



MAA MathFest

A Capstone Course in Statistics using Local Data

Dr. Heather Cook
Assistant Professor of Statistics
University of Southern Indiana

USI's Statistics Capstone

- Semester long student project
- Students work together
 - Groups of 2-3
- Student led
 - Students choose topics
 - Each student learns something new (statistical)
- Goal/Deliverables:
 - Open Colloquium Presentation
 - Written Article

Guidelines for Assessment and Instruction in Statistics Education (GAISE) 2016

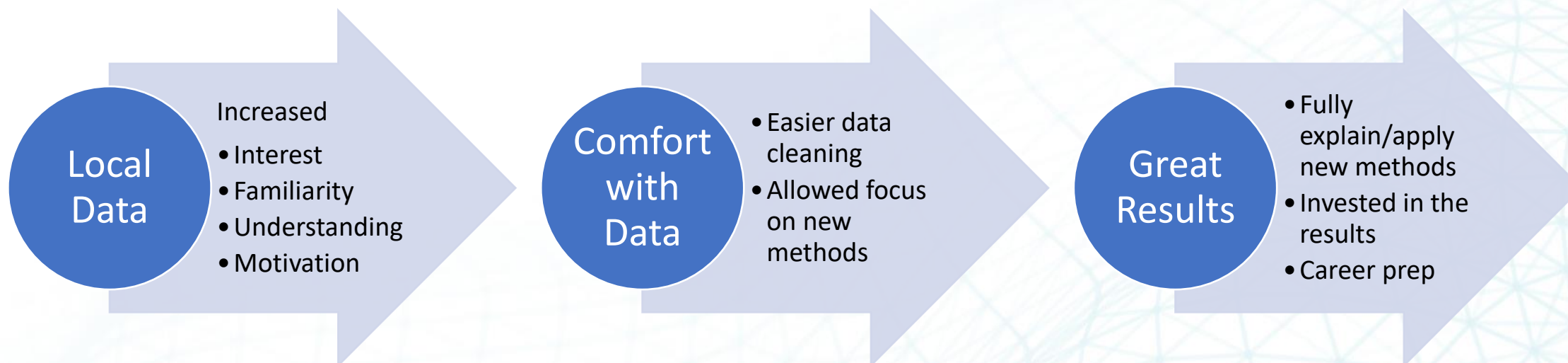
Follows GAISE guidelines

1. Teach statistical thinking.
 - Teach statistics as an investigative process of problem-solving and decision making.
 - Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

Spring 2021 Capstone

- Predicted housing prices from Fall 2020
- Two locations:
 - Evansville, IN
 - Nashville, TN
- Regression Methods Used:
 - Stepwise, LASSO, Ridge
 - Bagging, Random Forests, Boosting
- Manually collected data from Realtor.com
 - Sold price
 - Number of bedrooms, **bathrooms**, and stories
 - **Square footage of house** and lot
 - Year built
 - Images in listing
 - **Property tax in 2019**
 - Crime rate
 - Renovations
 - Garage type
- Good model performance

Spring 2021 Student Experiences



Local Data

- Increased
 - Interest
 - Familiarity
 - Understanding
 - Motivation

- Having new real world unexplored data excited them
- Applicable to their lives
- Showed they could use statistics for their own lives
- Strong connection to data and question they were trying to answer

Comfort with Data

- Easier data cleaning
- Allowed focus on new methods

- Collecting own data
 - Gave new appreciation
 - Knew the data intimately
- Interjected themselves into the project “This could be us”
- Easier to understand
 - They knew the areas and other aspects of the data

Great Results

- Fully explain/apply new methods
- Invested in the results
- Career prep

- Students wanted to answer their question and had a common goal
- Helped tie everything together since it’s “all under one umbrella”
- Felt better prepared for their careers since one team project
 - Rather than multiple individual questions or smaller projects

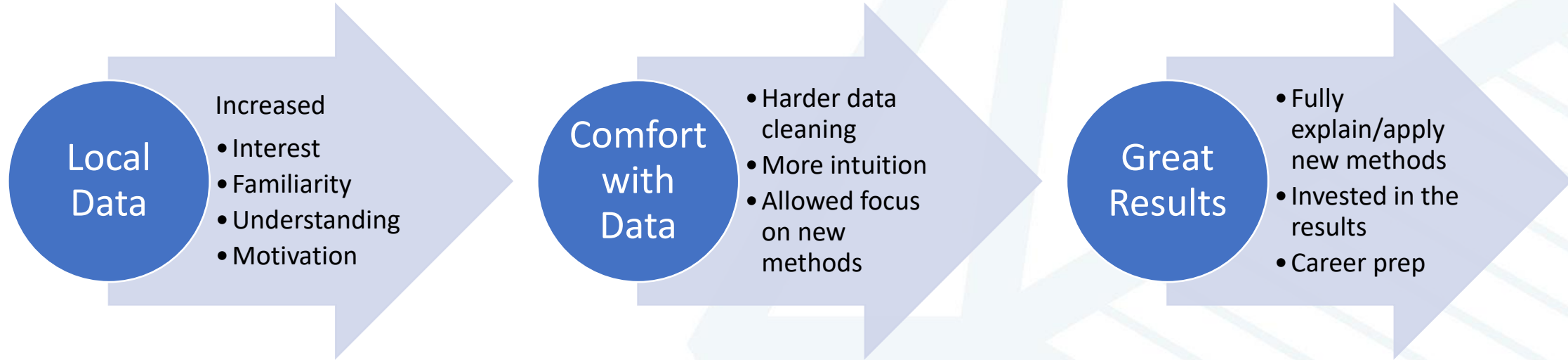
Spring 2021 Student Experiences

- Negatives:
 - Possibility for human error when collecting own data
 - Can't generalize results to larger population
 - One student didn't have ties to Evansville nor Nashville
 - Besides where USI is located
 - Less influence by personal connections
 - Still motivated by the data and question

Spring 2024 Capstone

- Predicted number of daily fire department calls and classified type of call
- All 2023 data from Evansville, IN
- Methods Used
 - Regression: Stepwise, LASSO
 - Classification: Multinomial Logistic, Random Forests
- Poor model performance
- Evansville Fire Department data (from EFD)
 - Date, time, type of call, location, main responding station
- Evansville Weather data (from Weather Underground)
 - Date, time, wind speed, pressure, precipitation, weather condition, humidity
- Duplicates, missing data, two sets merged based on time, new variables created

Spring 2024 Student Experiences

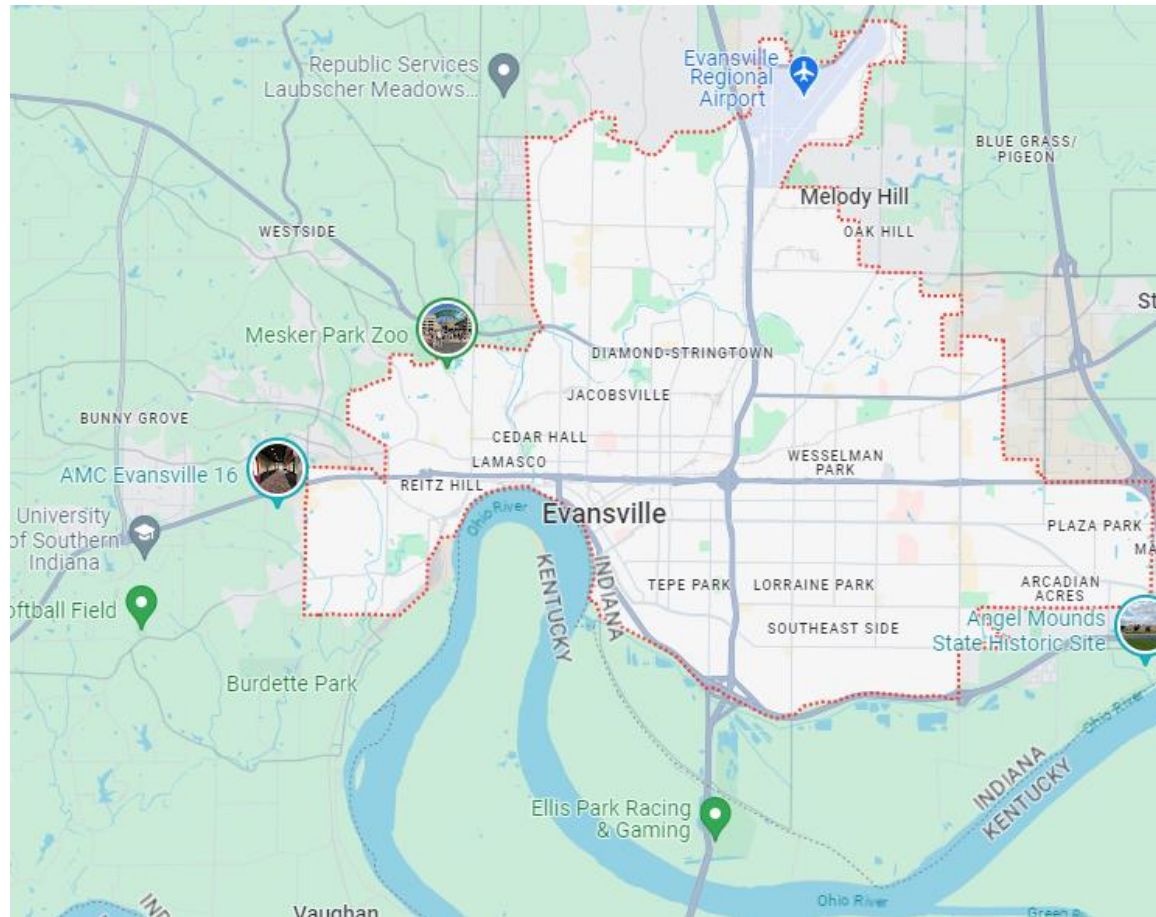


- Increased
 - Interest
 - Familiarity
 - Understanding
 - Motivation
- If had questions, local resources
- Talked to family and friends about the project
- Students saw importance of their work and wanted others to see that too
- More interesting and engaging since can relate to it

- Comfort with Data
 - Harder data cleaning
 - More intuition
 - Allowed focus on new methods
- Had time to fully understand methods and go deeper
 - With multiple question assignments, can't dwell on one problem too long
- Weren't held up by understanding data after data cleaning
- Thought about the methods in terms of the data they were using
 - Chose methods that fit the questions and thus the data

- Great Results
 - Fully explain/apply new methods
 - Invested in the results
 - Career prep
- New insights
 - Even though completed proper techniques, still get poor performance
- Big picture
 - Whole process helped realize how projects work
 - Build on previous work
- Presented results to EFD

Spring 2024 Student Experiences



- **Negatives:**
 - Extensive data cleaning took valuable time
 - Limited data from other sources
 - No traffic data to add
 - Weather data was all from airport (North side of town)
 - Can't generalize results to larger population
 - One student was from KY and was relieved we didn't go with IREAD-3 scores
 - IREAD-3 is specific to IN
 - Lacking personal connection

Summary

- Local data helps create **personal connections with data**
- Knowing the data well allows **time to focus on learning methods**
- Seeing the whole process allows them to tie it all together and **see the big picture**

Key Quotes

- “Having a personal connection just goes a long way, ... just in general. I feel like stuff that you read or do, or whatever, if you have a personal connection to it, you’re just going to remember it.”
- “From the beginning, I kind of knew that I wanted to have data from the community so that we would have results, or just like the paper would be about something that other people would want to read from here.... I just wanted it to be a little bit more important ... than other projects that [I’ve] work[ed] on.”
- “[The] class that prepared me the most was probably the capstone class because it prepared me to actually ... do a project instead of just like doing just questions.”
- “It’s not just something you learn, you get your degree, and then you forget about, [it’s] stuff that you can use for your whole life.”

Thank You!

Any questions?