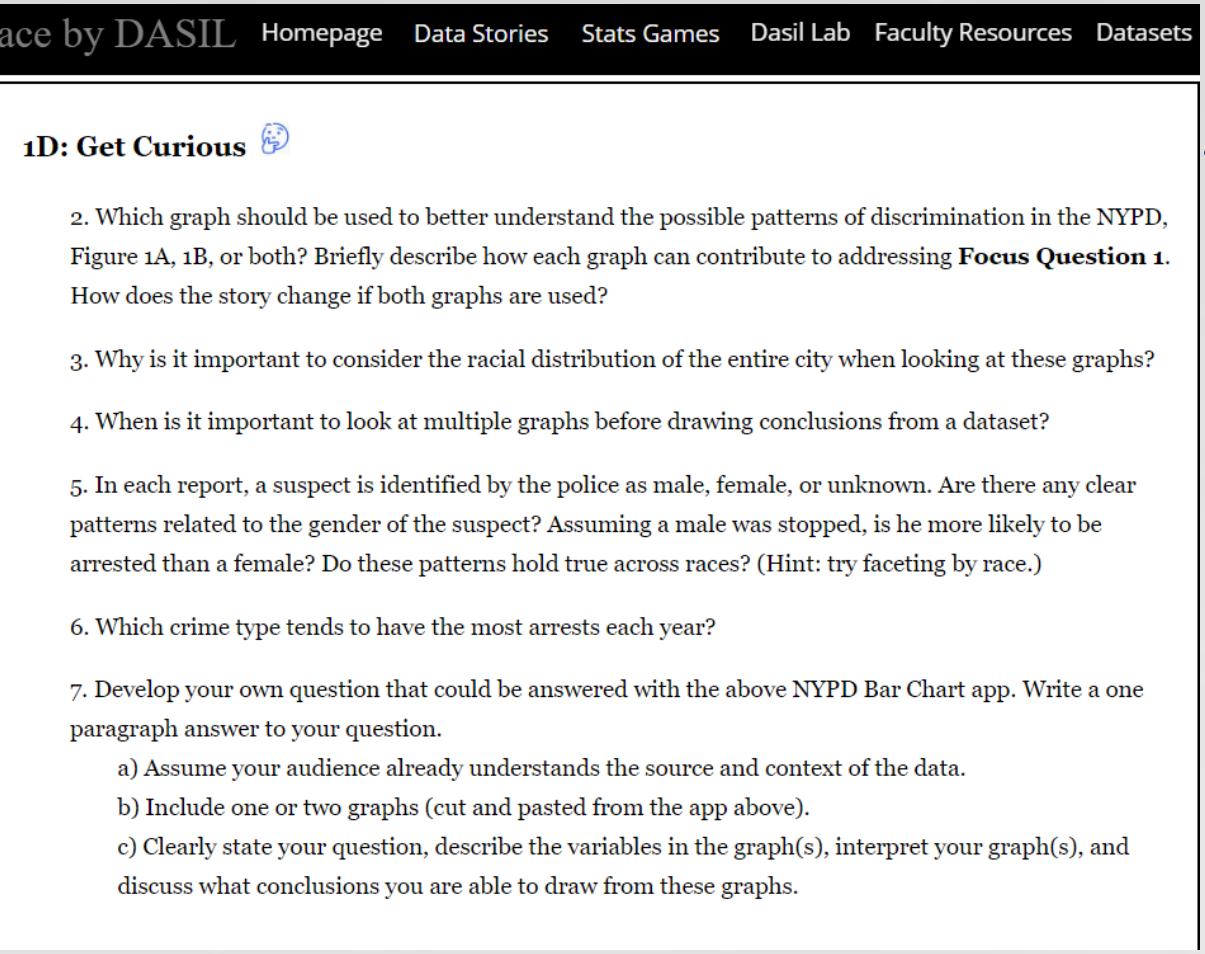
- What is the source?
- What is the context?
- What assumptions are we making?

Connecting math to:

- context of the study
- core reasoning skills

Articulate uncertainty in context when making decisions (drawing appropriate conclusions)

Explore the data



ace by DASIL Homepage Data Stories Stats Games Dasil Lab Faculty Resources Datasets

1D: Get Curious

2. Which graph should be used to better understand the possible patterns of discrimination in the NYPD, Figure 1A, 1B, or both? Briefly describe how each graph can contribute to addressing **Focus Question 1**. How does the story change if both graphs are used?

3. Why is it important to consider the racial distribution of the entire city when looking at these graphs?

4. When is it important to look at multiple graphs before drawing conclusions from a dataset?

5. In each report, a suspect is identified by the police as male, female, or unknown. Are there any clear patterns related to the gender of the suspect? Assuming a male was stopped, is he more likely to be arrested than a female? Do these patterns hold true across races? (Hint: try faceting by race.)

6. Which crime type tends to have the most arrests each year?

7. Develop your own question that could be answered with the above NYPD Bar Chart app. Write a one paragraph answer to your question.

a) Assume your audience already understands the source and context of the data.
b) Include one or two graphs (cut and pasted from the app above).
c) Clearly state your question, describe the variables in the graph(s), interpret your graph(s), and discuss what conclusions you are able to draw from these graphs.

What other information can we gain from this data?

Develop your own question:

- Give space to explore questions and unique approaches
- Communicate these decisions with their peers
- Advanced class: Find an example where two NYPD graphs appear to tell different stories using the same data.

Grinnell DataSpace | RStudio Server | Grinnell DataSpace

dataspace.sites.grinnell.edu/index.html

Registrar | Grinnell... | Imported From IE | JMM Minicourses: T... | Blackboard | stat2labs | Login | Qualtrics | GET TEXT | Stats + Stories | Stats Game Grinnell | Calendar | Amherst | IMS with R

Exploring Racial Disparities in New York City's Stop-and-Frisk Policies

Many individual cases have been found to show evidence of discrimination by police officers. However, showing evidence of systematic patterns (e.g. clear evidence of bias across an entire city) is more difficult. Before drawing conclusions with any dataset, it is essential to consider the type of data available ...

Statistically Grounded

Do you have the power to predict the best location to sell Coffee?

(Under Construction)

The Greenhouse Game

Can you grow the best crops?

The Racer Game

Interactive Model Building and Hypothesis Testing

SELECT A TRACK

Oval Track

Speeding up for the more experienced players!

BACK

RACE

Goals for Dataspace

Integrate examples that are "real to the students" (Gould, 2010)

- Find patterns that matter (tell a story with your data)
- Deeper meaning and insights so that better decisions can be made.

Critical Thinking: (GAISE guidelines)

- Address common misunderstandings
- bias, conditional probabilities, assumptions, abstraction

Transition from small/carefully vetted data to large/messy data (2014 Curriculum Guidelines)

Active Learning (Laursen and Rasmussen(2019))

- Deep Engagement, peer-to-peer learning

Why Change our courses?

- Students have input into the research process and the outcome is not known a priori to either the students or the instructors.. the *study becomes real to the students in very new ways*[1]
- Contribute to a student's *sense of responsibility*, *ownership of his or her piece of the project*, and the importance of his or her *contribution to a broader picture*[2]
- They *take action* based upon those decisions and *defend their decisions* against their peers
- Bridge the gap from smaller, focused textbook problems to real problems.

Learning gains similar in kind and degree to gains reported by students in dedicated summer research programs[2]"

1) Lopatto, D., Undergraduate Research as a High-Impact Student Experience, Association of American Colleges and Universities, Spring 2010, Vol. 12, No. 2, http://www.aacu.org/peerreview/pr-sp10/pr-sp10_Lopatto.cfm
2) Cynthia A. Wei and Terry Woodin Undergraduate Research Experiences in Biology: Alternatives to the Apprenticeship Model, *CBE Life Sci Educ*, Vol. 10, 123–131, Summer 2011

We are looking for story contributors and class testers!

dataspace.sites.grinnell.edu

USING DATA & STORIES TO BRING DISCIPLINES TOGETHER

Shonda Kuiper
Grinnell College
kuipers.grinnell.edu

Students take ownership for their decisions

- Train students to experiment with the data, find their own patterns, and ask their own questions.

  Students learn to handle larger/messier datasets.

  Students have input on what questions are asked even with a common dataset

- Challenge students to think carefully about data and the models they choose to build.

  Interpretation of the numbers is just as important as the calculations!

Learning is essentially hard; it happens best when one is deeply engaged in hard and challenging activities
- Papert

Papert, Seymour (1998, June). Does easy do it? Children, games, and learning. Game Developer Magazine, p. 88.

DataSpace

A place for educators to ease the difficulty that students have in adjusting from traditional homework to real-world case studies and research projects, particularly early in their academic careers.

- **Stories are presented in multiple parts** so that the level of difficulty is **accessible** for students in high school or any undergraduate course.
- **Data visualizations** allow students to clearly see patterns in data and recognize the importance of data in decision-making.
- **Interactive apps** with corresponding **guided questions** train students to ask new questions beyond the initial visualizations.
- User interactivity allows for the exploration of data and encourages **students to ask their own unique questions** about the research and corresponding data.
- **Discussion questions** lead the students to think deeply about properly interpreting data, the impacts of the research, and the importance of clearly communicating results.
- A repository of cleaned and easily accessible data is available for **more advanced work or additional research projects**.