



Responding to local data support challenges through a global student DataSquad

LIBER 2022, Odense July 7 2022

Deborah Wiltshire: GESIS Institute for the Social Sciences

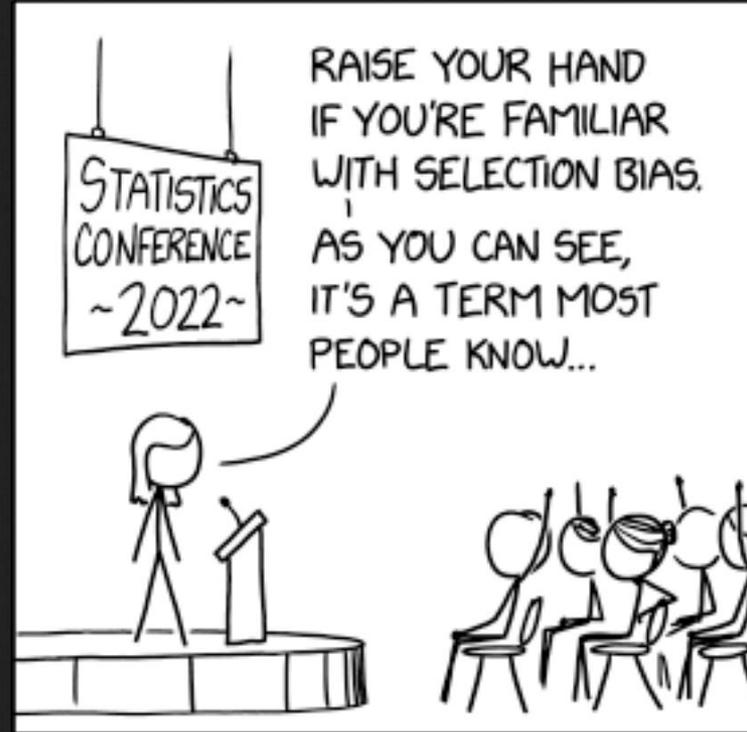
Paula Lackie: Carleton College

Tim Dennis: UCLA

Elizabeth Parke: University of Toronto

Local Data Support Challenges

- The demand for expert data management & manipulation is growing
- Data librarians & data archives role is provide data support
- Demand for support often out strips resources



The dilemma of managing training & recruitment

Issue number one:

- Data librarians & data archives have the skills to provide support but not the time
- We often lack resources
- We often struggle to recruit new people with the right skills

Issue number two:

- Students may receive data skills training but may be gaps
- Theory often differs from the reality!
- Traditional teaching doesn't offer practical work experience
- BUT steady stream of budding data scientists eager for real data experience

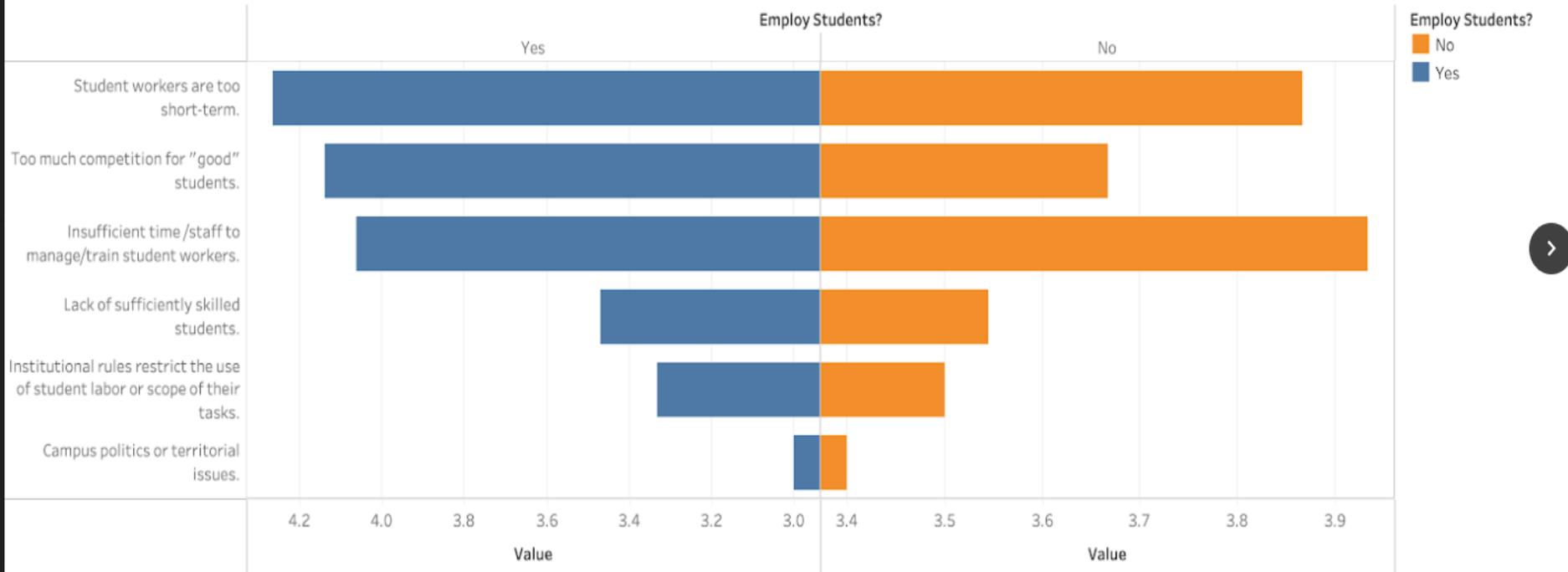
Surveying global data support services

- Carried out a survey in 2021
- 100+ data support services responded from across the global
- Around 70% employed students to help deliver data support services
- Questions focused on the provision of data support services & the barriers to using student labour

Surveying global data support services

Average Rank of Obstacles Grouped by Whether They Employ Students

(5 = Biggest Obstacle, 1 = Smallest Obstacle)



Introducing DataSquad



- Started in the 1990's at Carleton College
- 2014 DataSquad was born
- Helps libraries and data service provide services
- Gives students valuable work experience

[Carleton ITS DataSquad \(carletondatasquad.bitbucket.io\)](http://carletondatasquad.bitbucket.io)

Facilitating Data Processes At Carleton

DataSquad

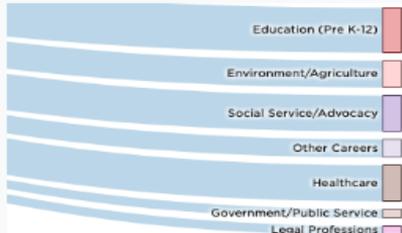
NEED HELP? 
SEND US AN EMAIL

OR 
REACH OUT TO THE QRC



Tableau Presentation & Support for a Sociology Class

*Sociology and
Anthropology Department*



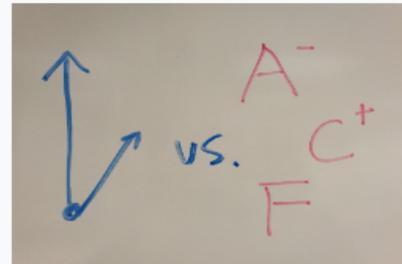
Data visualization for the Career Center

Career Center



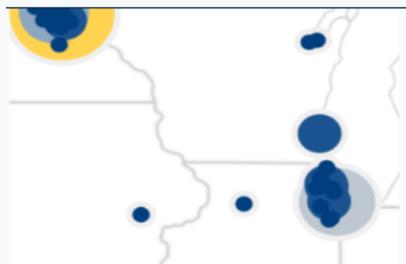
Data Security Service

*Institutional Research
Board (IRB)*



Make Grading Easier

Psychology Department



Visiting Accepted Students Map

Office of Admissions



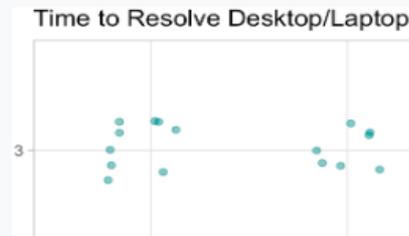
Data Visualization

*ITS Helpdesk and
Academic Technology
Support*



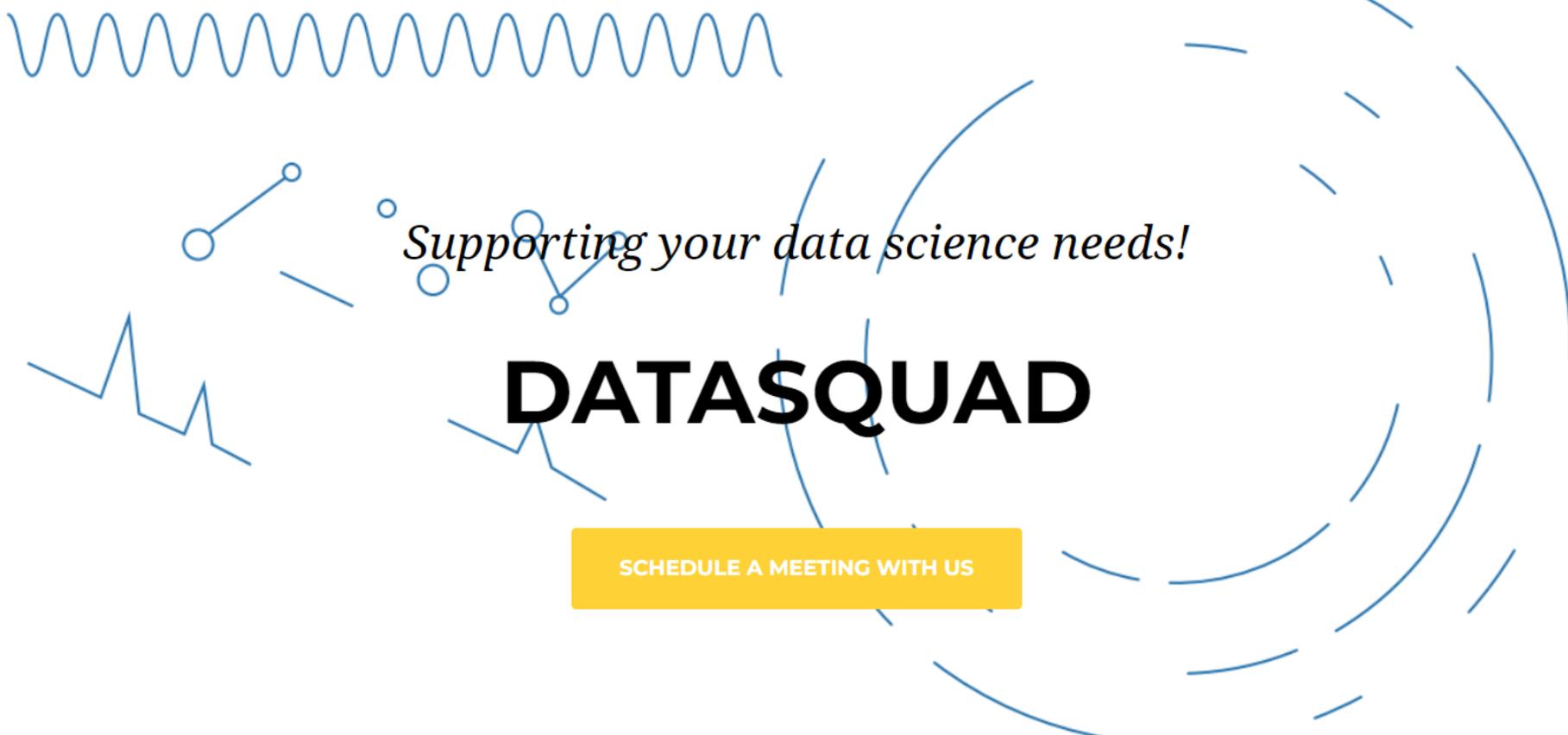
Duplicate Faculty List into WHD Data

Academic Technology



ITS Survey Analysis

ITS

The background features several decorative blue line-art elements: a wavy line at the top left, a network graph with nodes and edges in the middle left, a jagged line graph at the bottom left, and several concentric curved lines on the right side.

Supporting your data science needs!

DATASQUAD

SCHEDULE A MEETING WITH US

Dr. Terence Keel: Analyzing Civilian Deaths at the Hands of Police

Consultant: Ethan and Julia

Data was collected from the community where civilians submitted their own confessions and autopsies were analyzed to understand what was being written as cause of death.

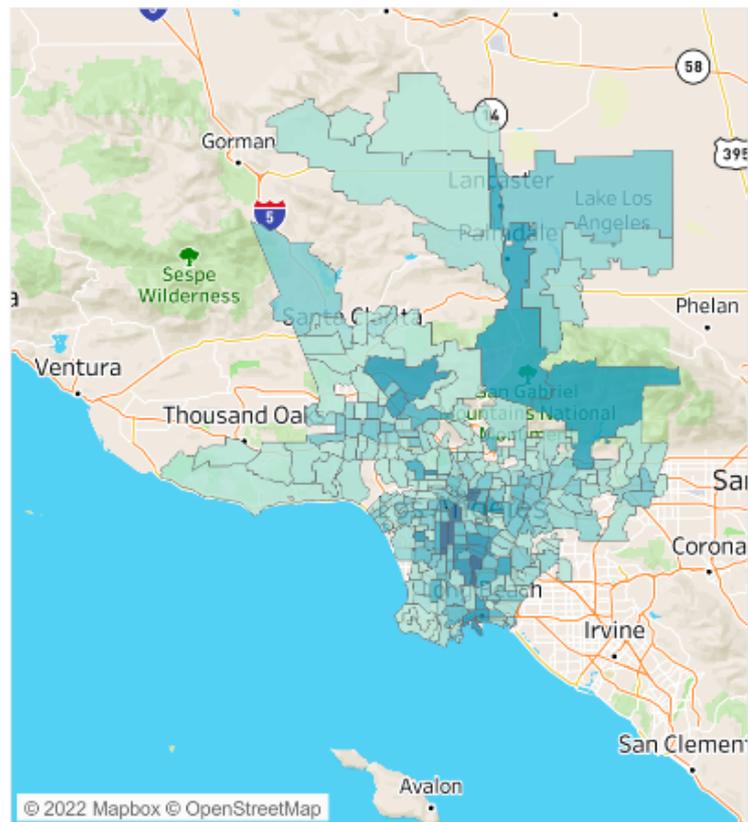
- **Duties:**

- Gathered and cleaned data to get the names of civilians

- Compare between two datasets to find new names that were previously hidden

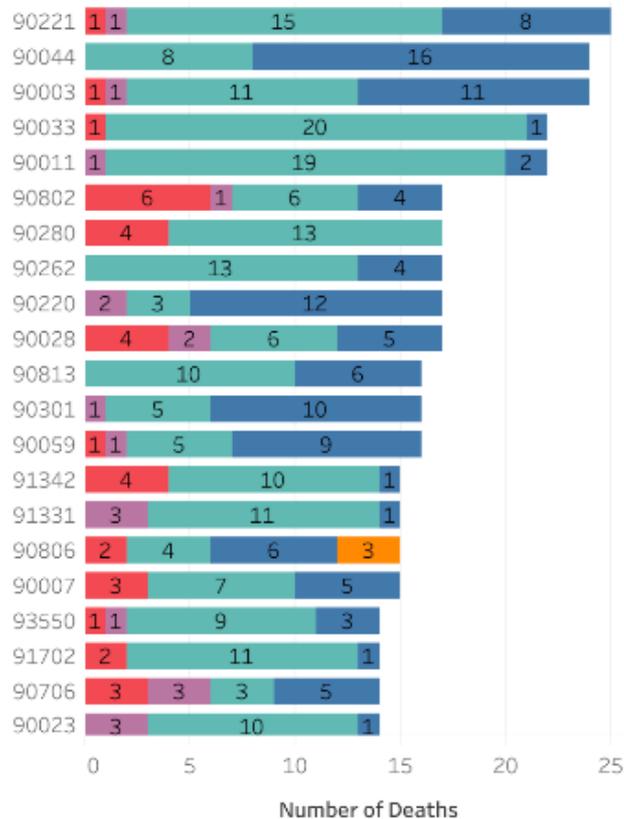
- Conduct Chi-squared tests to determine which states or localities had a different number of previously unreported deaths

Map of Fatal Encounter Deaths in Los Angeles (2000-Feb2022)



Deaths by Race & ZipCode in LA Bar Graph

Location of ..



Number of Deaths
1 25

Race with imputations

- Asian/Pacific Islander
- Black
- Hispanic/Latino
- Middle Eastern
- Native American/Alaskan
- Race unspecified
- White

Kate Mcinerney: Helicopter Flight Data

Consultant: Ethan

Initial consultation used to determine how to get location address from latitude and longitude and convert time into correct timezone

- **Duties:**

- Get a unique flight ID to join the position and flight data frames

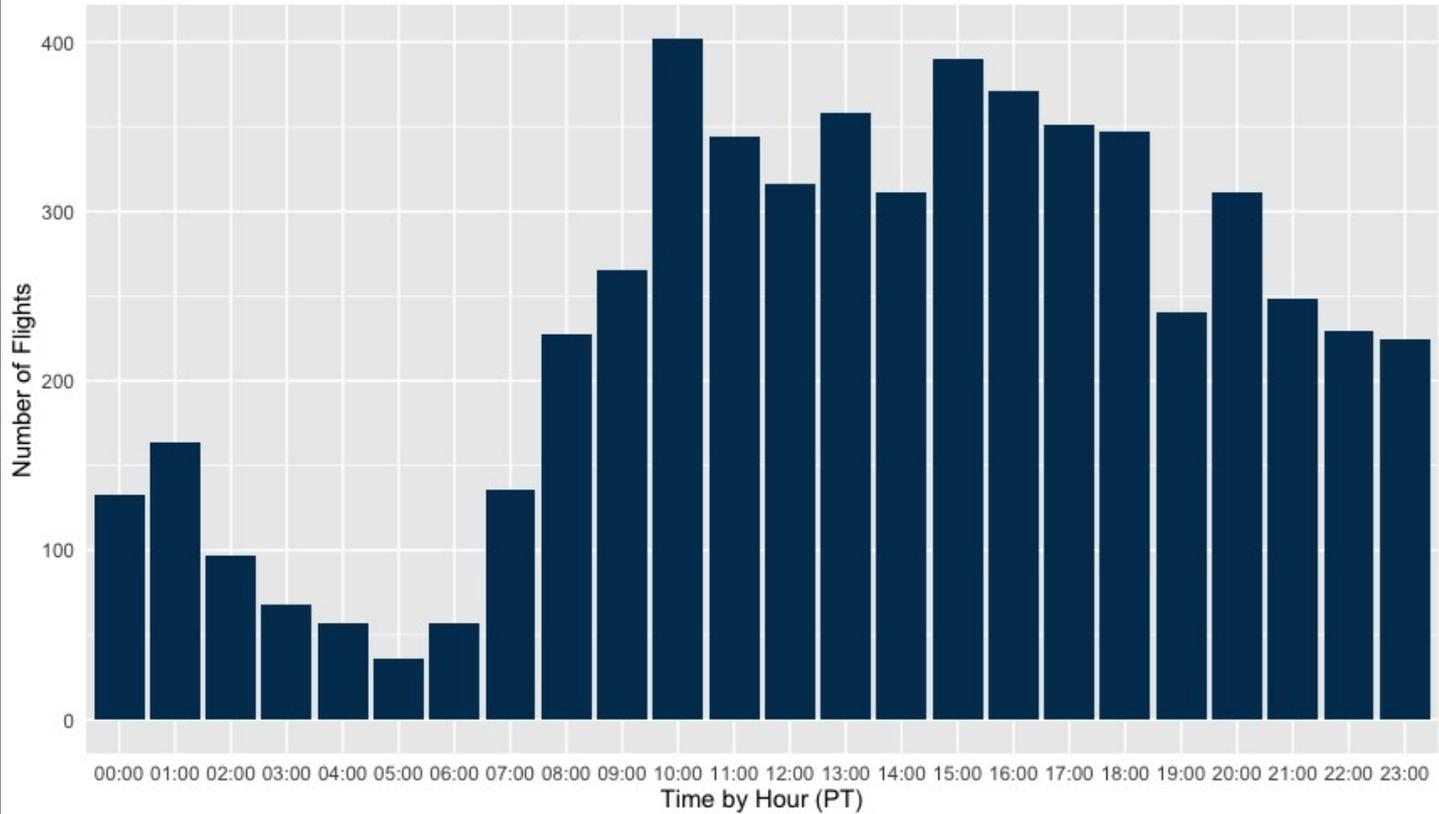
- Created more efficient way to read in and clean dataset

- Created method to detect outliers in altitude based on rolling range

- Provided examples of exploratory data analysis in terms of visualizations and summary statistics

- Explained code whenever confusion arose

Number of Flights in a Given Hour



Total Number of Flights: 2812
Data for flights from October 2019

DataSquad International

- Network of data support organization managers who employ students in the delivery of their services
- Support the distribution of the highly flexible DataSquad model
- Based on the fundamental principles of Open Science, employing and nurturing novices in their skills in data science

Thank you for listening!