TEAM 1
MENTAL HEALTH ANALYSIS

JAN 13th 2018

GROUP MEMBERS:

Aiyana Liz Mathew, Aruna Amaresan, Aswathy Mohan, Michael Ilelaboye

AGENDA

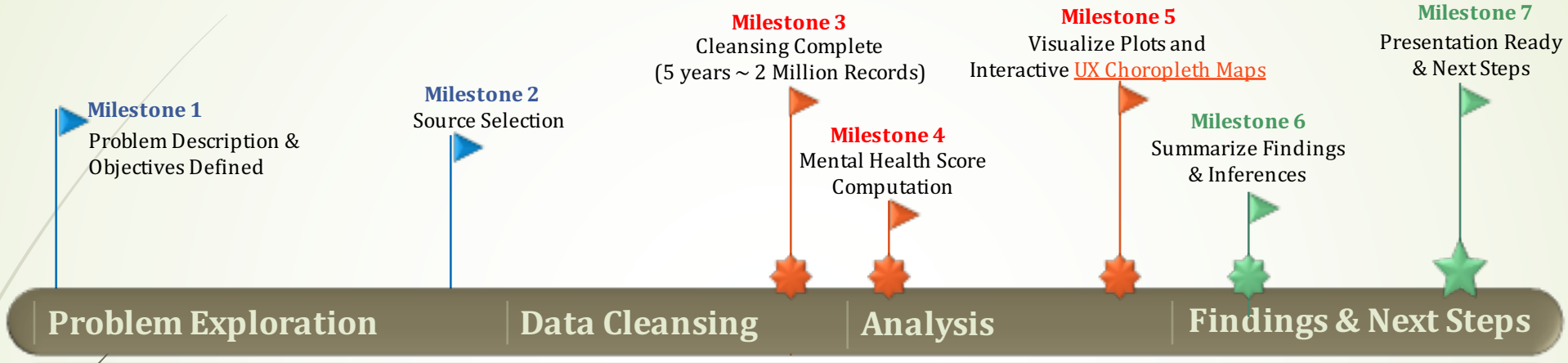
- GOALS
- SOURCES, ASSUMPTIONS & CONSTRAINTS
- APPROACH
- FINDINGS & NEXT STEPS (UX DEMO)

- THE PROCESS (DEMO)
- CHALLENGES AND OPPORTUNITIES

GOALS

- Identify how different factors affect Mental Health across all US population
- Identify how Mental Health vary across US (vs. New Jersey)
  - The US - By State
  - The US – By Rural vs Urban
  - By Gender
  - By Age
  - By Race
  - By Employment Status
- Check Trends Across last 5 years for Mental Health factors in the same areas (2012-16)
- Finally, check hypothesis & Identify Target groups

APPROACH:
THE JOURNEY & MILESTONES



PROBLEM & SOURCE

- Decide best source datasets
- Un-biased and Un-skewed was key
- Looked at Kaggle, Data.Gov & CDC

DECIDE & CLEANSE

- Decision on handling Nulls & others
- Mapping of FIPS code to State Code
- Processed last 5 year
- Picked 32/275 (11%) columns
- Mapping of SAS Metro Status Code to Rural vs Urban

ANALYZE & PLOT

- Deciding Mental Health State by a "Mental Health" score
- Using libraries that provide interactive UX like ipywidgets, plotly, seaborn and others
- More than 50+ graphs generated

INTERPRET & VISION

- Patterns across years
- Patterns in NJ vs US (Similarities vs Differences)
- Conclusions & Next Steps

SOURCES, ASSUMPTIONS & CONSTRAINTS

- **Sources:**

CDC (Center for Disease Control and Prevention) has a BRFSS system that collects telephone surveys across 50 US states, D.C., Guam and Puerto Rico (LAN & Mobile lines)

- **Assumptions:**
  - Reliable Data Survey System – In Existence for 33 years (~468K+ records on average across 5 years)
  - Number of Questions – Not all respondents willing to answer all significant questions (Appx. 275 questions per year)
- **Constraints:**
  - Any survey: Susceptible to State of mind at that particular instant of time of survey respondent
  - Phone survey: Susceptible to human error

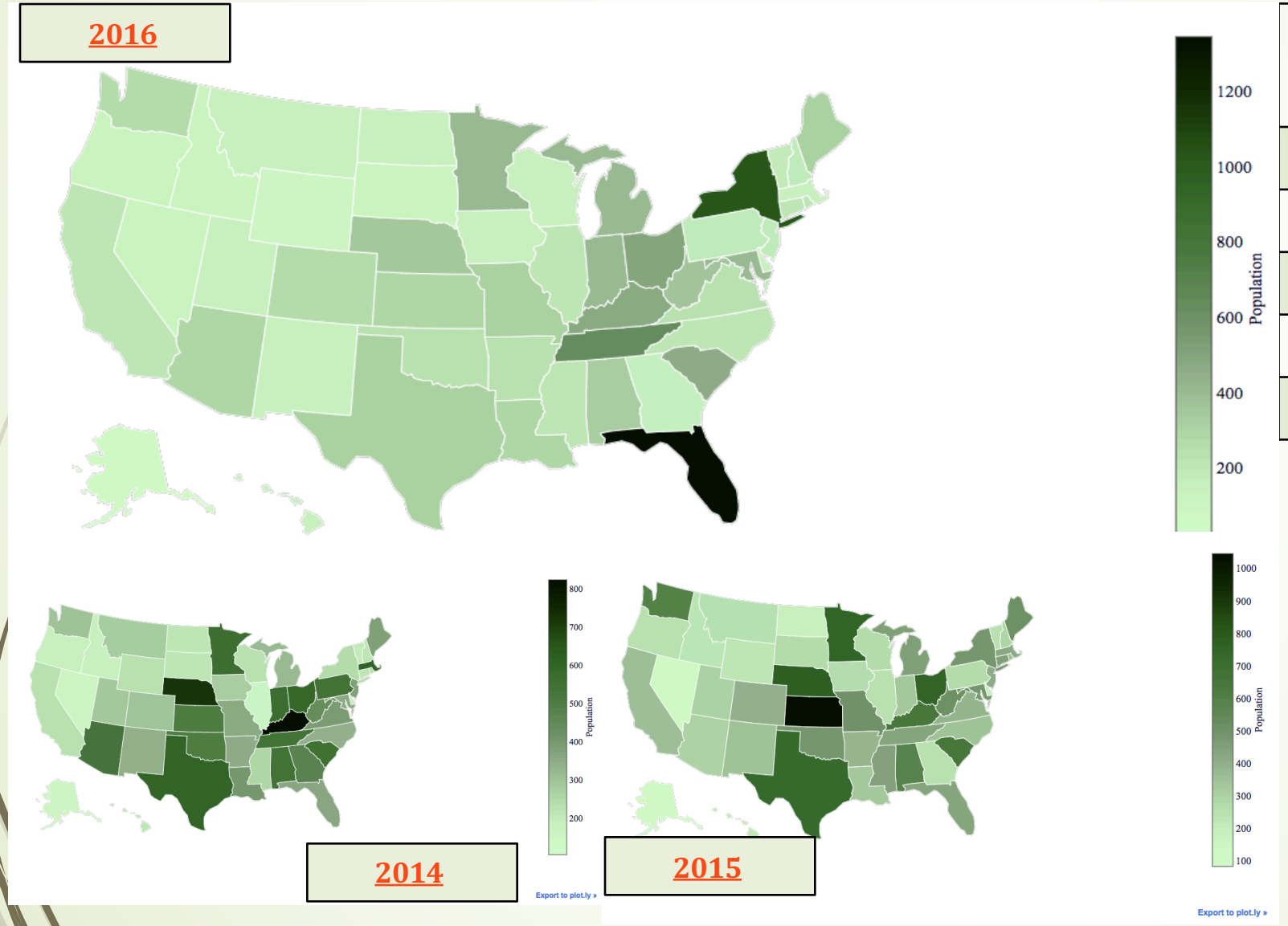
APPROACH: TOCHECK: IF WITH PROCESS MANAGING DATA VOLUME – STATS

Year	Volume of Records	File Sizes (SAS XPT files) (Initial)	% Reduction (↓) in "Cleansed Files" Sizes (Final)
2012	471,860	0.85 GB	0.06 GB (↓ 93%)
2013	486,843	0.80 GB	0.06 GB (↓ 93%)
2014	460,590	0.65 GB	0.06 GB (↓ 90%)
2015	437,462	1.14 GB	0.05 GB (↓ 96%)
2016	482,451	1.07 GB	0.06 GB (↓ 94%)

APPROACH:
DEFINTION OF "MENTAL HEALTH" STATE

- **"Mental Health Score": 23 Variables used to calculate it.** Factors Included in coming up with a score:
  - Sleep Deprivation Levels
  - Chronic Disease Conditions
  - General Health – Including BMI + Physical Inactivity
  - Income
  - Depression Related Factors (Chronic Depression, StressDays, DepressedDays)
  - Smoking Frequency
  - Alcohol Consumption Frequency
  - Incase of Women - Pregnancy
  - Emotional Support Availability & Life Satisfaction
- **Result: Weighted Average** Mental Health Score with higher weights given for factors historically found to highly impact "mental health" state.
  - Example: Stress and Depression
- **Higher Mental Health Score means** Highly Impacted Mental Health State

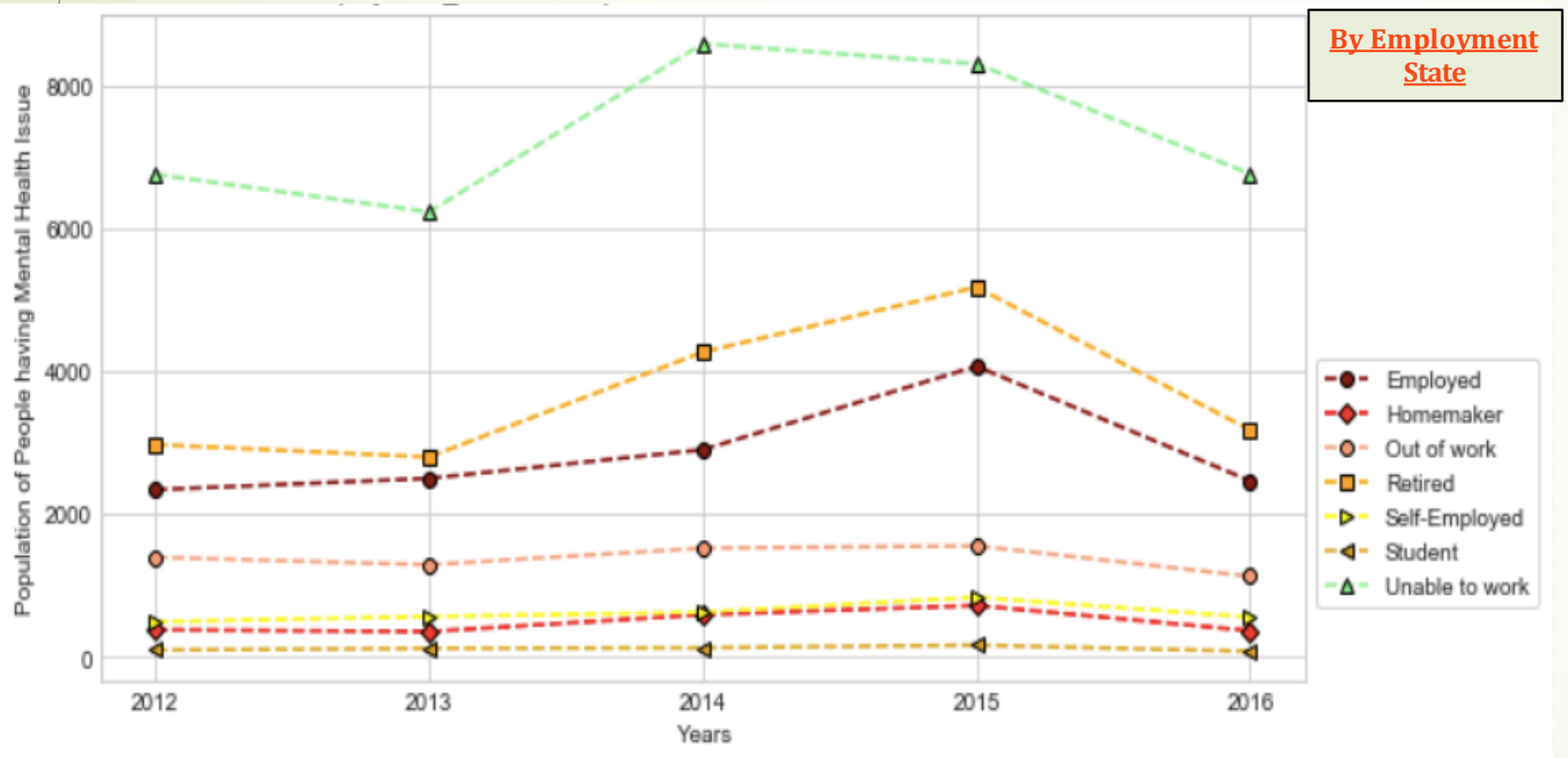
FINDINGS: MENTAL HEALTH STATE BY YEAR ACROSS US STATES (For details, click here)



Year	Highly Impacted State	Least Impacted State	NJ
2012	Ohio	Alaska	↓ 64%
2013	Florida	Alaska	↓ 80%
2014	Kentucky	Nevada	↓ 54%
2015	Kansas	Nevada	↓ 60%
2016	Florida	Alaska	↓ 85%

- Notice that Alaska and Nevada are always least impacted states whereas Highly Impacted varies
- East vs West nuances
- NJ vs. Highly Impacted State
- Interactive UX Demo

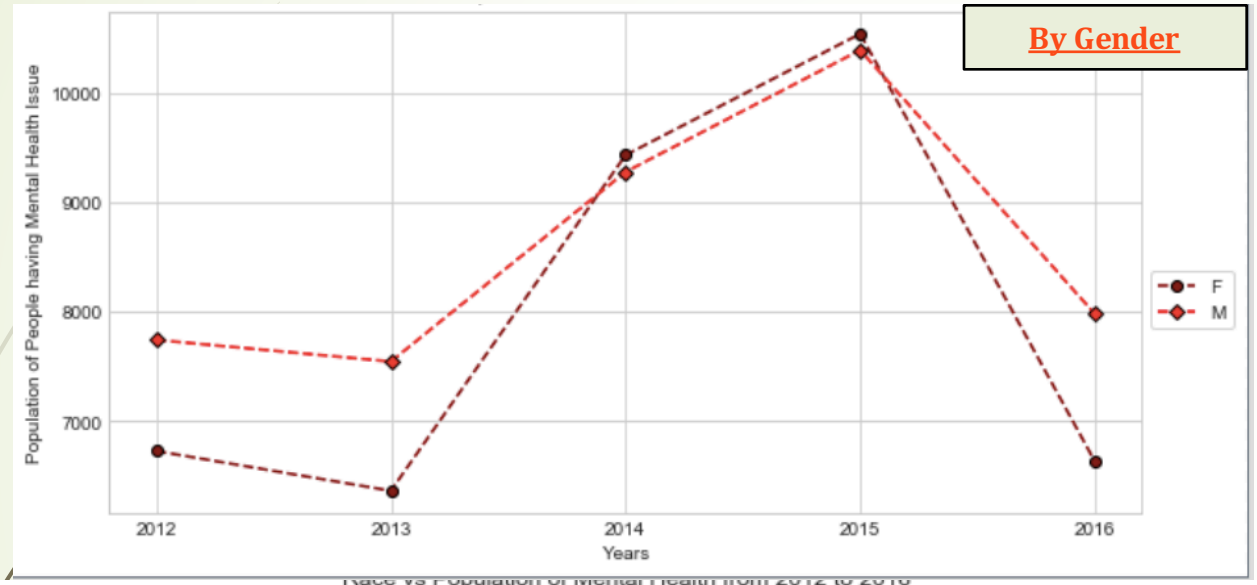
FINDINGS: MENTAL HEALTH ACROSS YEARS BY EMPLOYMENT STATE



- Unable to work vs out of work
- Higher Mental Health scores for Unable to work group.
- Lowest scores – Students
- Jump in 2014-15 attributed to External factors. E.g., the 2008 housing crisis impact till 2016 Q1

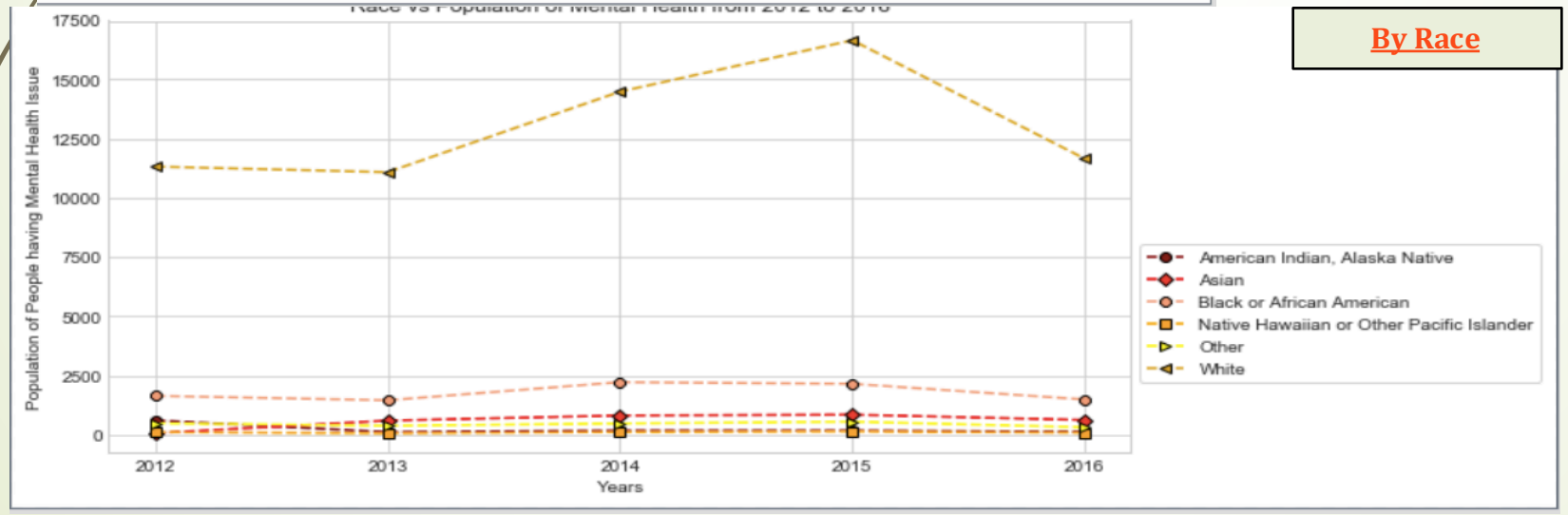
FINDINGS: MENTAL HEALTH ACROSS YEARS BY GENDER AND RACE

By Gender

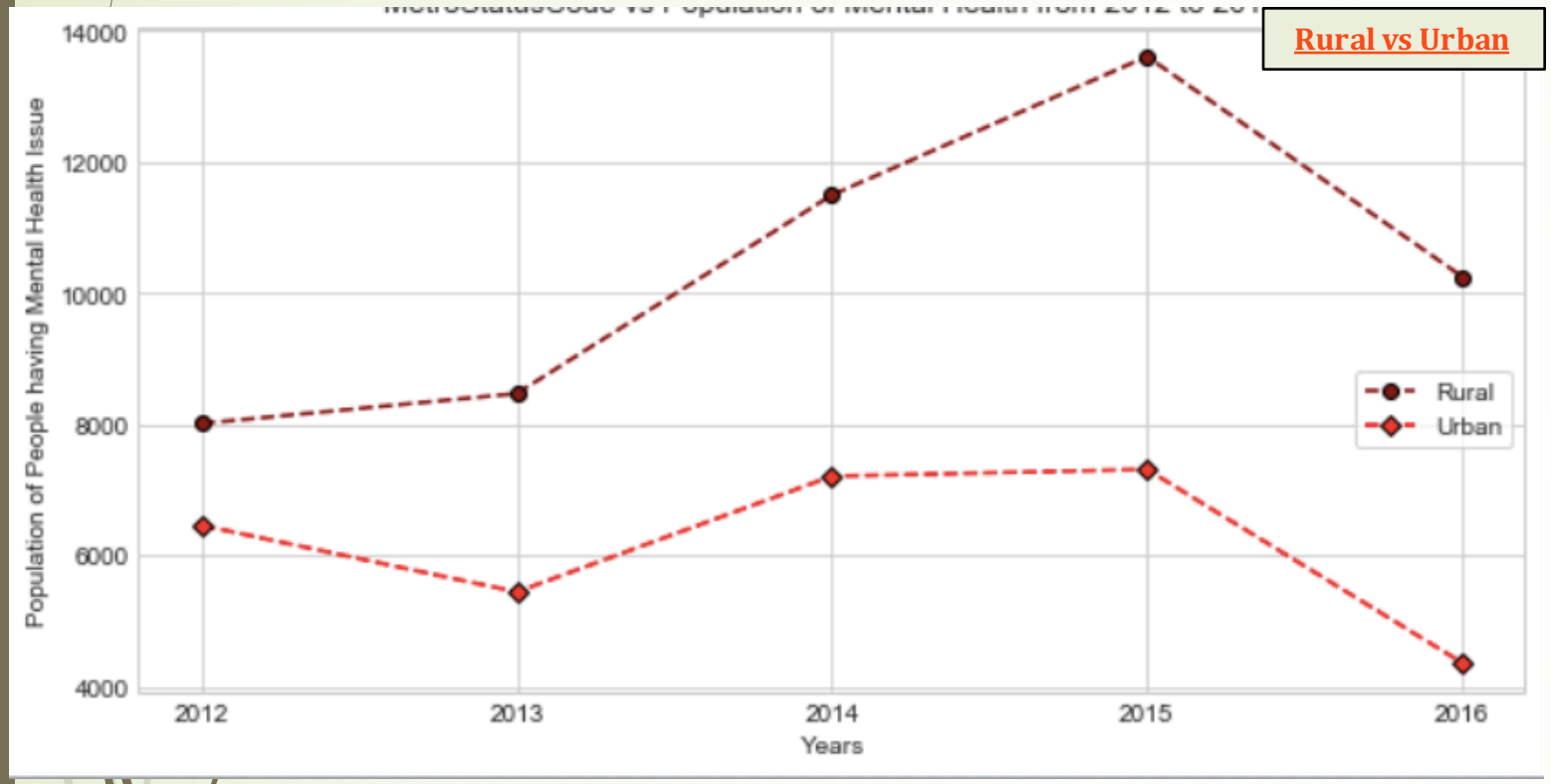


- Females more highly impacted than males except in 2014-15
- Count of White Americans impacted more than other races. Seconded by African-Americans. Count negligible in others
- Constraint: Reason for above – Possible that the Phone survey may have been with more white Americans than folks of other race or ethnicity – Need to Investigate

By Race



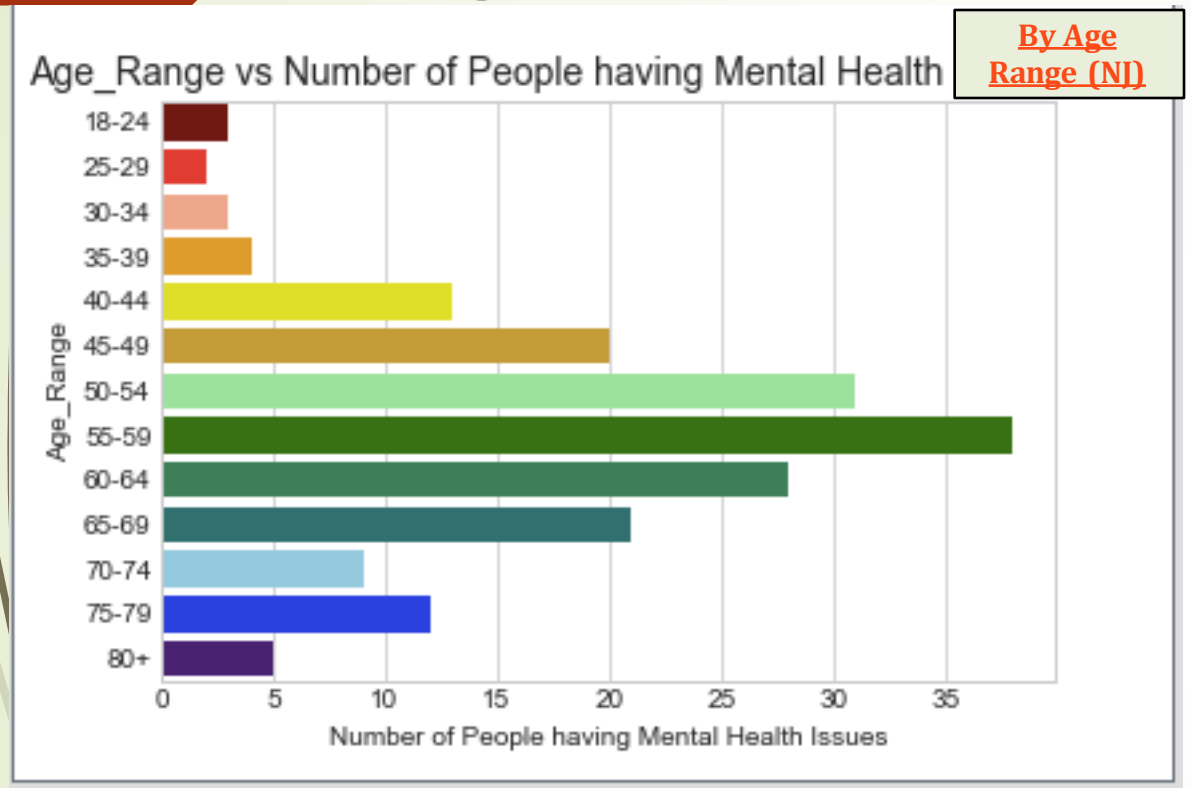
FINDINGS: MENTAL HEALTH ACROSS YEARS BY RURAL vs. URBAN & AGE

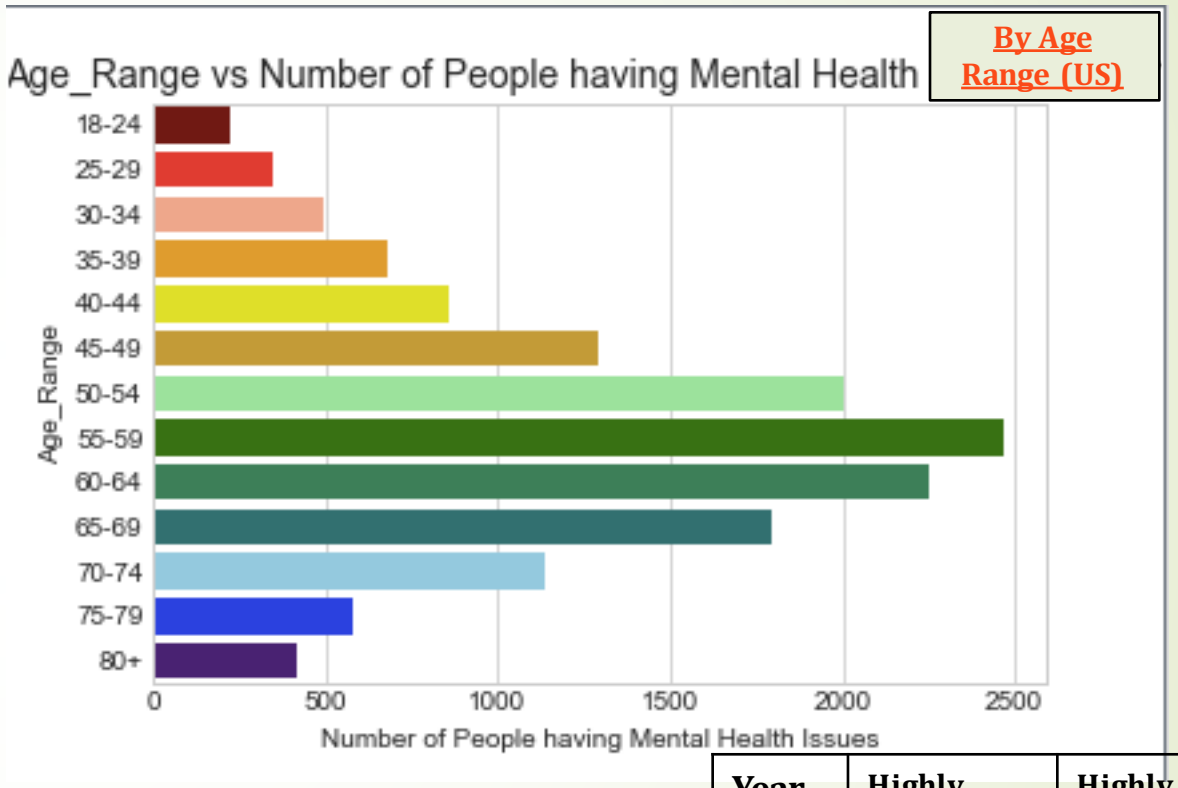


- Counter to Expectation – Trends across years shows Rural areas more highly impacted
- This may be different state stats (see here)
- Investigate more on here.
- By Age Buckets – Highly Impacted are in age-range of 55-59. Least impacted are in age range of 18-24. For details click here.

FINDINGS: MENTAL HEALTH BY YEAR

On Age (US vs NJ) For details, click here.



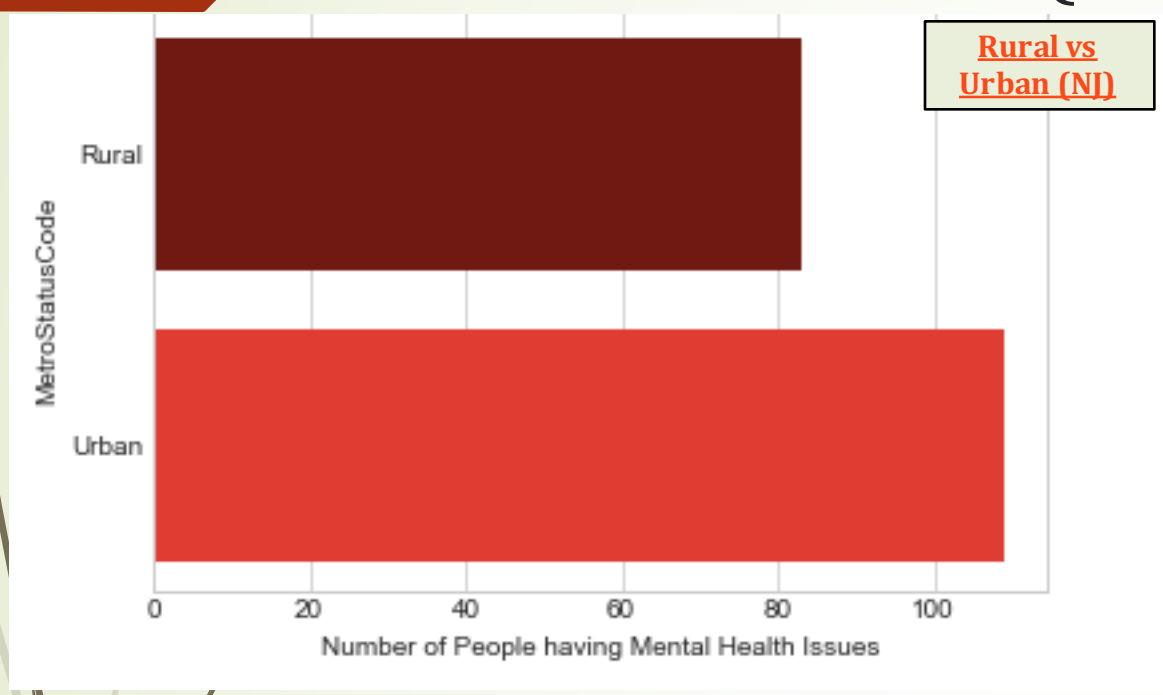


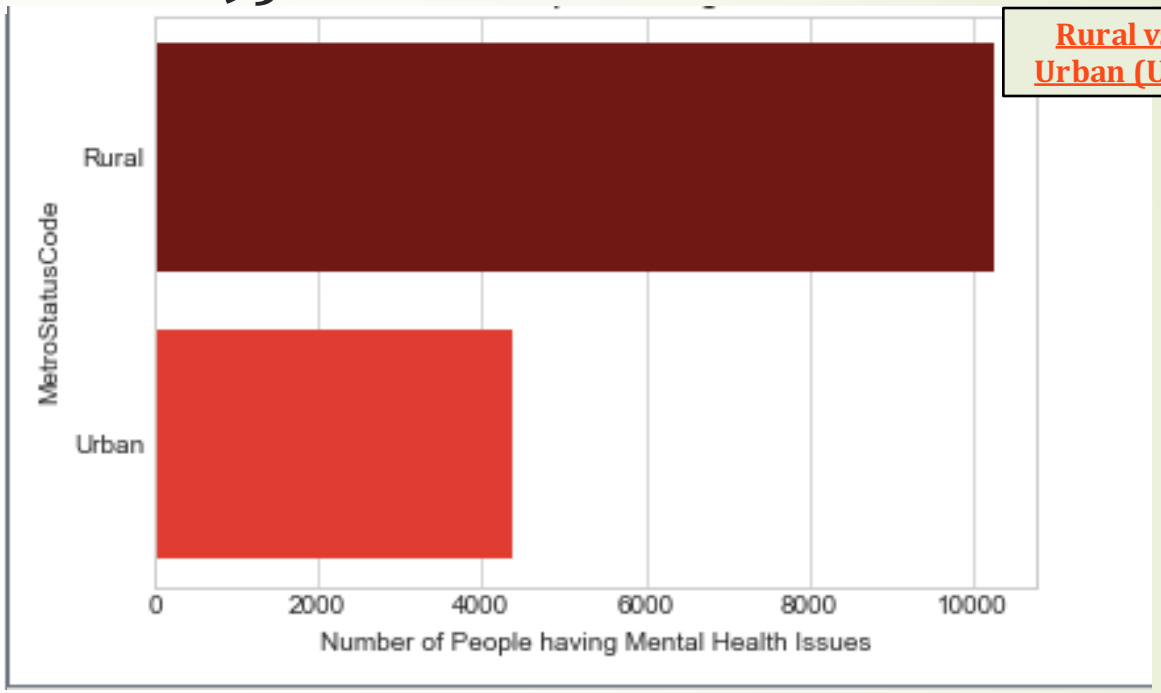
- Comparison of NJ vs US trends 60 different charts done across 5 years (12 per year)
- In NJ and the nation: Lowest impacted Age range continues to be steady at 18-24 age
- In NJ vs US: The Highly impacted age range increase from 2012-14 and then settled

Year	Highly Impacted Age (NJ)	Highly Impacted Age (US)
2012	60-64	55-59
2013	65-69	
2014	55-59	
2015	50-54	
2016	55-59	

FINDINGS: MENTAL HEALTH BY YEAR

On Rural vs Urban (US vs NJ) For details, click here.





- Across last 5 years, NJ has always shown Urban population as more highly impacted than Rural population – This is in contrast to US trends

SUMMARIZED FINDINGS & NEXT STEPS

Summarized Findings:

- Across States:
  - West vs East Nuances
- In NJ vs US Analysis:
  - **On Gender:** Male are more highly impacted across years compared to Female (narrow margin)– In Contrast to US except for 2016. Looks like US is following the trend seen in NJ
  - **On Health Coverage**: Contrast to belief, Population density having health coverage seem to be highly impacted than people with no coverage (Trends similar in NJ and US)
  - **On Employment State:** NJ Trends follow the nation trend of high impact seen with population unable to work (vs) lowest impact in Students
  - On Race: NJ Trends follows suit with national trend
- Jump in 2015 across the board – Higher stress reported across America External factors – e.g. After effects of housing crisis over many years, discrimination

Conclusions:

- The Target group for mental Health related marketing campaigns would be white Americans who are unable to work and have health coverage. The age range and targeting rural vs urban areas would depend on state vs nation wide campaigns

NEXT STEPS

- Investigate the factors influencing jump in 2015 more
- Investigate reason for Rural vs Urban impacts across US
- Focusing on Drugs – Target ads to the target group identified
- From a Governmental aspect: See why the health coverage is not truly showing a impact is a better mental health score.
- Investigate why the highly impacted states changes across years

THE PROCESS

- **Data Cleansing** (Show Code)
- **Analysis & Plots**
  - Calculate Variable =

    Smoker(10) + Alcoholic(10)+ General Health(10)+ Sleep Deprivation(10)+Chronic Disease(10)+BMI(10)+Physical Inactivity(10)+ Pregnancy(10) +Income(5) + Emotional Support Availability(2.5)+ Life Satisfaction(2.5)
  - To convert to a score of 20:
    AverageDepression=[Chronic_Depression(10)+StressDay(10)+DepressedDays(10)]*(2/3)
  - **Total Weight**=[(Calculate Variable + AverageDepression(20))/110(female)] or 100(male)]*100
  - Libraries used: UX Choropleth Maps, plotly, pandas, numpy, seaborn, matplotlib
  - **Demo** (Interactive Bar graph, UX Maps and Line graphs)

CHALLENGES & OPPORTUNITIES

Challenges:

- Original Scope included effectiveness of Drugs and governmental policies - Narrowed it down
- Massive volume of Data and managing to work with it on gitHub
- Exploration of APIs and which would work best for the Dataset we are looking at
- Summarizing the findings across 100+ graphs and sharing inferences

Opportunities:

- Investigate Gun control policies and security regulations in areas having the target group population higher in number
- Investigate how governmental policies can help target group that are unable to work