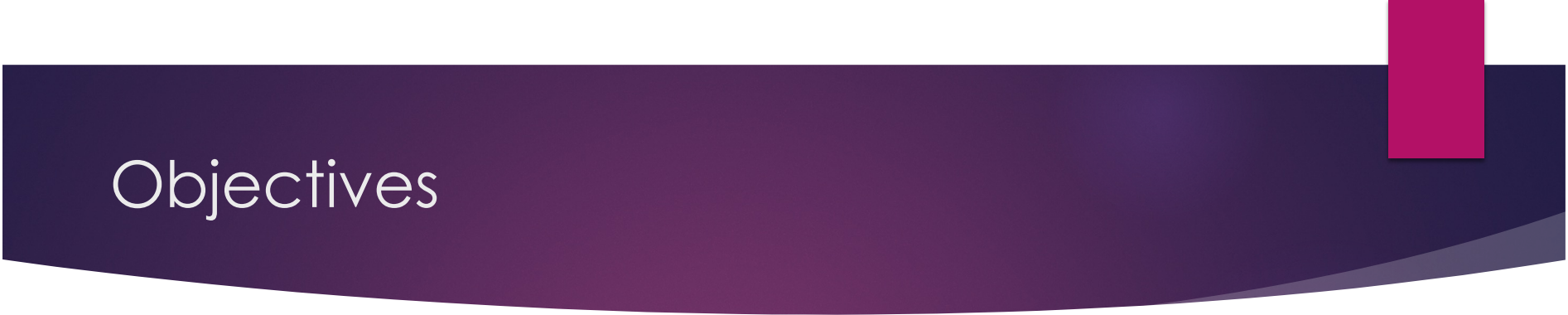




Turning Numbers into Action: A Hands-On Analytics Workshop for Public Health Practice

RANDI FORAKER, PHD

CAROL KOOB & MARY CHRISTIE



Objectives

- ▶ Introduce the unified dataset for Missouri mortality and socioeconomic outcomes
- ▶ Walk through a 10-step analysis to explore county-level data
- ▶ Generate summary statistics and maps for decision-making / reporting

Data Sources

US_CENSUS_ACS_5_YR_SUMMARY
(2017–2021)

County Health Rankings &
Roadmaps — Missouri (2025)

HDPulse NIMHD — Missouri
Mortality Table (2019–2023)

Data Preparation (Cleaning & Joining)

Standardized column names and types
(fixed percents/ratios, trimmed NAs)

Converted provider ratios to
'population per 1 provider' (numeric)

Harmonized county names/IDs to join
across sources

Saved a single, analysis-ready file:
MissouriCompiledData.csv

Data Prep continued

- ▶ Load tidyverse:

- ▶ Read in ACS Dataset → Split county variable

- ▶ ````{r} ACS_Data_modified <- ACS_Data %>% separate(COUNTY, into = c("COUNTY", "STATE"), sep = ",") ````

- ▶ Read in Mortality dataset → Select only county and death variable and rename

- ▶ ````{r}Missouri_Death_modified <- Missouri_Death %>% select(County, Age.Adjusted.Death.Rate.....deaths.per.100.000) %>% rename(COUNTY = County, ALL_MORTALITY_AGE_ADJUSTED = Age.Adjusted.Death.Rate.....deaths.per.100.000)````

- ▶ Repeat with Disease-specific datasets

- ▶ Read in County Health Rankings dataset → (Added "County" to all county names before import)

Data Prep continued

- ▶ Joined each dataset using “Left Join”

- ▶ ````{r}mo_death <- ACS_Data_modified %>% left_join(Missouri_Death_modified, by = "COUNTY")````

- ▶ ````{r}mo_cancer <- mo_death %>% left_join(Missouri_Cancer_modified, by = "COUNTY")````

- ▶ ````{r}mo_withdeaths <- mo_cancer %>%
left_join(Missouri_HeartDisease_modified, by = "COUNTY") %>%
left_join(Missouri_KidneyDisease_modified, by = "COUNTY") %>%
left_join(Missouri_Diabetes_modified, by = "COUNTY")````

- ▶ Repeat

Data Prep continued

- ▶ Removed some variables for clarity

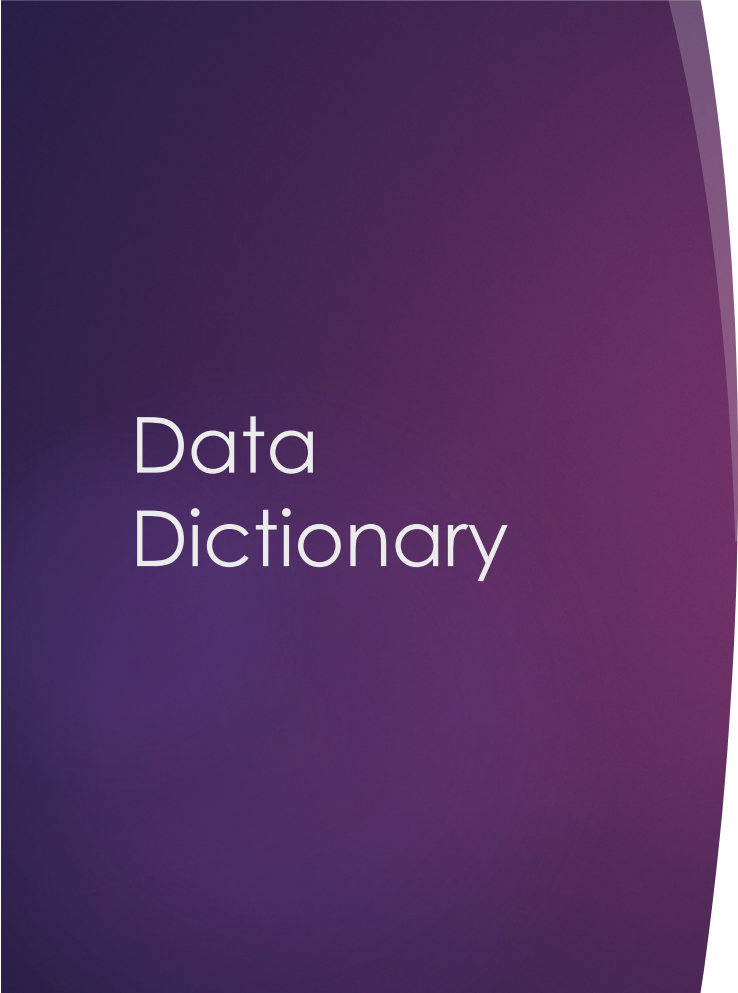
- ▶ ````{r}MO_otherdata <- `2025_MO_CNT_RAT` %>% select(-c(FIPS, State))`
 - ▶ `mo_withdeaths_edited <- mo_withdeaths %>% select(-c(GEOID, STATE))````

- ▶ Convert ratios to numeric

- ▶ ````{r}testdata <- MissouriCompiledData %>%
 separate(RATIO_PRIMARY_CARE_PHYSICIANS, into =
 c("RATIO_PRIMARY_CARE_PHYSICIANS", "REMOVE1", "REMOVE2"), sep = ":") %>%
 select(-c(REMOVE1, REMOVE2))testdata$RATIO_PRIMARY_CARE_PHYSICIANS <-
 as.numeric(testdata$RATIO_PRIMARY_CARE_PHYSICIANS)````

Data Prep continued

- ▶ Renamed variables so naming scheme matches existing dataset
 - ▶ ```{r}MO_otherdata_renames <- MO_otherdata %>%`
 - ▶ `rename(COUNTY = County,`
 - ▶ `PCT_LOW_BIRTH_WEIGHT = X..Low.Birth.Weight,`
 - ▶ `PCT_FAIR_OR_POOR_HEALTH = X..Fair.or.Poor.Health,`
 - ▶ `PCT_VACCINATED = X..Vaccinated,`
 - ▶ `RATIO_PRIMARY_CARE_PHYSICIANS = Primary.Care.Physicians.Ratio,`
 - ▶ `RATIO_MENTAL_HEALTH_PROVIDERS = Mental.Health.Provider.Ratio,`
 - ▶ `RATIO_DENTISTS = Dentist.Ratio,`
 - ▶ `PCT_WITH_ANNUAL_MAMMOGRAM = X..with.Annual.Mammogram,`
 - ▶ `PCT_SEVERE_HOUSING_PROBLEMS = X..Severe.Housing.Problems,`
 - ▶ `PCT_HOUSEHOLDS_WITH_BROADBAND_ACCESS = X..Households.with.Broadband.Access,`
 - ▶ `PCT_SOME_COLLEGE = X..Some.College,`
 - ▶ `PCT_UNEMPLOYED = X..Unemployed)```



Data Dictionary



Defines every variable, unit, and meaning

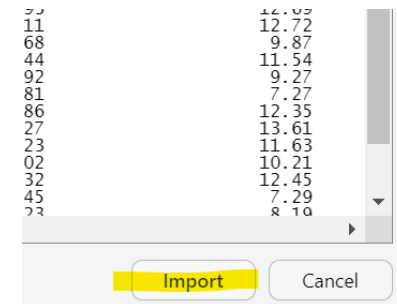
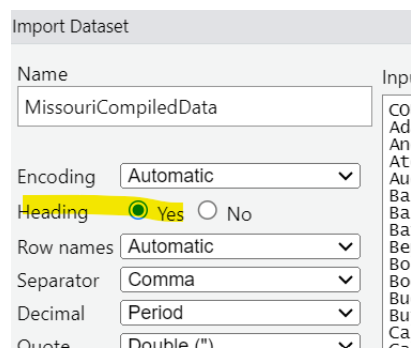
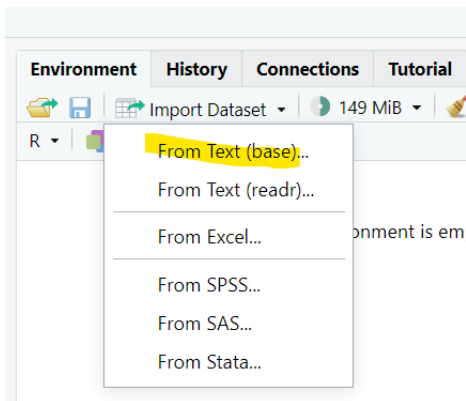
Clarifies mortality units: age-adjusted deaths per 100,000

Specifies provider metrics: population per 1 provider

Ensures consistent interpretation across the cohort

Step 1: Load the Dataset

- ▶ Read in the data set by clicking the Import Dataset button in the Environment tab to the right.
 - ▶ Select first option "From Text (base)".
 - ▶ Select the MissouriCompiledData.csv file to upload.
 - ▶ On next screen, click "Yes" by Heading and then "Import."



Step 2: Install & Load Packages

- ▶ Automatically install missing packages
- ▶ Load tidyverse, janitor, tigris, sf, gt, plotly, etc.
 - ▶ Either click CTRL+enter, line-by-line or highlight the set of code to run together

```
install_and_load <- function(packages) {  
  for (pkg in packages) {  
    if (!requireNamespace(pkg, quietly = TRUE)) {  
      install.packages(pkg, dependencies = TRUE)  
    }  
    library(pkg, character.only = TRUE)  
  }  
}  
install_and_load(c("tidyverse", "plotly", "sf", "tigris"))
```

Step 3: View MO County Names

- ▶ Review the names of all the counties in Missouri to determine how they are listed
- ▶ Code: `unique(MissouriCompiledData$COUNTY)`

[1] "Adair County"	"Andrew County"	"Atchison County"
[4] "Audrain County"	"Barry County"	"Barton County"
[7] "Bates County"	"Benton County"	"Bollinger County"
[10] "Boone County"	"Buchanan County"	"Butler County"
[13] "Caldwell County"	"Callaway County"	"Camden County"
[16] "Cape Girardeau County"	"Carroll County"	"Carter County"
[19] "Cass County"	"Cedar County"	"Chariton County"
[22] "Christian County"	"Clark County"	"Clay County"
[25] "Clinton County"	"Cole County"	"Cooper County"
[28] "Crawford County"	"Dade County"	"Dallas County"
[31] "Daviess County"	"DeKalb County"	"Dent County"
...	...	...

Step 4: Select County

- ▶ ## Step 4: Pick either your county or a county you are interested in learning about by pasting the county name in the code
- ▶ Code: `YOUR_COUNTY <- c("Phelps County")`
 - ▶ (Don't forget the quotes! County must match format in dataset.)

Step 5: Create a County Table

- ▶ Creates a table for your chosen county with all included variables
 - ▶ No editing to provided code needed

	Value
COUNTY	Phelps County
PCT_BELOW_POVERTY_LINE	18.81
PCT_NO_HEALTH_INSURANCE	8.59
MEDIAN_HOUSEHOLD_INCOME	48124
PCT_LESS_THAN_HIGH_SCHOOL	9.4
PCT_NO_VEHICLE_HOUSEHOLD	12.08
PCT_LIMITED_ENGLISH_HOUSEHOLD	1.53
PCT_HOME_OWNERSHIP	60.88
ALL_MORTALITY	1007
CANCER_MORTALITY	159.6
HEART_DISEASE_MORTALITY	188.9

Step 6: Review possible variables

- ▶ Now we are going to pick a variable of interest and look at how this variable ranks across counties.
- ▶ Review the variable names to know your options by running the names command.

```
[1] "COUNTY"  
[3] "PCT_NO_HEALTH_INSURANCE"  
[5] "PCT_LESS_THAN_HIGH_SCHOOL"  
[7] "PCT_LIMITED_ENGLISH_HOUSEHOLD"  
[9] "ALL_MORTALITY"  
[11] "HEART_DISEASE_MORTALITY"  
[13] "-----"
```

```
"PCT_BELOW_POVERTY_LINE"  
"MEDIAN_HOUSEHOLD_INCOME"  
"PCT_NO_VEHICLE_HOUSEHOLD"  
"PCT_HOME_OWNERSHIP"  
"CANCER_MORTALITY"  
"KIDNEY_DISEASE_MORTALITY"  
"-----"
```



Step 7: Enter the variable of interest

- ▶ Enter your variable of interest here.
- ▶ (Tip: You can copy and paste from your variable list above.)

Step 8:

- ▶ Create a table with your variable ranked by county from largest to smallest.
- ▶ (No editing to code needed)

COUNTY	Total
<i><chr></i>	<i><dbl></i>
1 Pemiscot County	27.5
2 Oregon County	25.4
3 Wayne County	23.6
4 Howell County	23.5
5 Shannon County	23.4
6 Mississippi County	23.4
7 Adair County	22.9
8 Barton County	22.1
9 Dunklin County	22.1
10 Carter County	22.0
11 Ozark County	21.9
12 Iron County	21.7
13 Douglas County	21.3
14 Ripley County	21.3
15 Butler County	21.3
16 Dallas County	21.1
17 Nodaway County	20.7
18 Morgan County	20.0
19 Wright County	19.7
20 St. Louis city	19.6



Step 9: Prep for Mapping

- ▶ Now we can visualize this on a map.
- ▶ Generate map data and join it into the data of your selected variable of interest above.
- ▶ (No editing to code needed)

Step 10: Mapping your County Variable

- ▶ Create your map!
- ▶ You can change the title and legend title of the graph by changing what is in the quotes on the first two lines.
- ▶ Code:

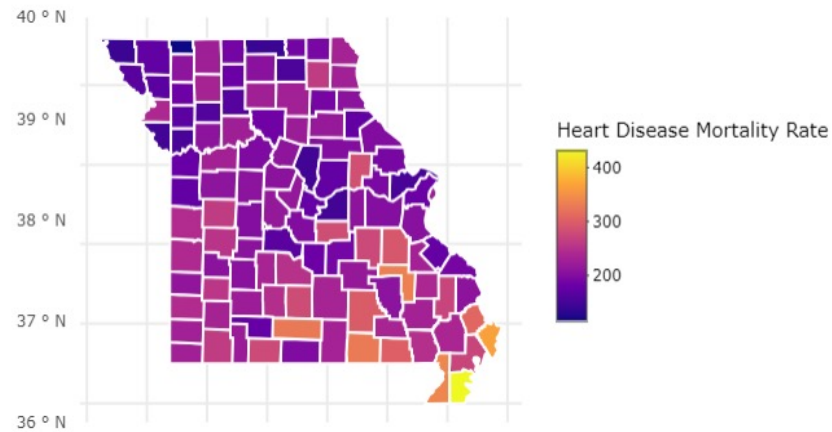
```
MAP_TITLE <- c("Percent Below the Poverty Line by County")
LEGEND_TITLE <- c("Percent Below the Poverty Line")
interactivemap <- ggplot(mo_map) + geom_sf(aes(fill = .data[[YOUR_VARIABLE]],
text = paste0("County: ", NAME,
"<br>Total: ", round(.data[[YOUR_VARIABLE]], 2))),
color = "white") + scale_fill_viridis_c(option = "plasma", na.value = "grey90") + theme_minimal() +
labs(fill = LEGEND_TITLE, title = MAP_TITLE)ggplotly(interactivemap, tooltip = "text")
```



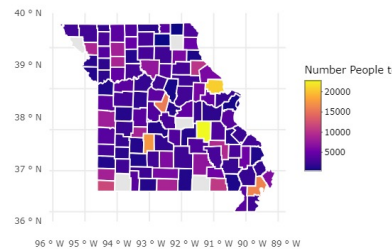

REPEAT!!!

- ▶ Interested in additional counties or variables?
- ▶ Go back to Step 4 to change YOUR_COUNTY and/or Step 7 YOUR_VARIABLE and rerun the code to look at different aspects of the data that interest you!
- ▶ (Reminder to update the title and legend for your graph to match your new variable.)

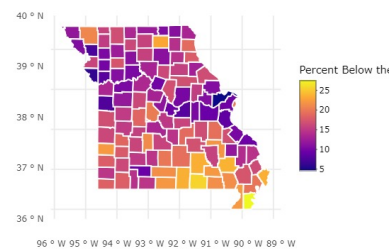
Heart Disease Mortality Rate



Number People to 1 Physician



Percent Below the Poverty Line by County



Multiple Maps



Limitations

- ▶ Datasets from different timeframes
- ▶ Some variables did not have county data (e.g., PPL_TO_PRIMARY_CARE_PHYSICIAN)



Thank you