

# DePaul Data Science Summer Academy

Educating Chicago Public High School Students about Data Science



# OBJECTIVES:

- Get high school students (particularly from under-served communities) excited about pursuing data science in college and careers.
- Teach basic data science concepts, tools & techniques with hands on applications for problems students can relate to.
- Introduce students to the wide variety of domain applications and career possibilities in data science through guest speakers.



- Center for Data Mining and Predictive Analytics (DaMPA)
  - Advance research and practice in data science
  - Nurture the next generation of computer and data scientists through innovative programs
  - Build partnerships with industry and non-profit organizations
- DePaul College of Computing and Digital Media (CDM)



# SPONSORS

SIGKDD and DaMPA



**Bamshad Mobasher**

Professor, School of Computing  
Director, Center for Web Intelligence



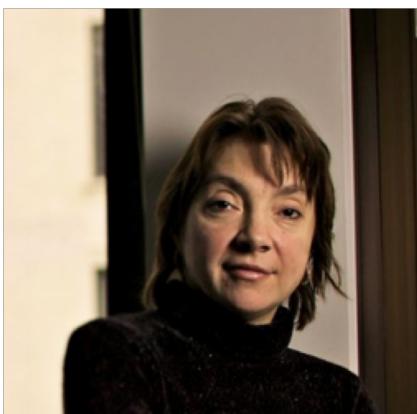
**Lucia Dettori**

Associate Dean, College of Computing  
Liaison with CPS



**Raffaella Settimi**

Associate Dean, College of Computing



**Daniela Raicu**

Professor, School of Computing  
Director, Center for Data Mining and Predictive Analytics  
Co-Director, Medical Informatics (MedIX) Laboratory

# Team

Bamshad Mobasher  
Daniela Raicu  
Raffaella Settimi  
Lucia Dettori



# Recruitment

## Specifically Targeted Chicago Public Schools

- Reached out to CPS high school computer science teachers to share information about our program
- Posted information on CPS online bulletin boards
- 173 applications for 15 spots were received

# SELECTION PROCESS

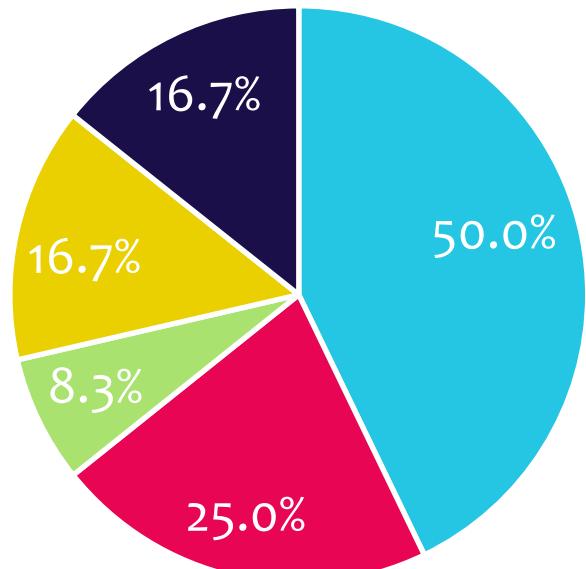
- **SCHOOL District Diversity**
  - 167 high schools in Chicago. Some among the best in the US and some are among the worst.
  - We selected candidates from a wide spectrum of school districts.
  - This ensured broad ethnic diversity as well as a wide range of economic backgrounds among students.
- **GENDER Diversity**
- **DESIRE**
  - Students were required to write a brief essay as to WHY they wanted to attend.
- **GPA**
  - Majority of participants had a GPA over 3.5 (on 4.0 scale). A few with 3.0 GPAs were accepted based on their essays and to increase diversity of schools.



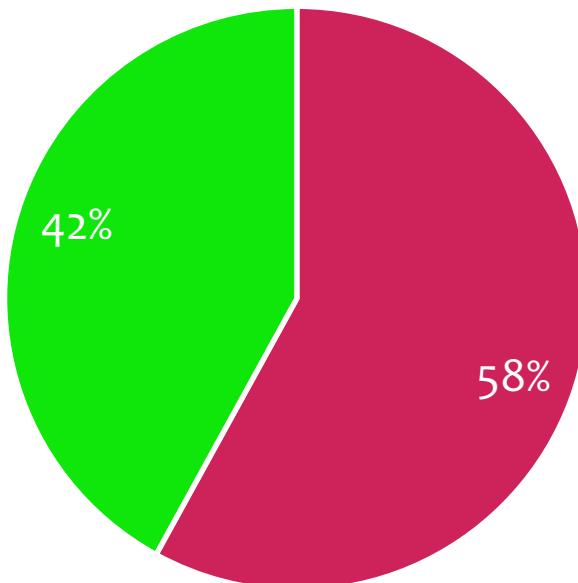
# STUDENT DEMOGRAPHICS

Economic data was not gathered, however 78% of the CPS population are economically disadvantaged

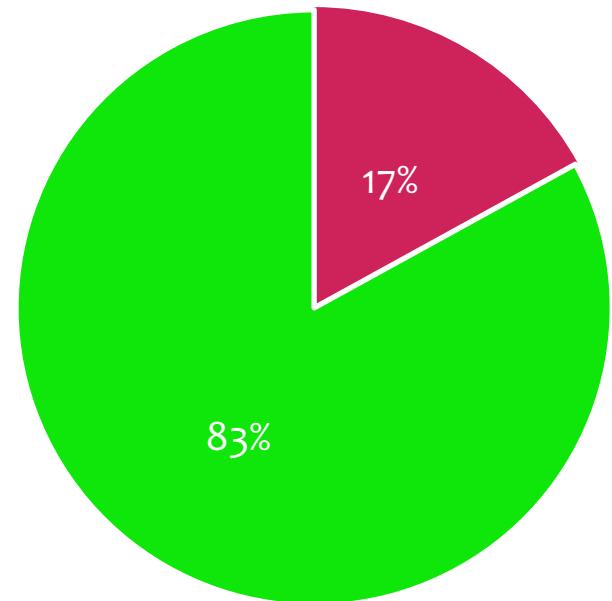
## Ethnicity



## Gender



## Age



- Hispanic   ■ Black   ■ Native Am.
- White   ■ Asian

- Female   ■ Male

- 17 yrs   ■ 16 yrs

Monday	Tuesday	Wednesday	Thursday	Friday
Intros & ABCs of Data Science	Data Exploration & Visualization	Classification Models (decision trees, KNN) & Image recognition	Cluster analysis & Segmentation; Recommender Systems	Student Group Presentations (5 groups)
Research project presentations by DePaul students	Tableau SPSS	Matlab SPSS	Python; Jupyter Notebook	Reception

# CURRICULUM

June 25-29, 2018  
9.30am to 3.30pm



# INDUSTRY SPEAKERS



# TOOLS & DATASETS

## TOOLS

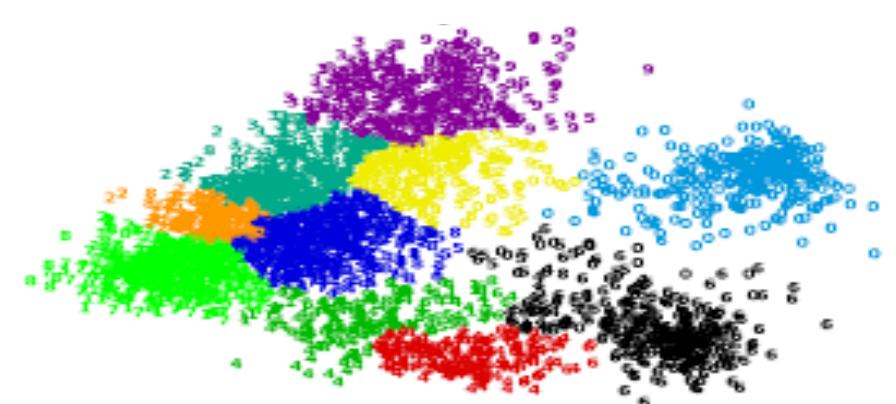
- MATLAB
- SPSS
- PYTHON
- TABLEAU
- SLACK

## DATASETS

- SPOTIFY (Used for overall group projects)
- DIVY (Chicago bike sharing) dataset
- Human Activity Recognition dataset
- ATT face recognition dataset (for images)

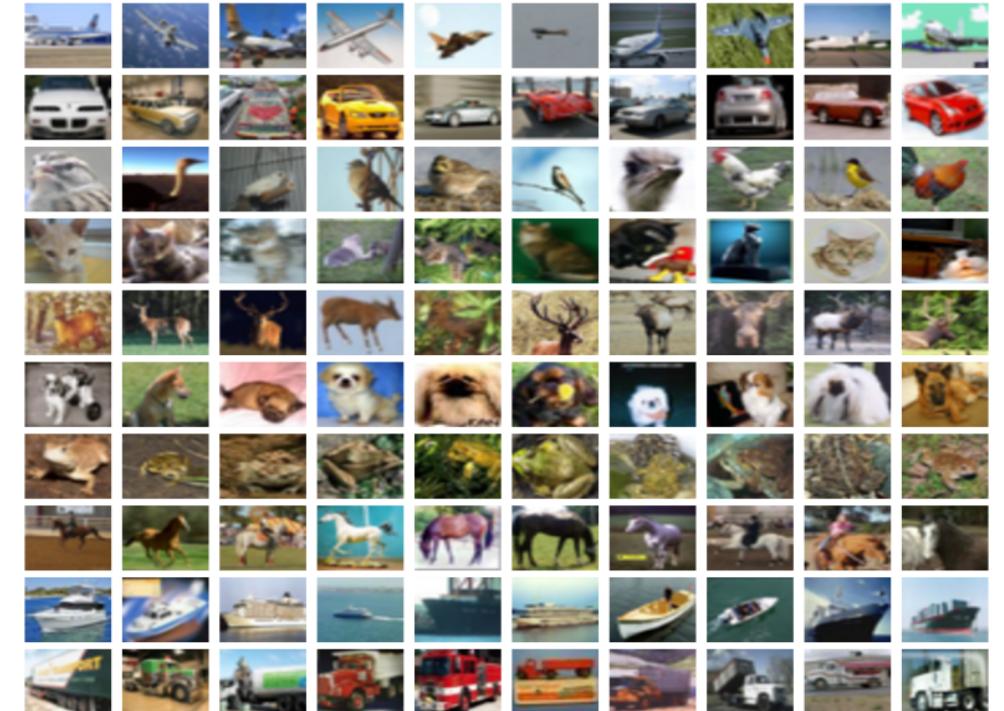
# Tuesday: Data Exploration and Visualization

- Focus on techniques to learn about the characteristics of data, identify general patterns and to visualize trends and patterns
- Tools: SPSS and Tableau
- Lab activity: Analyze Chicago Divvy bikes usage during the past year
- Group project activity: Characterize the Spotify data set



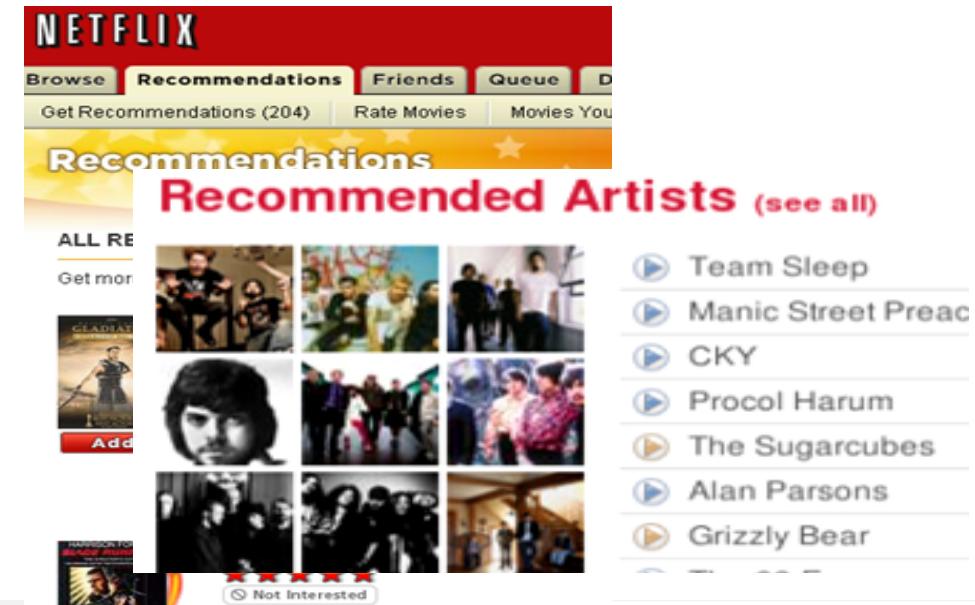
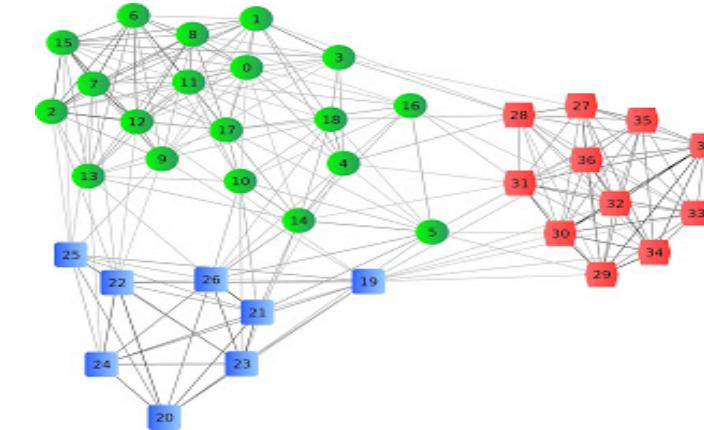
# Wednesday: Classification and Image Recognition

- Focus on discovering patterns in image and video data with applications to human activity recognition
- Tools: SPSS and Matlab
- Lab Activity: extracting pixel data for face recognition used in image classification
- Group project activity: build classifier to Spotify songs into existing categories based on song features



# Thursday: Distance-based Methods and Clustering

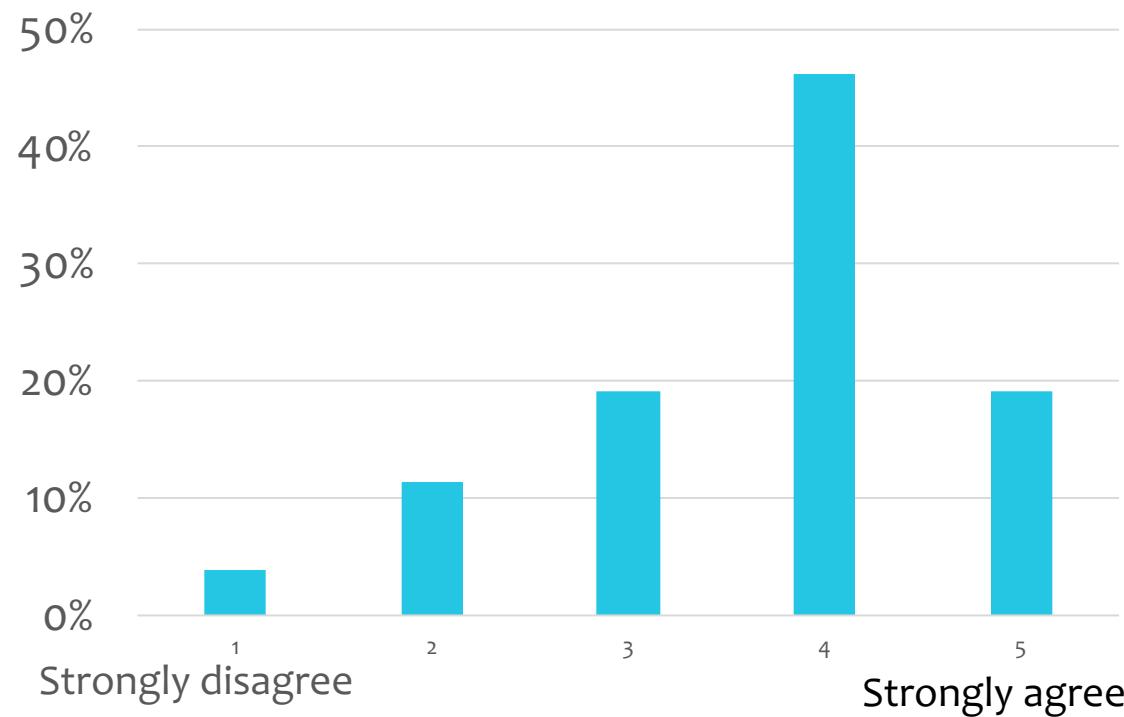
- Focus on applications in recommender systems & music playlist generation
- Jupyter Notebooks, Python, pandas, scikit-learn
- Lab activity using KNN for song recommendation; Kmeans clustering
- Group project activity: performing cluster analysis and visualization on Spotify song data



# STUDENT SURVEY RESULTS

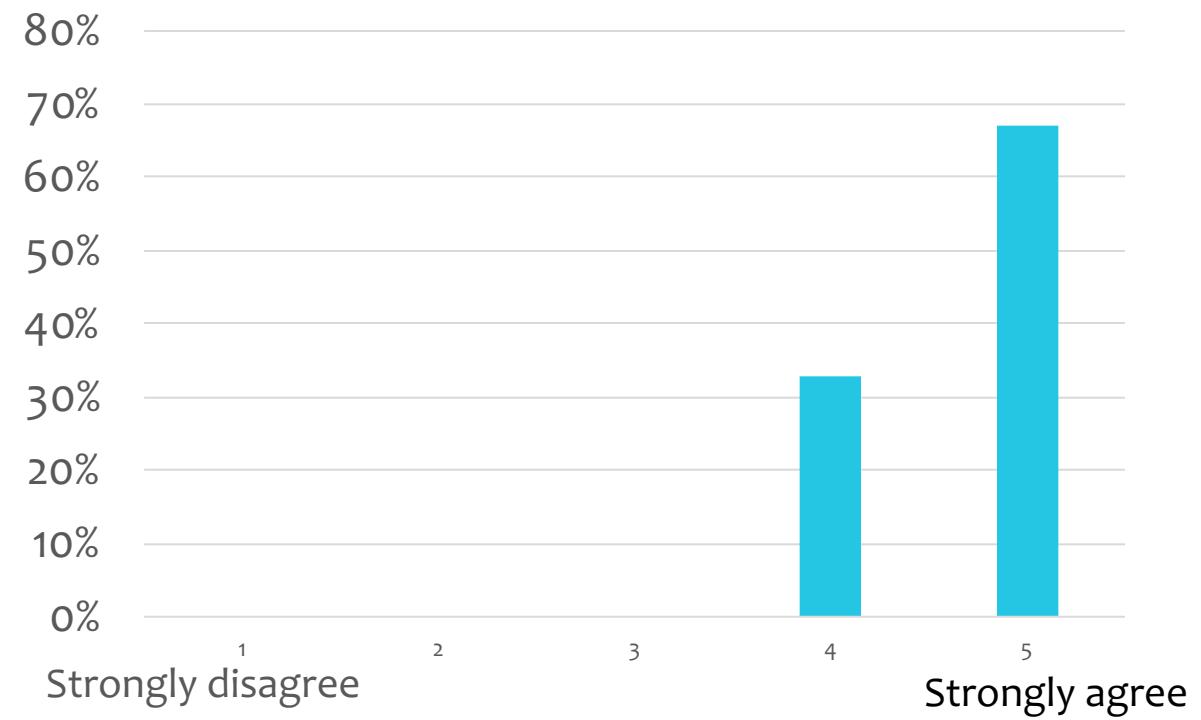
## BEFORE

"I know what Data Science is"



## AFTER

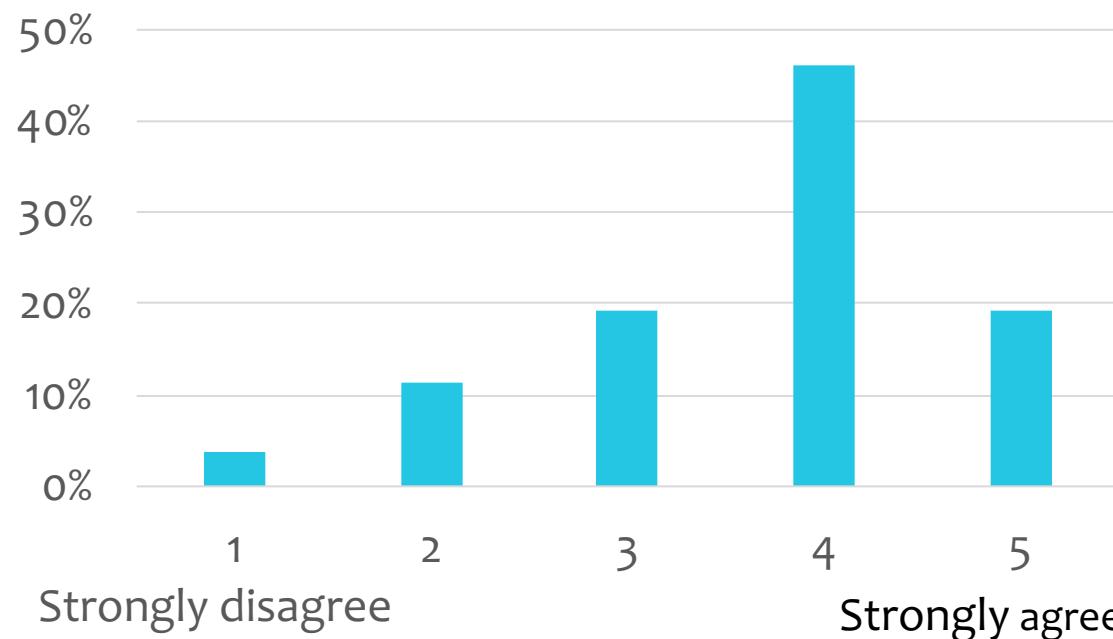
"I know what Data Science is"



# STUDENT SURVEY RESULTS

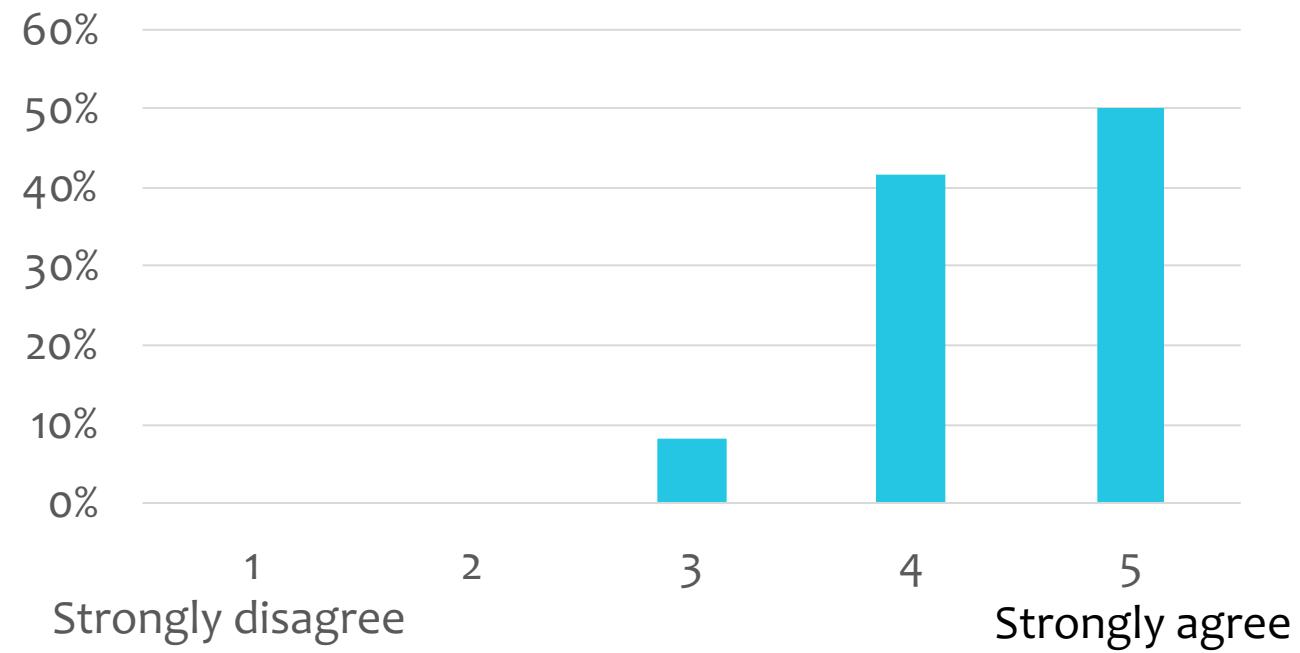
## BEFORE

“I am interested in a career in Data Science



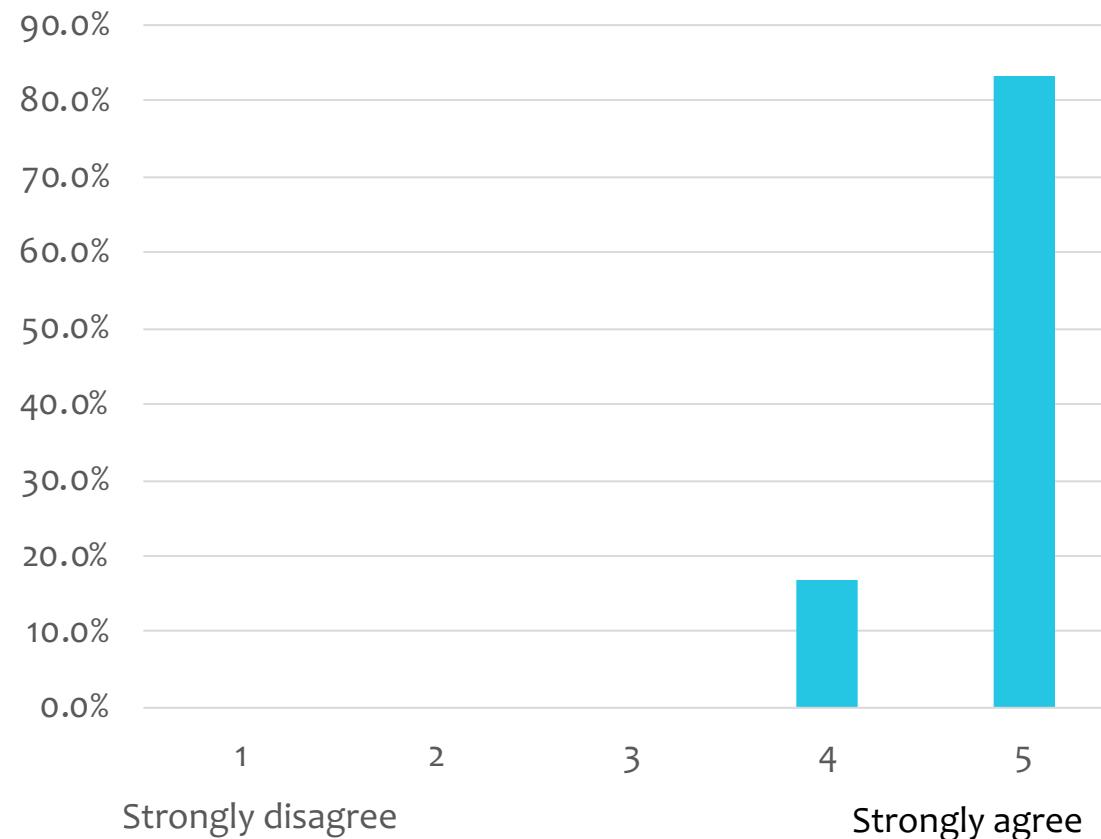
## AFTER

“I am interested in a career in Data Science

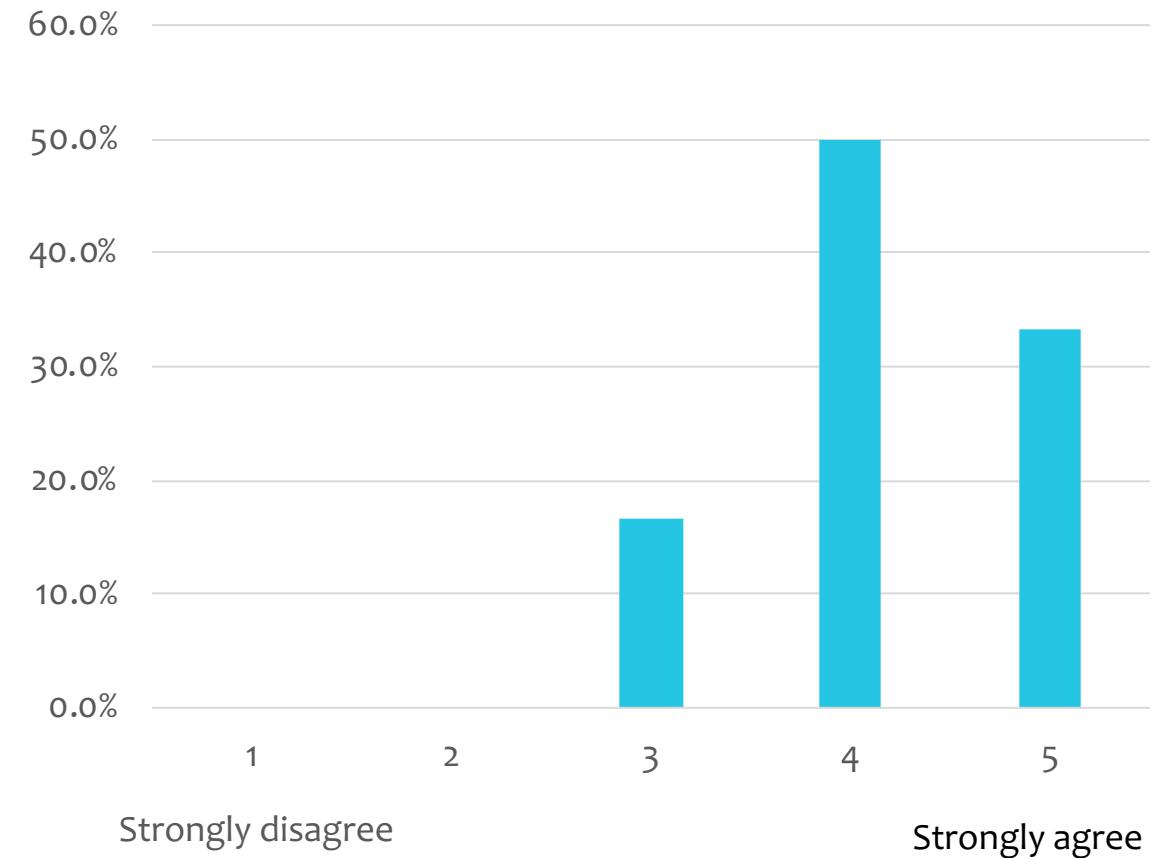


# POST-PROGRAM SURVEY RESULTS

I understand what it means to be a data scientist... what type of work they do.

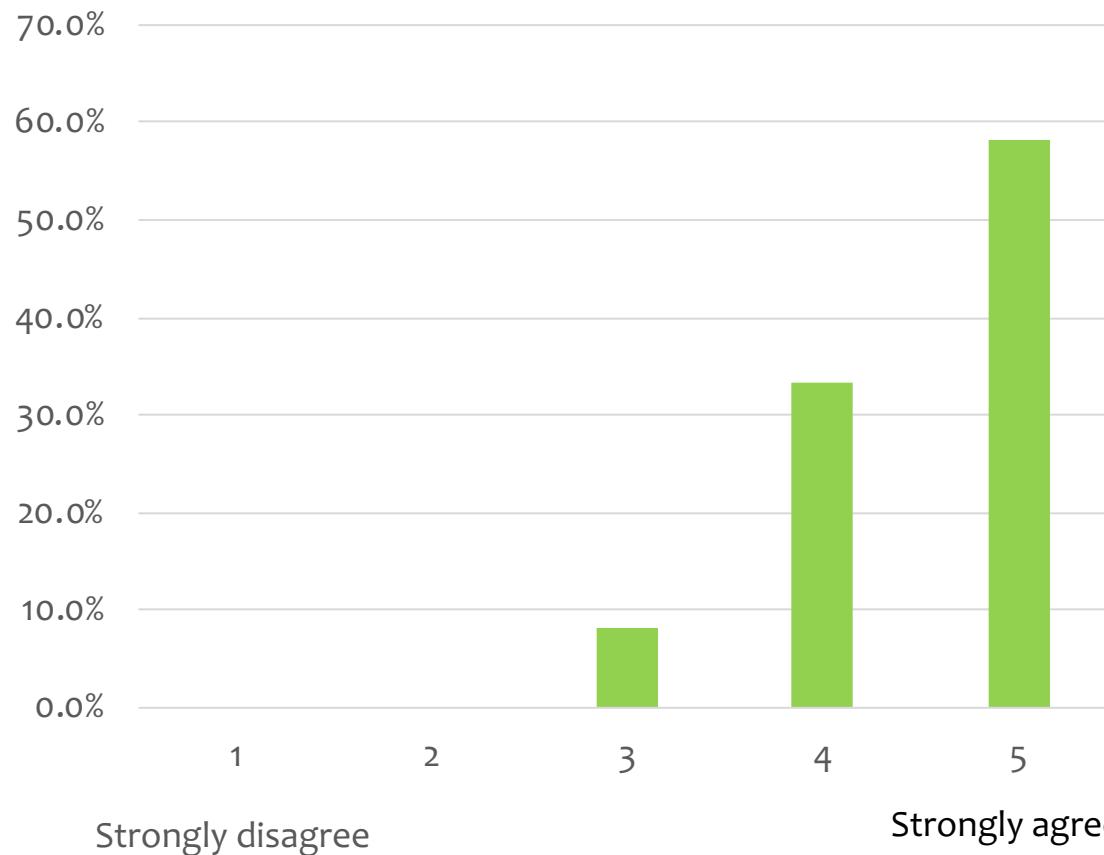


I plan to keep exploring data science in my free time.

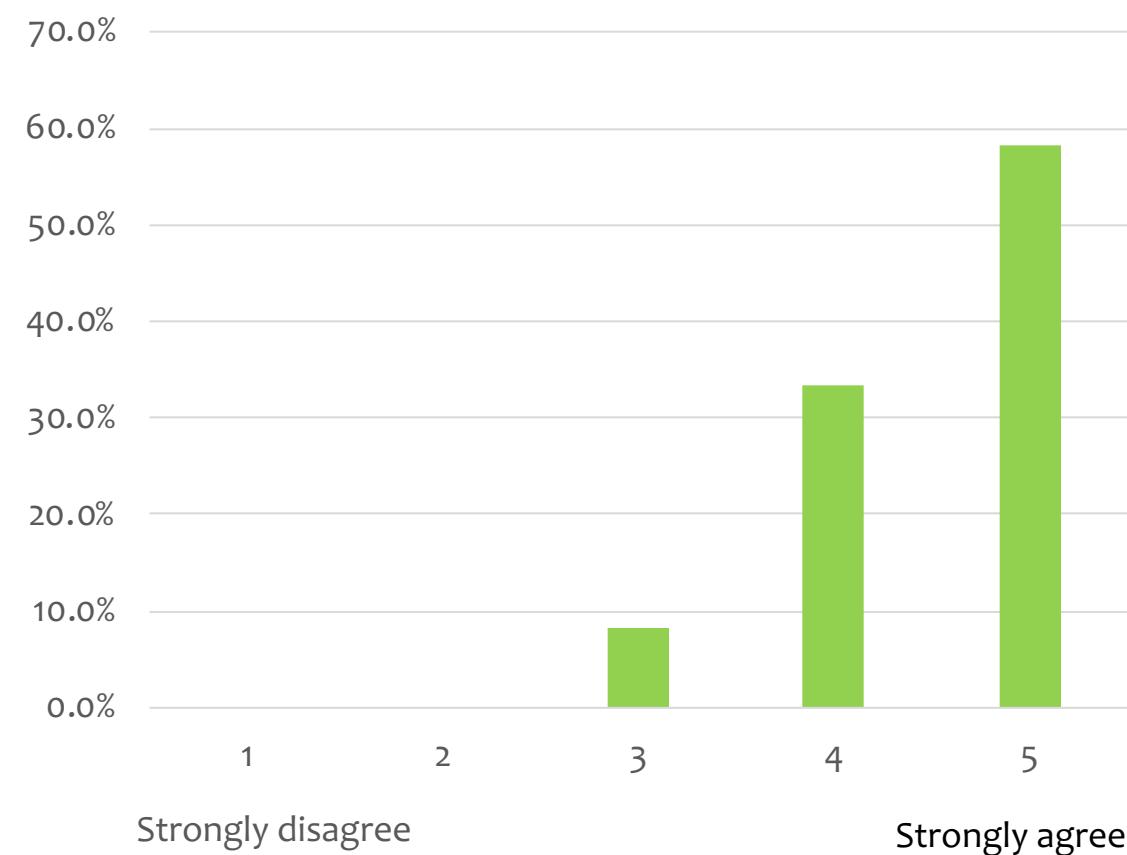


# POST-PROGRAM SURVEY RESULTS

I understand when a classification model should be used (i.e. decision tree).

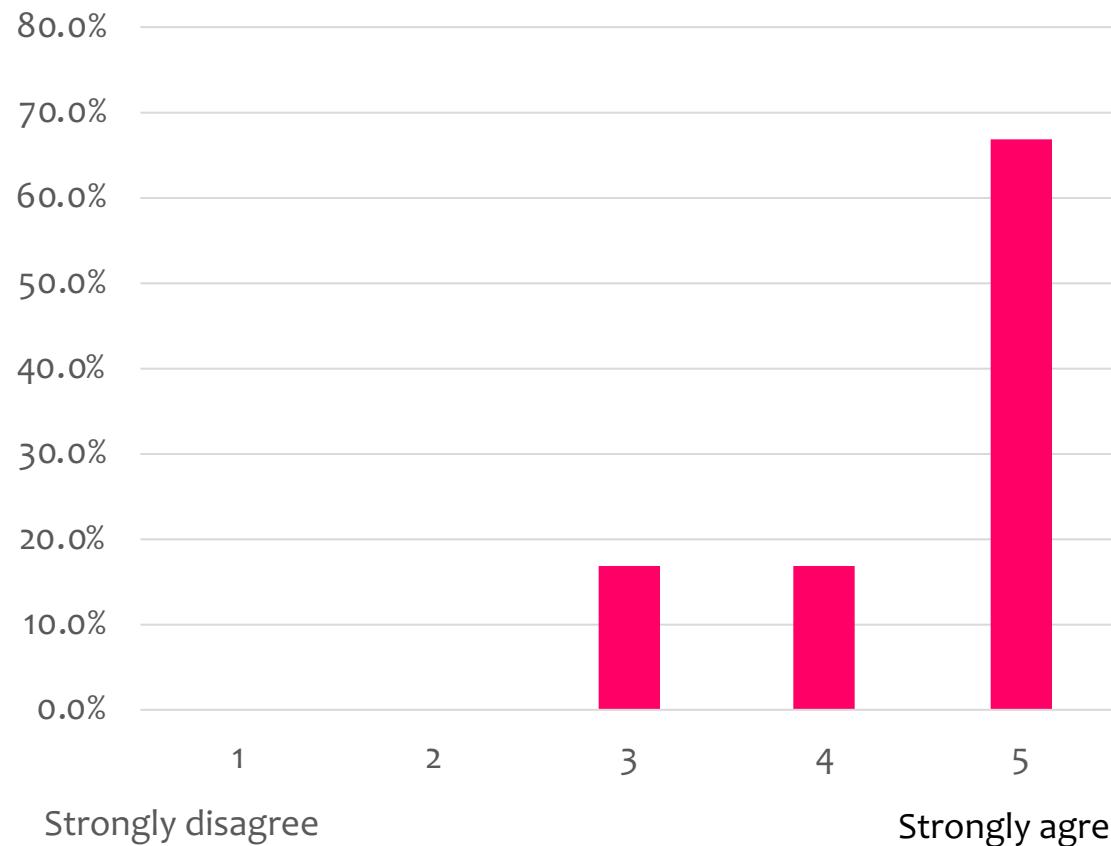


I understand when a distance-based model should be used (i.e. KNN).

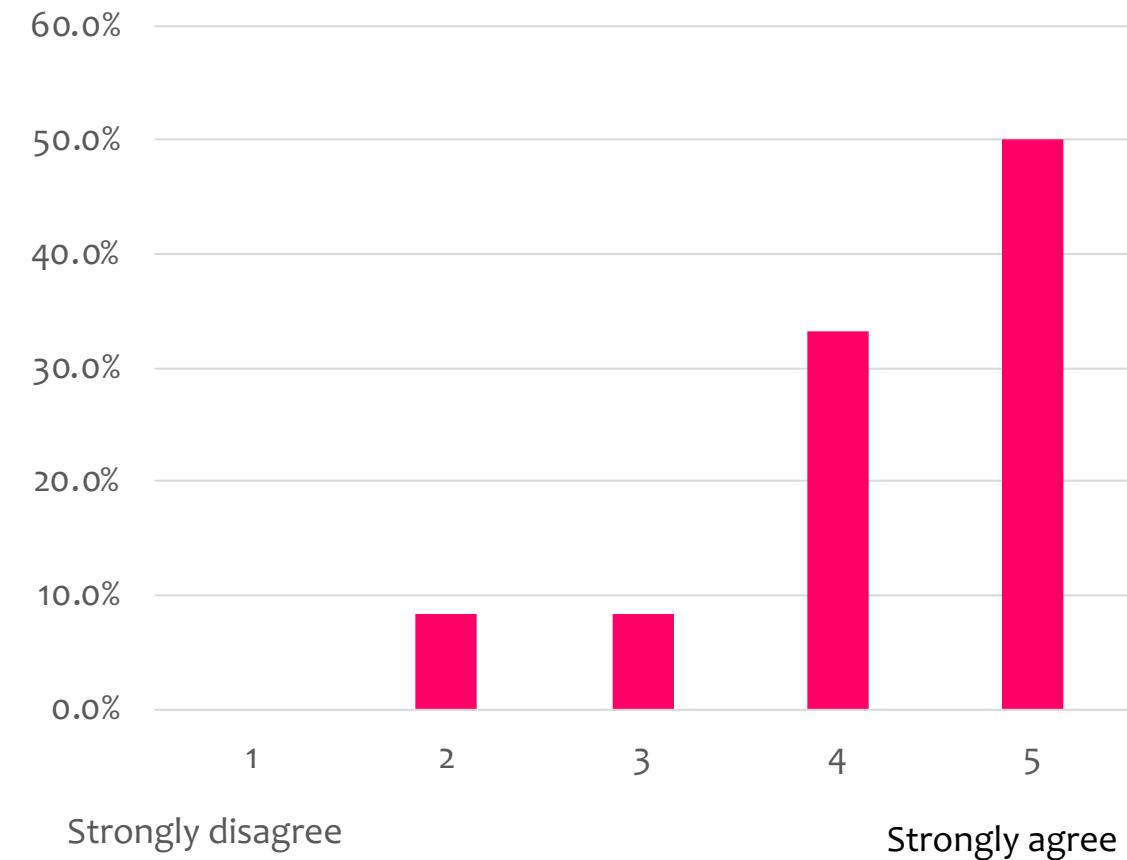


# POST-PROGRAM SURVEY RESULTS

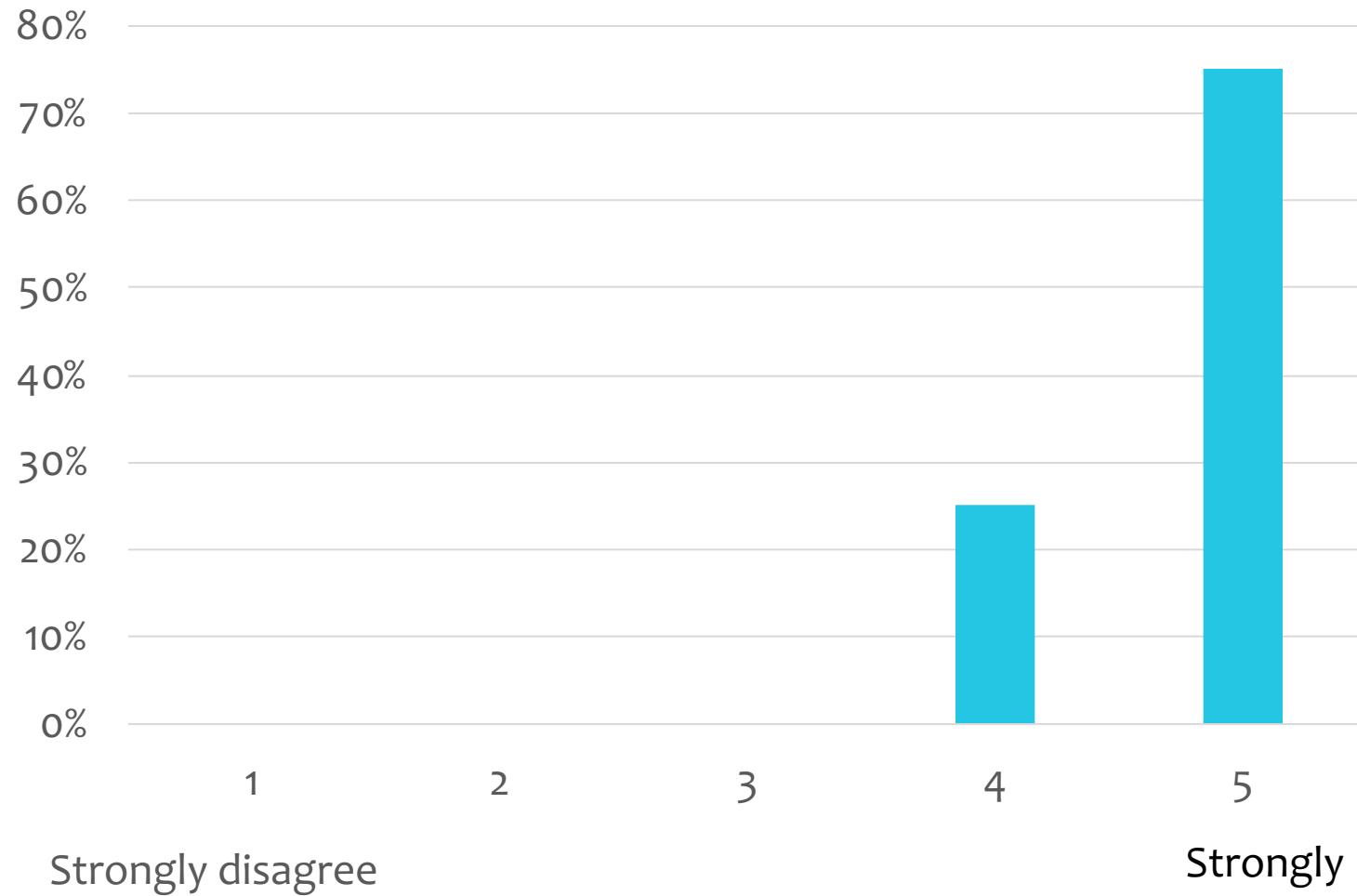
I know how to create a data visualization to create a story.



I understand what clustering is used for.



# SURVEY RESULTS – “I liked this camp” and “I learned a lot at this camp”



## STUDENT COMMENTS

All of the speakers [were my favorite], because they provided a perspective of someone who works in the field and taught us a lot of what they actually did.

I loved learning new applications and ways to analyze and visualize data.

It was fascinating to see how graphs could tell a story about data.

Great program! Wish it was 2 weeks!

Thank you for not dumbing it down!



# PLANS FOR FUTURE

Success of program and the strong demand demonstrates need for additional programs in the future. Especially to provide opportunities for students in economically disadvantaged communities.



# FUTURE PROGRAMS

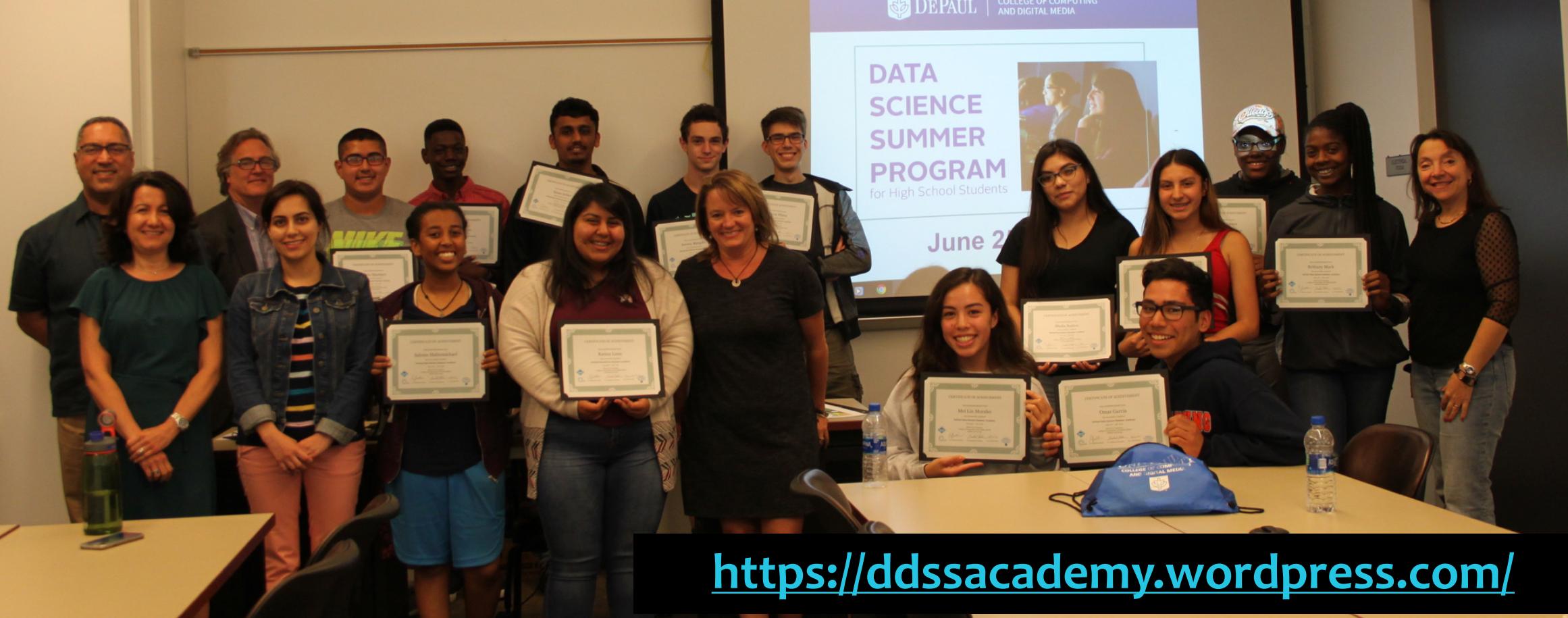
More engagement from Chicago high school teachers in the design of the curriculum

Include other datasets for social good

Even more hands on activities in a variety of settings



# LAST DAY



<https://ddssacademy.wordpress.com/>