Partnering for Progress AND Pandemic Projects (and beyond)

IASE Satellite Meeting – August 2021

A. John Bailer
baileraj@MiamiOH.edu
@john_bailer
@statsandstories

MIAMI UNIVERSITY

Abstract

Statistics majors and degrees often originated from mathematics departments and later in independent statistics departments. The emergence of data science and data analytics as publicly recognized activities and employment opportunities challenge us to consider our collaboration with a diverse collection of potential partners. The first part of this talk will focus on how partnerships can allow for novel degrees that expand our impact and reflect the changing skills needed in the workforce. The second part of this talk will address how experiential learning and classroom opportunities in statistics and data science can be enriched with problems from public health.

Outline

Part 1

1. History From math origins to statistics identity …
2. To connection with computer science … (data science)
3. To partners with for analytics

Part 2

4. Public Health, Pandemics and Experiential Learning

Thanks! Professor Engel for the invitation to join you AND to my students

Part 1: Partnering for analytics degree

Part 1

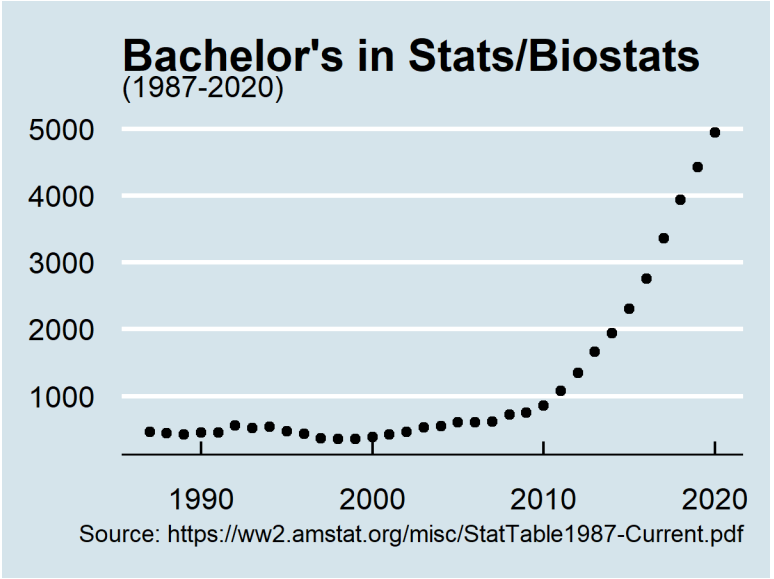
1. History From math origins to statistics identity …
2. To connection with computer science … (data science)
3. To partners with for analytics

assumptions:

* local experience used as a surrogate for general trends.
* Demand for data science and analytics outpaces supply

Part 1.1: History From math origins to statistics identity …

- Undergraduate Statistics degrees has relatively low enrollments in the late 1980s
- Master's degree considered by many as the degree needed to work as a 'statistician'



MIAMI UNIVERSITY

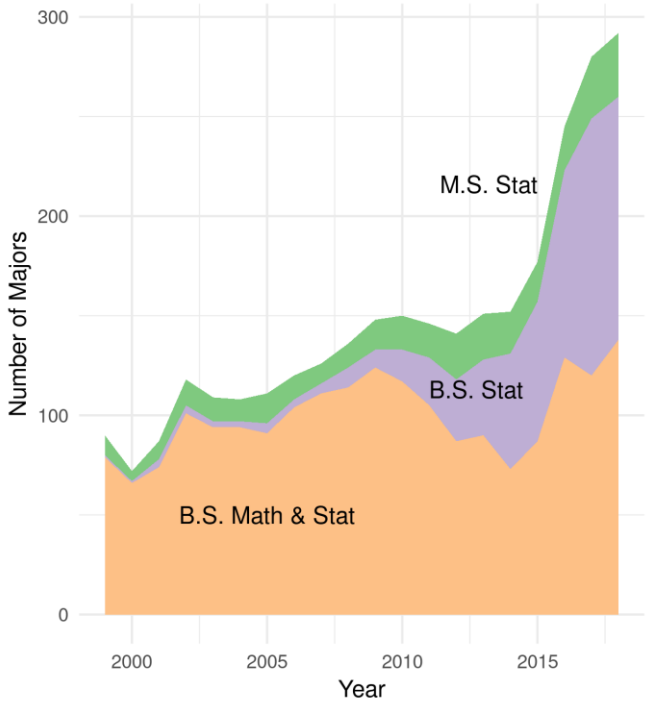
Part 1.1: BS Math & Stat (circa 2011)

- 31+ semester hours of MTH and STA 300+ courses
- Mathematics courses. All of these: Diff. Eq.; Abstract Algebra; Real (or complex) Analysis + At least one of these: Optimization; Combinatorics; Game Theory; Graph Theory; Math Finance; Numerical Analysis
- Statistics courses. Applied Statistics; Probability; Regression Analysis; At least one of these: Inferential Statistics; Experimental Design Methods.
- Electives to get to 31 hours
- Related courses: “a computer programming course”

Part 1.1: BS Statistics (circa 2011) – started in late 1980s

- 29+ semester hours of STA 300+ courses (Calc1-3,LA)
- Statistics courses. Applied Statistics; Probability; Statistical Programming*, Regression Analysis; Inferential Statistics; Experimental Design Methods + 3 courses from { Nonparametrics; SQC; Sampling; Multivariate; Data Practicum; Time Series; Categorical Data }
- Electives to get to 31 hours
- Related courses: “a computer programming course”
- * new course added in mid-2000s

Part 1.1: Majors (late 1990s to 2018)



Majors increasing in Statistics

BS Math & Stat relatively constant but BS Stat had dramatic growth

MIAMI UNIVERSITY

Part 1.1: Majors (late 1990s to 2018)

- U.S. News and World Report in their 2021 rankings reported Statistician #6 overall, #5 in the Best STEM Jobs and #2 in Best Business Jobs. Data Scientist was ranked #8 overall, #6 in Best STEM Jobs and #2 in Best Technology Jobs.
- Forbes ranked Data Scientist #1 and Data Analyst #31 in their list of Best Jobs in America for 2019.
- How do our stat degrees connect with data science and data analytics?

Part 1.2: BS Data Science & Statistics (Summer 2018 rev.)

- Core: Calc 3, Linear Algebra, Pgm Fundamentals (CSE), intro to stat modeling^, prob. Statistical Pgm, Reg Analysis; Inf Statistics
- Data Science Track: OOP, Data abstraction / data structures, database systems, mng big data, adv. data viz#, stat learning# + Bayesian# or time series + 2 of optimiz, graph th
- Statistics Track: Expt'l Design, Data Practicum + 2 additional stat classes + 1 simulation/optim class + related hours
- #new courses added in mid-2000s / ^revised in late 2010s
- Key department partner: CSSE – one track has almost CS minor

Part 1.2: Issues

- Current major still has significant math prerequisite requirements and computing science components
- You don’t need to be an engineer to drive a car. Can we help enhance content areas with analytics preparation?
- Intro stat has been taught been many departments and in many divisions.
- Can new partnerships be identified?
- Can a new major be defined with these partners?
- ANSWER: Yes (or my talk would be much shorter!)

Part 1.2: Issues

- You Don’t Have to Be a Data Scientist to Fill This Must-Have Analytics Role – Henke, Levine, McInerney (HBR, Feb 2018) https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role
- [analytics] *translators help ensure that the deep insights generated through sophisticated analytics translate into impact at scale in an organization. By 2026, the McKinsey Global Institute estimates that demand for translators in the United States alone may reach two to four million.*

Part 1.2: Issues

- *In addition to their domain knowledge, translators must possess strong acumen in quantitative analytics and structured problem solving.*
- *need to know what types of models are available (e.g., deep learning vs. logistic regression) and to what business problems they can be applied… be able to interpret model results and identify potential model errors, such as overfitting.*

Part 1.3: BA Data Analytics (Fall 20) – CORE + Concentration

CORE

- Professional Communication (course from ENG)
- Math Foundations for Data Analytics (course from MTH)
- Intro to Programming and Scripting for DA (STA course)
- Building, Managing and Exploring Data Sets in Analytics (STA)
- Intro to Stat Modeling (STA or ISA/POL classes)
- Data Ethics (PHL, CSE, JRN, ENG pick list)

Part 1.3: Math Foundations for Data Analytics

- Math concepts and terminology needed for statistical programming and data analysis. Topics include: systems of linear equations and matrix algebra; graphs and networks; logic and Boolean algebra; sets and probability; power, polynomial, exponential, logarithmic and trigonometric functions; basics of differential and integral calculus, including partial derivatives; elementary principles of continuous optimization; numerical methods. Emphasis on contexts related to data and programming.

Part 1.3: BA Data Analytics (Fall 2020) + CONCENTRATIONS

Concentrations (so far)

1. Geospatial Analytics (Geography)
2. Bioinformatics (BIO, MBI)
3. Sports Analytics (SLM – Sports Leadership and Marketing)
4. Social Data (POL, GTY)

Future? Data Journalism? Digital Humanities?

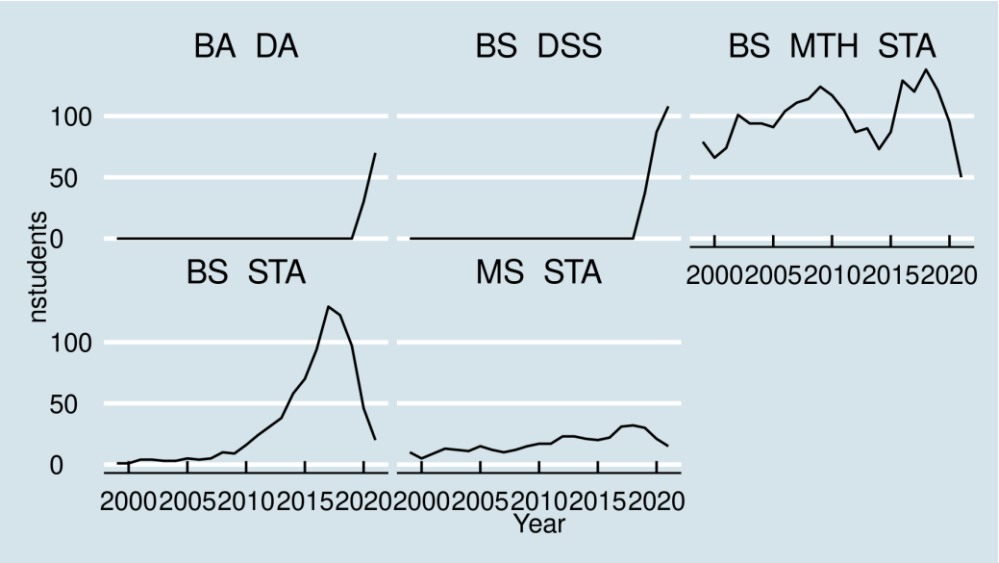
Part 1.3: BA Data Analytics Notes

a. Concentrations should have content foundation + advanced methods courses + adv. computational courses
b. Adviser for entering students in STA, concentrations will advise students more in later years of study
c. Steering committee with department reps
d. Other concentrations can be added in the future (e.g. data journalism)
e. Business analytics is separate degree in School of Biz

Part 1.3: BA Data Analytics Current Status

BA Data Analytics –> from n=0 in 2019 to n=70 in Fall 2021 Future?

BS Data Science & Statistics growing at expense of BS Stat



MIAMI UNIVERSITY

Evolution of degrees / curricula / courses

- *Relationships are like sharks, they have to keep moving forward or they die. And I think what we have on our hands is a dead shark* (from the movie *Annie Hall*)
  [credit: Photo by Glenda from Pexels]
- Replace ‘Relationships’ by ‘curriculum’?

Part 2: Public Health – experiential learning

Public Health, Pandemics and Experiential Learning

What learning opportunities emerge from public health challenges?

Assertion:

- Clients can enhance the experience in data practicum classes and for other classes including data visualization classes.
- Engage hearts first and heads will follow

Part 2: Getting Started – inviting clients

Dear Colleagues,
Do you or your office have data that would benefit from better analysis and visual display? Do you have a complicated story involving numerical summaries in which visualization might lead to insight? Do you have data that you haven't fully investigated but you believe might contain the nugget of an interesting story? If you are interested in help addressing these issues, you are invited submit a project idea for consideration.

Part 2: Getting Started – inviting clients

This Fall semester, I am teaching a section of an advanced data visualization course (…) populated by undergraduates and graduate students representing a diverse set of backgrounds including business, design, finance, psychology and statistics. This course focuses on the construction of well designed data displays that tell accessible stories from data. A major component of this class is a project that will be conducted for an external client.

Part 2: Getting Started – inviting clients

{ logistics + data description follow … }

- A short title
- Goal of the analysis (e.g. dashboard displaying important data; website with interactive visualization; a story for possible print/web publication)
- If possible, provide at least one or two specific questions to be answered by the analysis;
- Data to be analyzed, if available (e.g. spreadsheets, CSV files)

Part 2: Getting Started – inviting clients

{ logistics + data description follow … }

- A short title
- Goal of the analysis (e.g. dashboard displaying important data; website with interactive visualization; a story for possible print/web publication)
- If possible, provide at least one or two specific questions to be answered by the analysis;
- Data to be analyzed, if available (e.g. spreadsheets, CSV files)

Part 2: Case Studies

Case Study 1: Ohio COVID-19 cases – client: me (+ health dept.)

Case Study 2: Overdose deaths – client: county coroner
{ if time permits }

Part 2: Coronavirus Cases

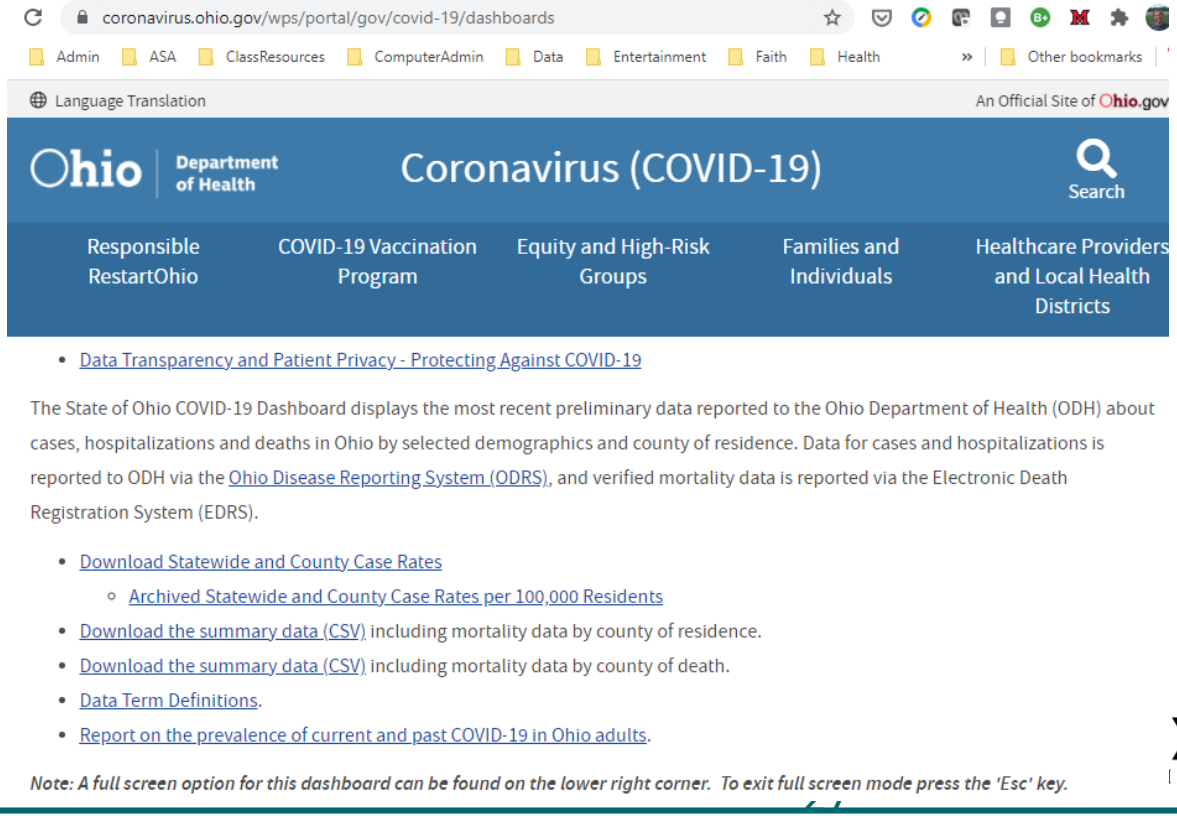
Context: Case Study 1: Working with Ohio Pandemic Data

- Teaching data viz during a pandemic when ALL my classes online
- Challenge to bring clients to class
- Seeing brilliant visualizations by the Financial Times, Our World in Data and other sites
- Hoping to connect with local experience – what's happening where I live?
- Ability to scaffold the experience

Part 2: Coronavirus Cases

State dashboard includes data that is updated!

368K rows



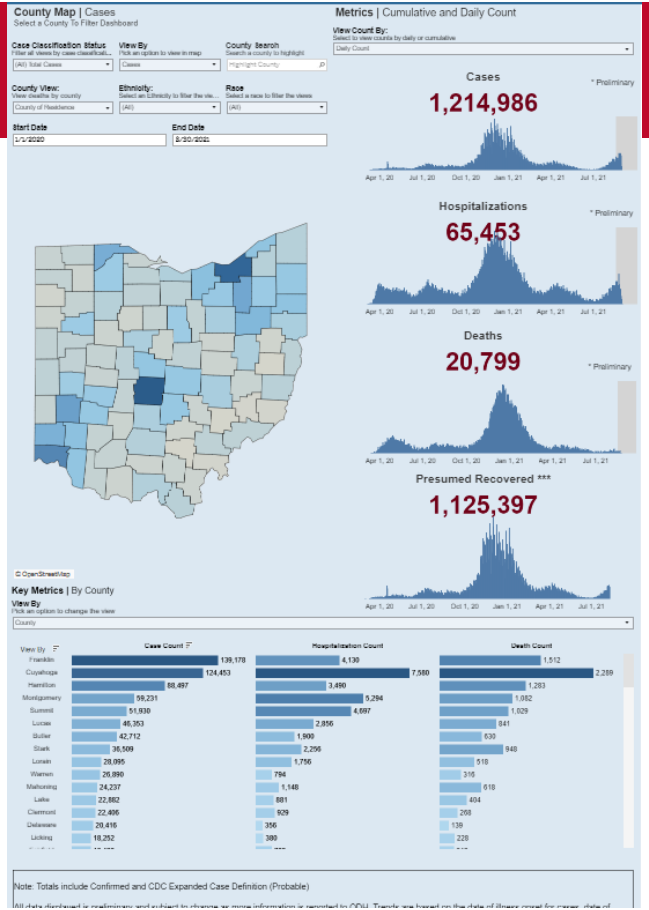
UNIVERSITY

Part 2: Coronavirus Cases

State dashboard includes variety of figures – choropleth map, vertical bar time series, horizontal bars

Features:

- Calculations need to build display data sets
- Color scaling for map (darker = more cases)
- Annotations (counts, shading grey – underreporting)



MIAMI UNIVERSITY

Part 2: Coronavirus Cases

- Teaching Strategy [tools used R, tidyverse, ggplot2]

Homeworks

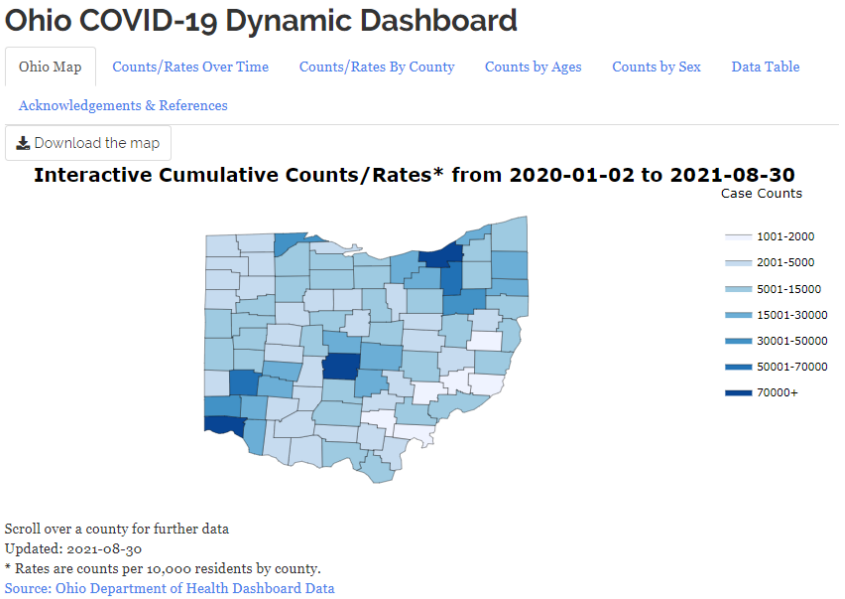
- Data Preparation [tidyverse – dplyr, tidyr, forcats]
- Time Series – [geom_col, geom_ma – also fct_reorder]
- Map – also scaling of colors [with cuts]
- Arranging graphs [patchwork, grid_arrange]

Projects

- Static Dashboard [generate static Ohio dashboard]
- Interactive Dashboard [Shiny – tab version with features]

Part 2: Coronavirus Cases -

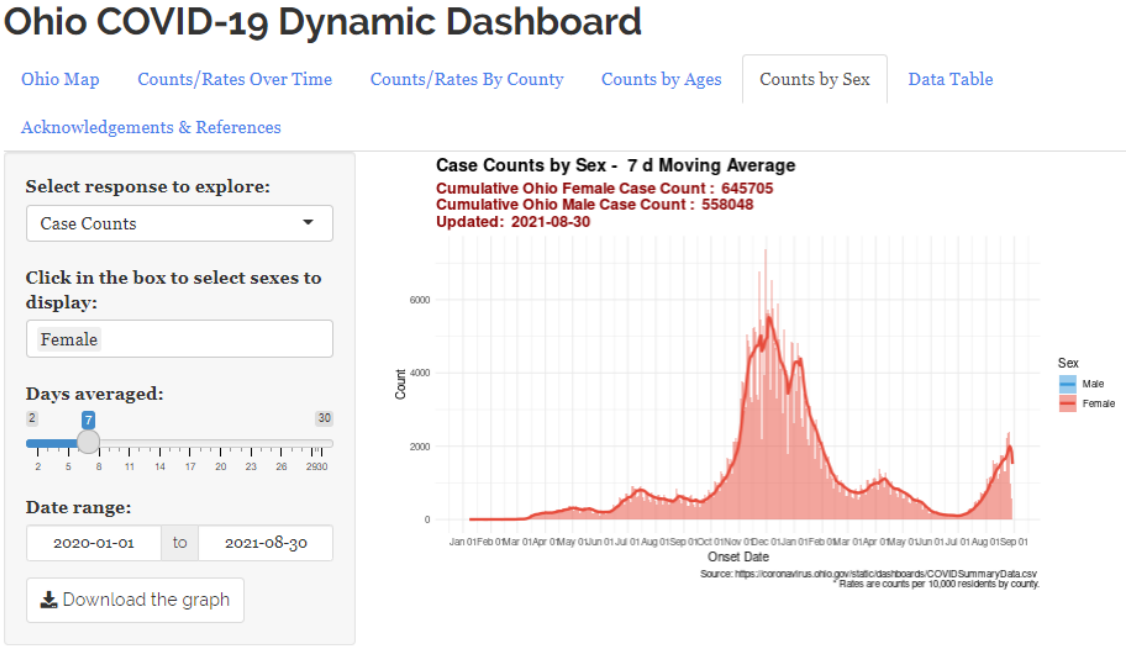
After class, team continued to work on this to produce a dashboard that contains elements not included in the Ohio dashboard



MIAMI UNIVERSITY

Part 2: Coronavirus Cases -

Great opportunity to consider what people might want to learn from these data – also how you can explore ideas such as moving averages



MIAMI UNIVERSITY

Part 2: Coronavirus Cases - Issues

- Location of data sets changed in middle of Fall 2020 semester
- Structure of data sets changed in Spring 2021
- Both provided ‘teachable moments’
- Current status – matching counts with Tableau (or not)
- Next projects – how do counties compare with respect to vaccination history?

Part 2: Case Study 2: Overdose deaths

Client 2: Working with a county coroner

- Butler County coroner wanted to understand patterns drugs found in people who died of drug overdoses
- Client for both a data visualization class and a data practicum class with students continuing to word on the project as independent studies

Part 2: Case Study 2: Overdose deaths - issues

- Geocoding of locations of deaths
- Ethics of what can be displayed (all data are for deceased)
- Data structure changes over the years
- Frequent collaboration with client needed to clarify which drugs could / should be grouped for producing displays

Conclusions

- Analytics and Data Science provide an opportunity for Statistics – expand current partnerships (CS, Math) and find new partnership opportunities (Biology, Geography, Sociology, Political Science, English) that might lead to new majors!
- Public health problems provides engaging and challenging experiential learning opportunities for our analytics, data science and statistics students

Contact information / Questions?

Contact information:

John Bailer

Email: baileraj@miamioh.edu
URL: http://www.users.miamioh.edu/baileraj
@john_bailer
@statsandstories

References

https://coronavirus.ohio.gov/wps/portal/gov/covid-19/dashboards # Ohio Dashboard (Tableau)

https://dataviz.miamioh.edu/COVID-OHIO/ # Ohio Dashboard (Shiny app – class – BETA)

Tuiyott A., Clements B., Bailer A.J., Mannix L.K. and Bailer J.F. (2020): Web Application to Investigate Butler County Overdose Death Data Ohio Journal of Public Health 3(1)
https://ohiopha.org/wp-content/uploads/2020/06/OJPH-2020-31-Tuiyott.pdf
App link: http://dataviz.miamioh.edu/Butler_County_Overdose_Deaths/

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

References

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

W Chang, J Cheng, JJ Allaire, C Sievert, B Schloerke, Y Xie, J Allen, J McPherson, A Dipert and B Borges (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny

TL Pedersen (2020). patchwork: The Composer of Plots. R package version 1.1.1. https://CRAN.R-project.org/package=patchwork

B Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

M Dancho and D Vaughan (2021). tidyquant: Tidy Quantitative Financial Analysis. R package version 1.0.3. https://CRAN.R-project.org/package=tidyquant