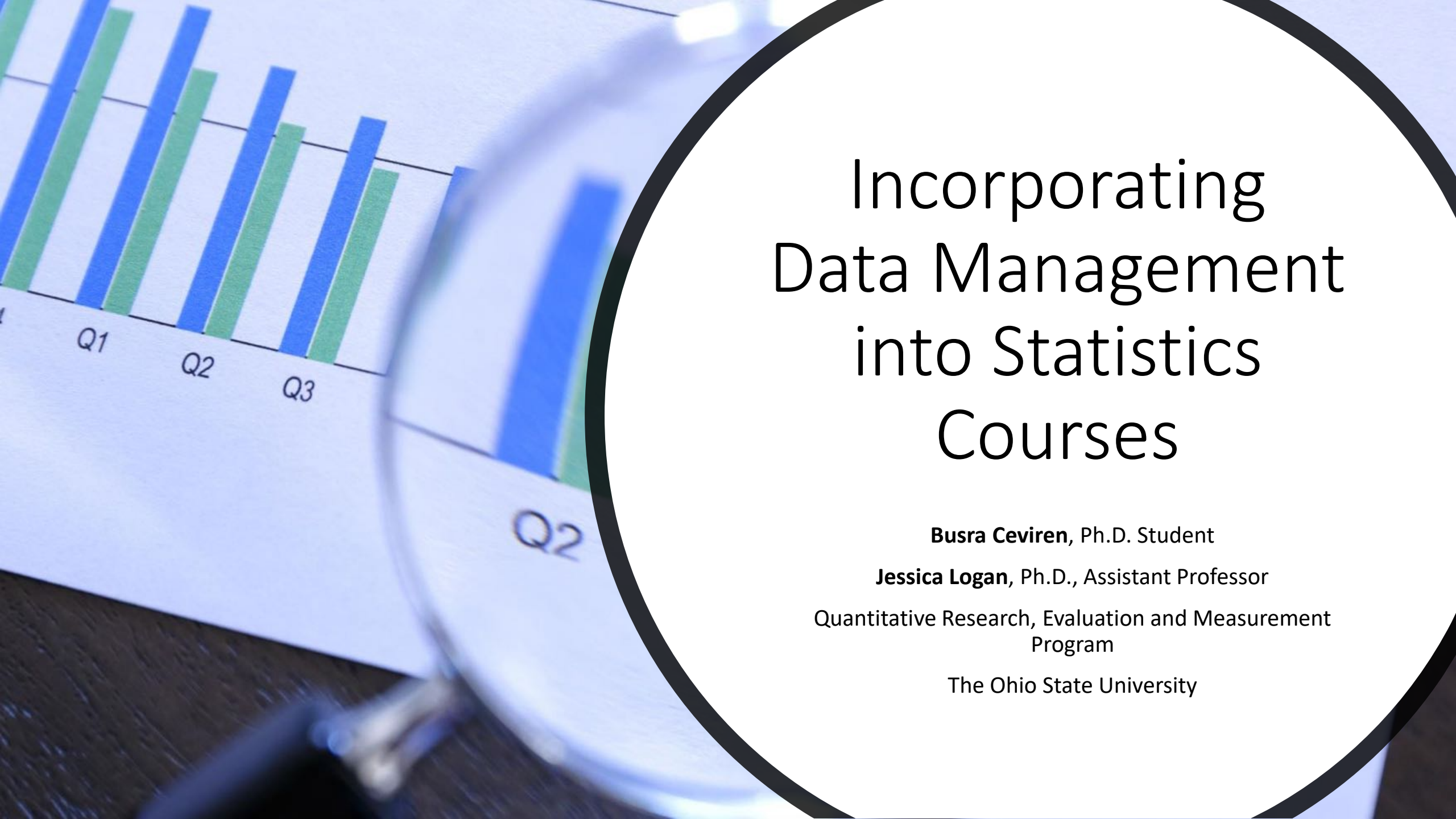


Incorporating Data Management into Statistics Courses

Busra Ceviren, Ph.D. Student

Jessica Logan, Ph.D., Assistant Professor

Quantitative Research, Evaluation and Measurement Program

The Ohio State University

Lifecycle of topic

Introductory statistics courses usually introduce concepts using small, tidy, example datasets.

Students learn data management skills through trial and error and struggle in various areas as they manage their research data.

Integrating data management skills into statistics courses would prepare students for real world of data analysis.

Research data management

Research data management (RDM) can be defined as a set of practices to handle information collected and created during research.[1]

It is a part of the research process and an ongoing activity throughout the data life cycle.

In context of statistics course, research data management refers to data structure, data entry, merging and wrangling or manipulation to allow for different statistical test to be computed.

[1] Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights*, *32*(1).

Why is RDM important for students?

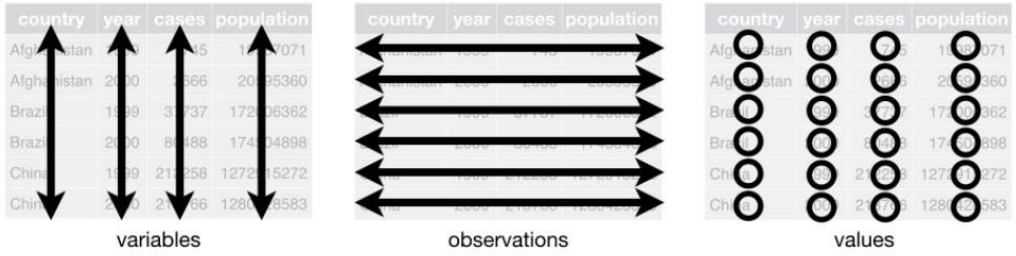
Reinforces key statistical concepts in the real world of data analysis

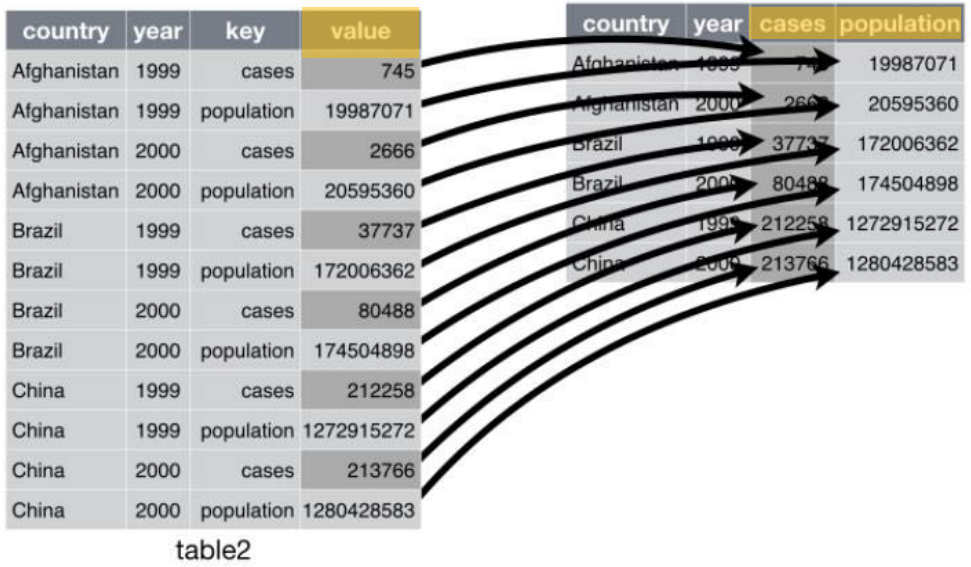
Prepares students to apply their knowledge to new datasets

Why is RDM important for students?

RDM reinforces key statistical concepts in the real world of data analysis.

- Data entry
- Data structure
- Central tendency





country	year	key	value
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362
Brazil	2000	cases	80488
Brazil	2000	population	174504898
China	1999	cases	212258
China	1999	population	1272915272
China	2000	cases	213766
China	2000	population	1280428583

table2

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Source of images: https://r4ds.had.co.nz/tidy-data.html

Why is RDM important for students?

RDM reinforces key statistical concepts in the real world of data analysis.

- Data entry
- Data structure
- Central tendency
- Correlation

Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
-2	-1	0	1	2

Why is RDM important for students?

RDM reinforces key statistical concepts in the real world of data analysis.

- Data entry
- Data structure
- Central tendency
- Correlation
- t-test

pre.sav [DataSet6] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Gra

1 : ID 1

	ID	pretest	var
1	1	86.00	
2	2	91.00	
3	3	83.00	
4	4	94.00	
5	5	72.00	
6	6	73.00	
7	7	95.00	
8	8	68.00	
9	9	90.00	
10	10	95.00	
11	11	94.00	
12	12	82.00	
13	13	70.00	
14	14	86.00	
15	15	93.00	
16	16	73.00	
17	17	77.00	
18	18	85.00	

post.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Grap

1 : ID 1

	ID	posttest	var
1	1	96	
2	2	85	
3	3	95	
4	4	96	
5	5	79	
6	6	82	
7	7	82	
8	8	98	
9	9	90	
10	10	96	
11	11	97	
12	12	80	
13	13	75	
14	14	88	
15	15	99	
16	16	80	
17	17	80	
18	18	95	

Why is RDM important for students?

RDM prepares students to apply their knowledge to new datasets.

- Data management errors can easily lead to violations of statistical conclusion validity.[1]
- Students have more experience manipulating data and preparing data for analysis as we integrate data management principles into statistics courses.
- Students can apply data management skills in future research.

[1] Brown, A. W., Kaiser, K. A., & Allison, D. B. (2018). Issues with data and analyses: Errors, underlying themes, and potential solutions. *Proceedings of the National Academy of Sciences, 115*(11), 2563-2570.

Contact Information

Busra Ceviren

ceviren.1@osu.edu

Jessica Logan

logan.251@osu.edu