

**CANCER SURVIVORSHIP OUTCOMES IN
INDIVIDUALS WITH LUNG CANCER: AN ANALYSIS
OF THE BEHAVIORAL RISK FACTOR
SURVEILLANCE SYSTEM (BRFSS)**

DS 5100 GROUP PROJECT

Uyen Nguyen (gmd8sq)

Andy Ortiz (eao7r)

Lee Ann Johnson (lj6gd)

JD Pinto (jp5ph)



OVERVIEW

- Introduction
- Data Source
- Data Cleaning
- Data Analysis & Processing
- Results
- Testing
- Conclusions



INTRODUCTION

- Lung cancer
 - 235,760 new cases
 - 131,880 new deaths
- Survivorship begins at diagnosis
- Relatively short survival after diagnosis
- Know less about survivorship compared to other cancers



RESEARCH QUESTIONS

The overall purpose of this project was to create predictive models to identify patient characteristics associated with guideline-concordant survivorship care.

- In a nationally representative sample of individuals with lung cancer, we will:
 1. Describe the demographics, physical, and mental health characteristics
 2. Examine associations between demographics, physical, and mental health characteristics and the survivorship outcomes (cancer care summaries, written cancer care summaries, and health insurance coverage)
 3. Model which demographic physical, and mental health characteristics predict survivorship outcomes



DATA SOURCE

- Behavioral Risk Factor Surveillance System (BRFSS)
- Annual, telephone-based
- 50 states + territories
- Topics covered
 - Preventative services
 - Health-related risk behaviors
 - Chronic health conditions
- Published by the Centers for Disease Control (CDC)
 - <https://www.cdc.gov/brfss/index.html>



DATA CLEANING: FIRST STEPS

- Merged 5 years of data: 2016, 2017, 2018, 2019, 2020
- Selected individuals with lung cancer
- Reduced dataset from 280 columns to 18
 - 14 predictors
 - 4 response variables



PREDICTOR VARIABLES

State of Residence

Sex

Age

Education level

Marital status

Employment status

Income level

Race

Mental health in previous 30 days

Physical health in previous 30 days

Activities of daily living in previous 30 days

Difficulty with cognition

Difficulty with mobility

Difficulty performing errands

RESPONSE VARIABLES

| | |
|--|--|
| Written Summary (CRVSUM) | Did any doctor, nurse, or other health professional ever give you a written summary of all the cancer treatments that you received? |
| Routine Cancer Check-Up (CSRVRTRN) | Have you ever received instructions from a doctor, nurse, or other health professional about where you should return or who you should see for routine cancer check-ups after completing treatment for cancer? |
| Adequate Health Insurance Coverage (CSRVINSR) | With your most recent diagnosis of cancer, did you have health insurance that paid for all or part of your cancer treatment? (“Health insurance” also includes Medicare, Medicaid, or other types of state health programs.) |
| Cancer Pain (CSRVPAIN) | Do you currently have physical pain caused by your cancer or cancer treatment? |



DATA CLEANING: FINAL STEPS

- Identified missing data
- After removal of missing data:
 - For 3 of the response variables, 498 rows were useful after removal of missing data
 - For 1 response variable, 697 rows were useful



DATA ANALYSIS & PROCESSING

- 1. Visualizations of predictor variables
- 1. Visualizations of response variables
- 1. Logistic regression assumptions and tests
- 1. Logistic regression

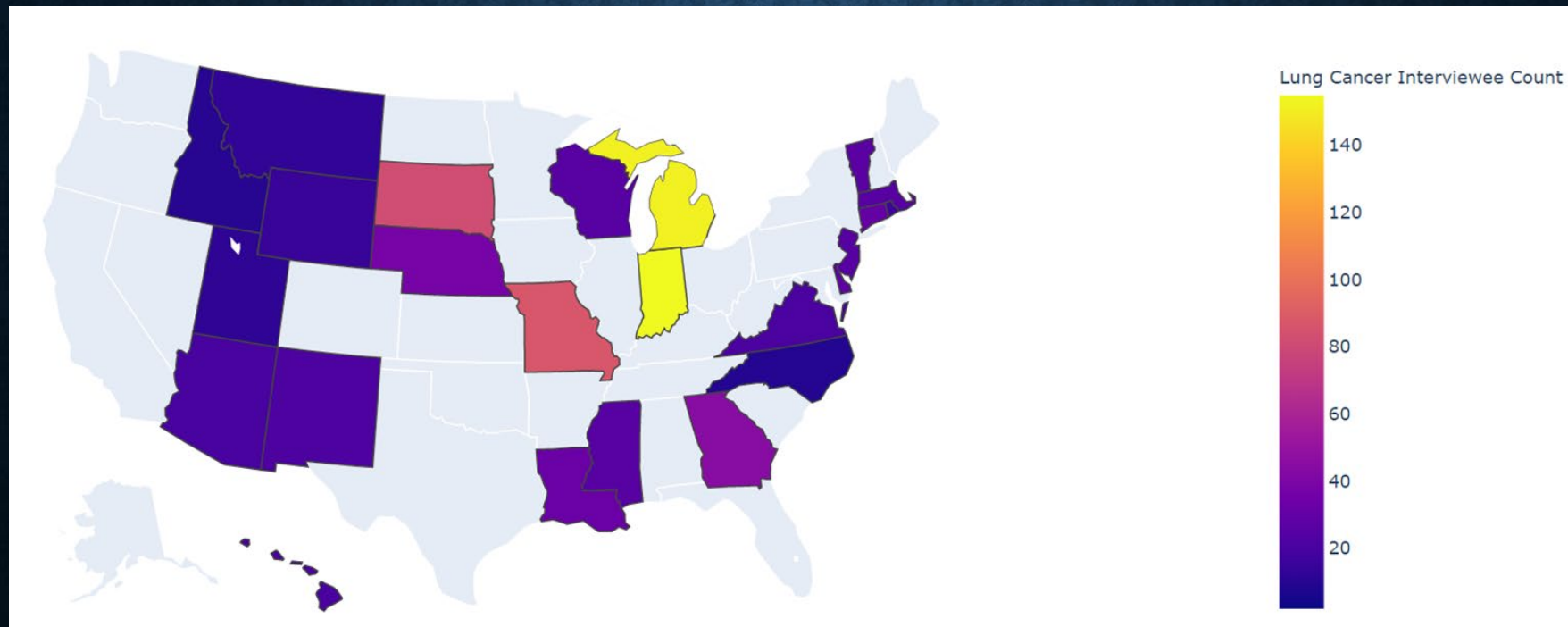


DATA ANALYSIS & PROCESSING

- Predictor and response variables were explored
- A selection of visuations will be presented for each type of variable

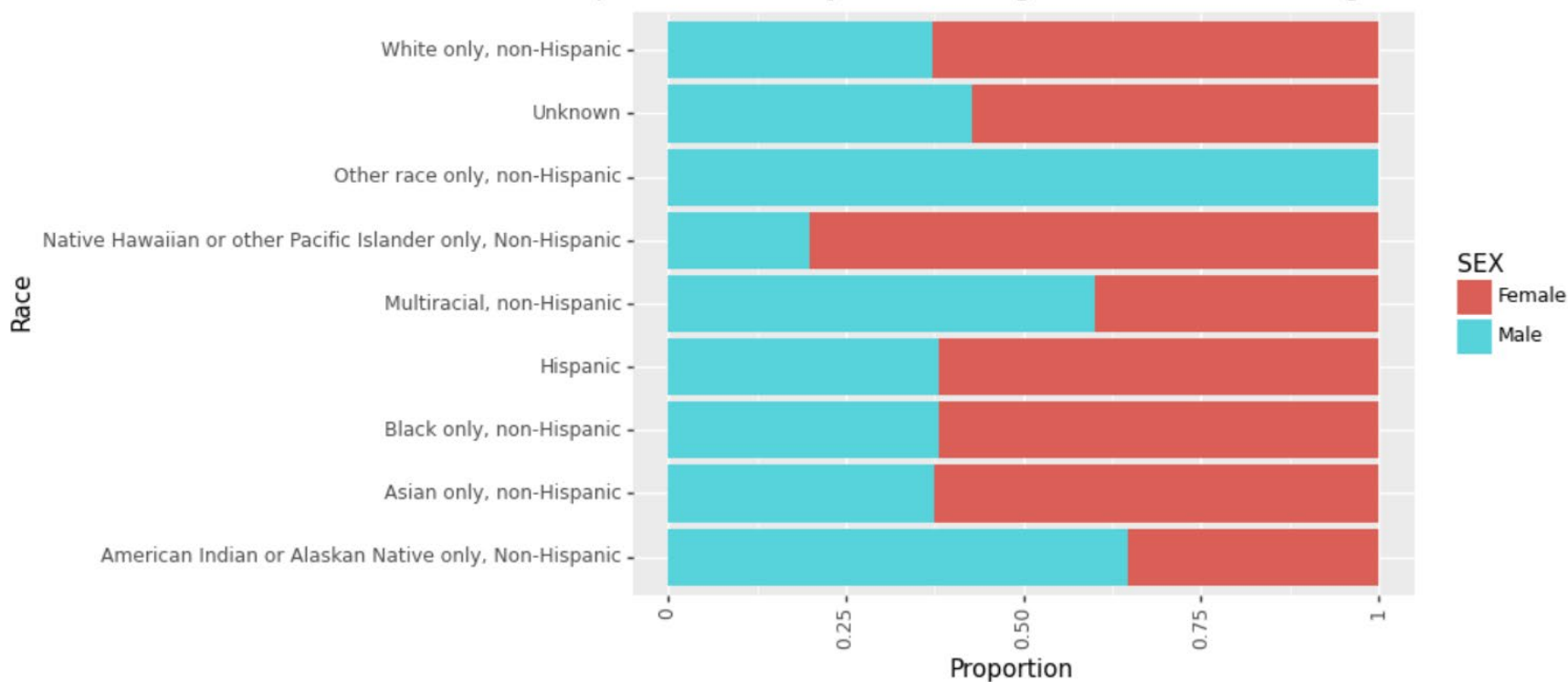


PREDICTOR: PARTICIPANTS BY STATE

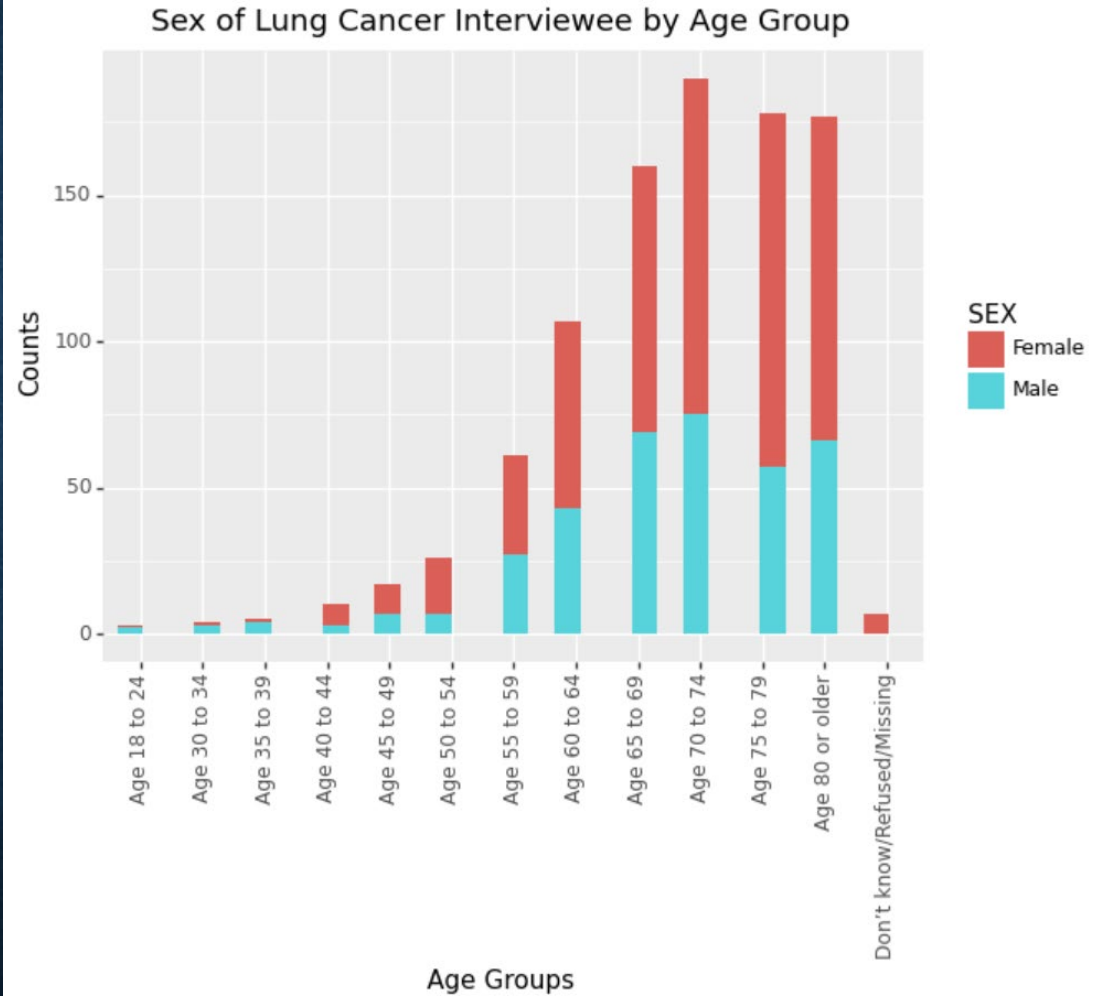


PREDICTOR: PARTICIPANTS BY SEX AND RACE

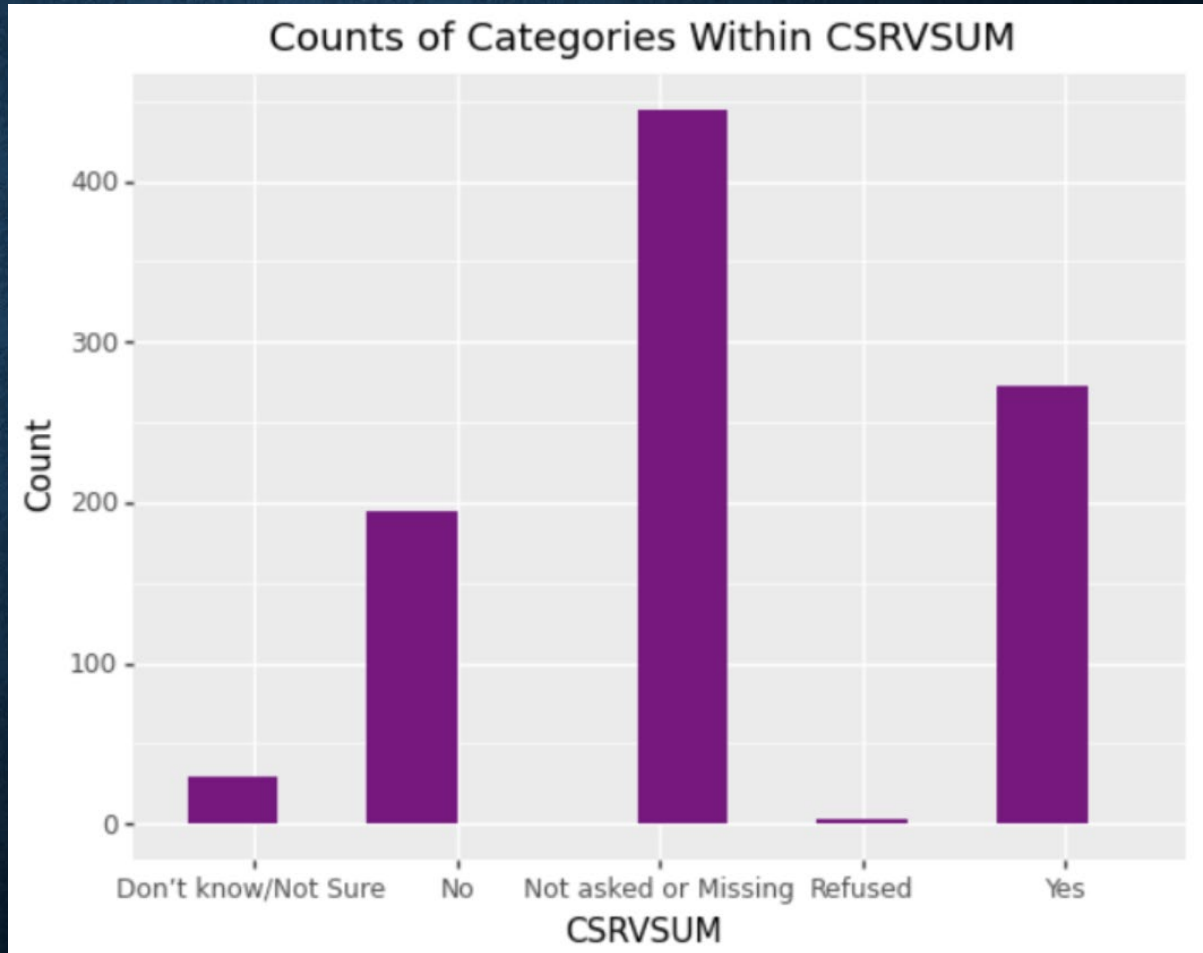
Proportion of Sex by Race among Interviewee with Lung Cancer



PREDICTOR: PARTICIPANTS BY AGE AND SEX



**RESPONSE:
CATEGORICAL
RESPONSES
FOR RECEIVING
WRITTEN
SUMMARY**





RESULTS FROM LOGISTIC REGRESSIONS

- Written Summary
 - Individuals with a college education are less likely to receive a written summary of cancer care.
 - Other variables were not significant within the model.

TESTING

- Converting SAS files to pandas dataframe
- Deleting unnecessary files

```
# function that checks for files to delete
def cleanup():
    import os
    try:
        os.remove("LLCP2016.XPT")
        os.remove("LLCP2017.XPT")
        os.remove("LLCP2018.XPT")
        os.remove("LLCP2019.XPT")
        os.remove("LLCP2020.XPT")
        return 'Files were removed.'
    except:
        return 'These files have already been removed.'
```

```
# function that converts XPT/SAS formatted files to pandas dataframes
def converter():
    try:
        brfss2016 = pd.read_sas("LLCP2016.XPT", format='xport')
        brfss2017 = pd.read_sas("LLCP2017.XPT", format='xport')
        brfss2018 = pd.read_sas("LLCP2018.XPT", format='xport')
        brfss2019 = pd.read_sas("LLCP2019.XPT", format='xport')
        brfss2020 = pd.read_sas("LLCP2020.XPT", format='xport')
        return 'All files were converted.'
    except:
        return 'File conversion was interrupted.'
```

```
import unittest
# from Group3ProjectCode.ipynb import cleanup, converter

class ProjectFunctionTestCase(unittest.TestCase): # inherit from unittest.TestCase

    def test_cleanup(self):
        self.assertEqual(cleanup(), 'These files have already been removed.')

    def test_converter(self):
        self.assertEqual(converter(), 'File conversion was interrupted.')

if __name__ == "__main__":
    # unittest.main()
    unittest.main(argv=[''], exit = False)
```



CONCLUSIONS

- Health-care providers should be more cognizant about providing written summaries to all individuals with lung cancer
- We can improve by adding more interactivity with respect to survey by year
- We can improve by adding more functionality that allows researchers to choose response variables and get results for each of them interactively

**THANK YOU FOR COMING TO OUR
TED TALK**