

# Investigating the Relationship Between Mental Health and Economic Inequality

**Team 6:** Mark Febrizio, Shumel Siraj, Alex Thiersch, and Xuan Zou

The George Washington University

DATS 6101: Introduction to Data Science

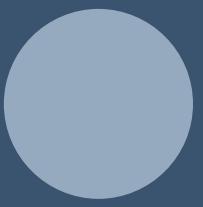


# Introduction

---



# Research Topic



## SMART Question:

- *What is the relationship between mental health and economic inequality within the United States from 2016 to 2021?*
- Null Hypothesis: There is no relationship between mental health and economic inequality that varies by geography.
- Alternative Hypothesis: There is a relationship between mental health and economic inequality that varies by geography.

○  
●  
○

## Background Information:

- Many scholars have studied the relationship between income inequality and human health, less research has focused on income inequality and people's mental health.
- An initial literature review suggests that the relationship between income inequality and mental health has not been conclusively answered ([Tibber et al. 2022](#)). Many articles use different measures and study different geographies.
- Therefore, this issue is worthy of further study

# Literature Review

**Bechtel et al. (2012):** determined that mental health is only adversely affected in the context of income inequality to a very small degree.

**Kelley & Evans (2017):** determined that there was no strong relationship between income inequality and subjective well-being.

**Layte (2012):** determined that there is an empirical consensus on the existence of an effect regarding income as a determinant of mental and physical health outcomes, but there is differing views on the explanation and causes of the effect

**Matthew & Brodersen (2018):** found a significant negative association between mental health and inequality.

**Picket & Wilkinson (2015):** determined that there is a strong causal link between income inequality and population, but a connection to mental health is less clear.

**Ribeiro et al. (2017):** found significant small effects that income inequality effects mental health.

**Sommet et al. (2018):** determined that income inequality affects the psychological health of individuals that are economically vulnerable.

**Tibber et al. (2022):** found that area-level income inequality is associated with poorer mental health.

**Zimmerman & Bell (2006):** found that the effect of income inequality on health may work through the differences in social capital among different racial groups.



# Description of the Dataset

## Dataset:

- Source: Robert Wood Johnson Foundation (RWJF)
- County Health Rankings & Roadmaps (2016 – 2021)
  - *Consists of economic and health data for all states at the county level in the United States*

## Variables:

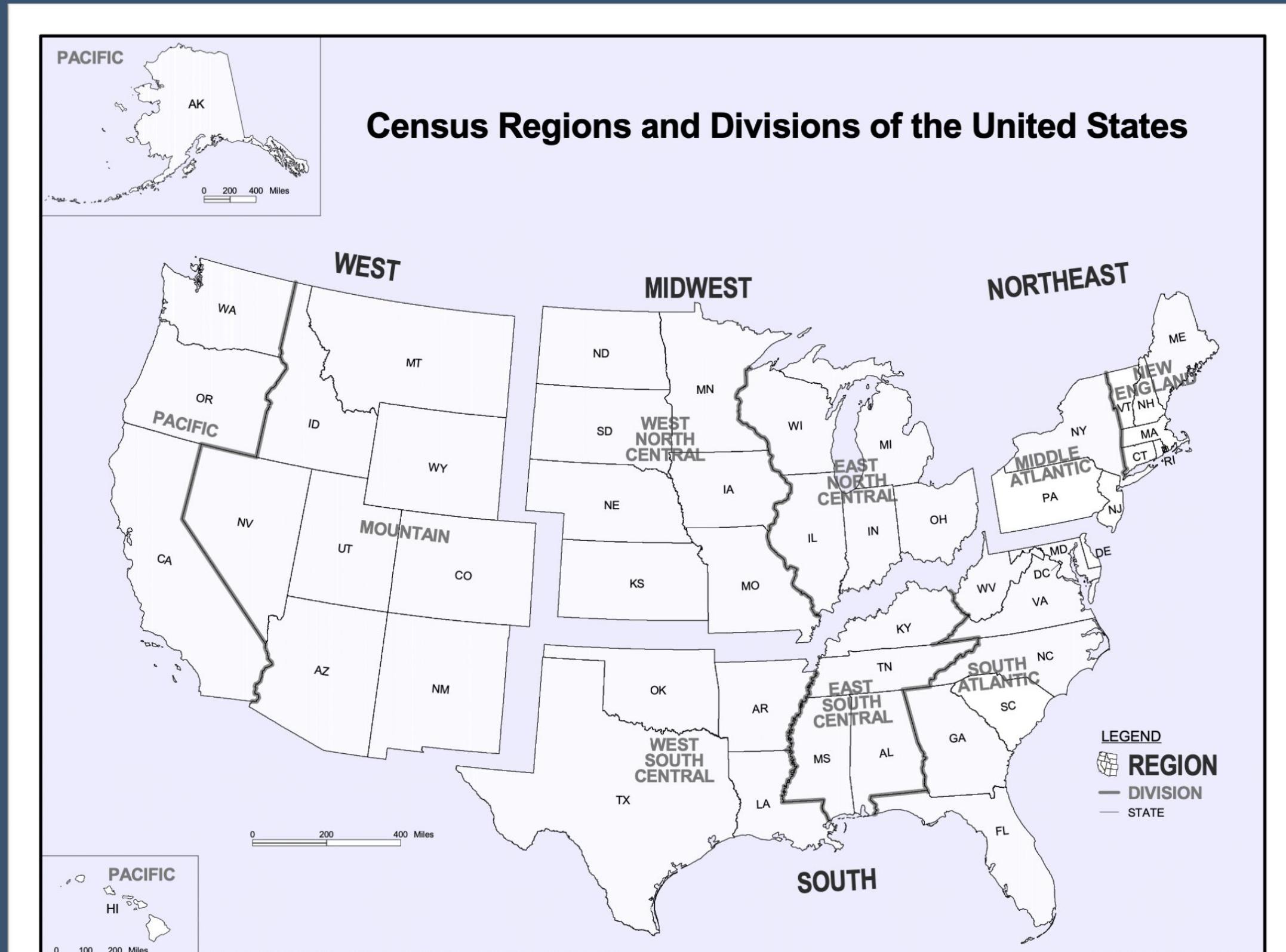
- Independent Variables:
  - **inequality** = Household Income Ratio
  - **median\_inc** = Median Household Income
  - **hs\_grad** = Percentage of high school graduates
  - **college** = Percentage of adults with some college education
  - **unempl** = Percentage of unemployment
  - **child\_poverty** = Percentage of child poverty
  - **single\_parent** = Percentage of single parent households
  - **severe\_housing** = Percentage of households with severe housing problems
  - **food\_index** = Food environment index
  - **mh\_providers** = Mental health provider rate
  - **pop\_provider\_ratio** = Mental health provider ratio
- Dependent Variables:
  - **mental\_health\_days** = Number of mentally unhealthy days
  - **mental\_distress\_rate** = Percentage frequent mental distress



# Regional Categories

## Four Regional Categories:

- Midwest
- Northeast
- South
- West



Source: U.S. Census Bureau

# Descriptive Statistics

## Dataset:

- RWJF County Health Rankings & Roadmaps (2016 – 2021)
  - *Consists of economic and health data for all states at the county level in the United States.*

Descriptive Statistics of Dataset														
	Var Num.	Count	Mean	Std. Dev.	Median	Trimmed Mean	MAD	Minimum	Maximum	Range	Skewness	Kurtosis	S.E.	
mental_health_days	1	18469	4.04e+00	6.92e-01	4.02e+00	4.03e+00	7.06e-01	2.100	7.29e+00	5.19e+00	0.223	-0.113	0.005	
mental_distress_rate	2	18469	1.26e-01	2.40e-02	1.24e-01	1.25e-01	2.40e-02	0.066	2.47e-01	1.81e-01	0.532	0.301	0.000	
inequality	3	18465	4.52e+00	7.33e-01	4.41e+00	4.45e+00	6.28e-01	2.543	1.20e+01	9.43e+00	1.254	3.431	0.005	
median_inc	4	18469	5.08e+04	1.36e+04	4.87e+04	4.93e+04	1.10e+04	21658.000	1.52e+05	1.30e+05	1.424	3.680	99.952	
hs_grad	5	16563	8.70e-01	7.90e-02	8.83e-01	8.78e-01	6.70e-02	0.025	1.00e+00	9.75e-01	-1.592	5.592	0.001	
college	6	18469	5.71e-01	1.16e-01	5.72e-01	5.72e-01	1.22e-01	0.152	9.11e-01	7.59e-01	-0.088	-0.263	0.001	
unempl	7	18469	5.00e-02	2.00e-02	4.60e-02	4.80e-02	1.70e-02	0.012	2.40e-01	2.28e-01	1.690	6.491	0.000	
child_poverty	8	18469	2.21e-01	9.10e-02	2.09e-01	2.14e-01	9.00e-02	0.024	7.47e-01	7.23e-01	0.695	0.536	0.001	
single_parent	9	18469	3.14e-01	1.06e-01	3.06e-01	3.08e-01	9.60e-02	0.000	8.72e-01	8.72e-01	0.696	1.250	0.001	
severe_housing	10	18470	1.42e-01	4.70e-02	1.37e-01	1.39e-01	3.70e-02	0.022	7.13e-01	6.91e-01	2.086	14.165	0.000	
food_index	11	18394	7.32e+00	1.18e+00	7.50e+00	7.44e+00	1.04e+00	0.000	1.00e+01	1.00e+01	-1.385	3.715	0.009	
mh_providers	12	17028	1.00e-03	2.00e-03	1.00e-03	1.00e-03	1.00e-03	0.000	2.40e-02	2.40e-02	3.457	22.562	0.000	
pop_provider_ratio	13	17028	2.00e+03	2.84e+03	9.90e+02	1.38e+03	8.97e+02	-957.000	5.49e+04	5.58e+04	4.291	34.700	21.797	

# Descriptive Statistics - Std. Dev. & Variance

- The small standard deviations and variances for each variable suggest that the data for each variable are relatively close to the mean.
- In other words, the data for most variables has low spread. Median income seems to be an exception.

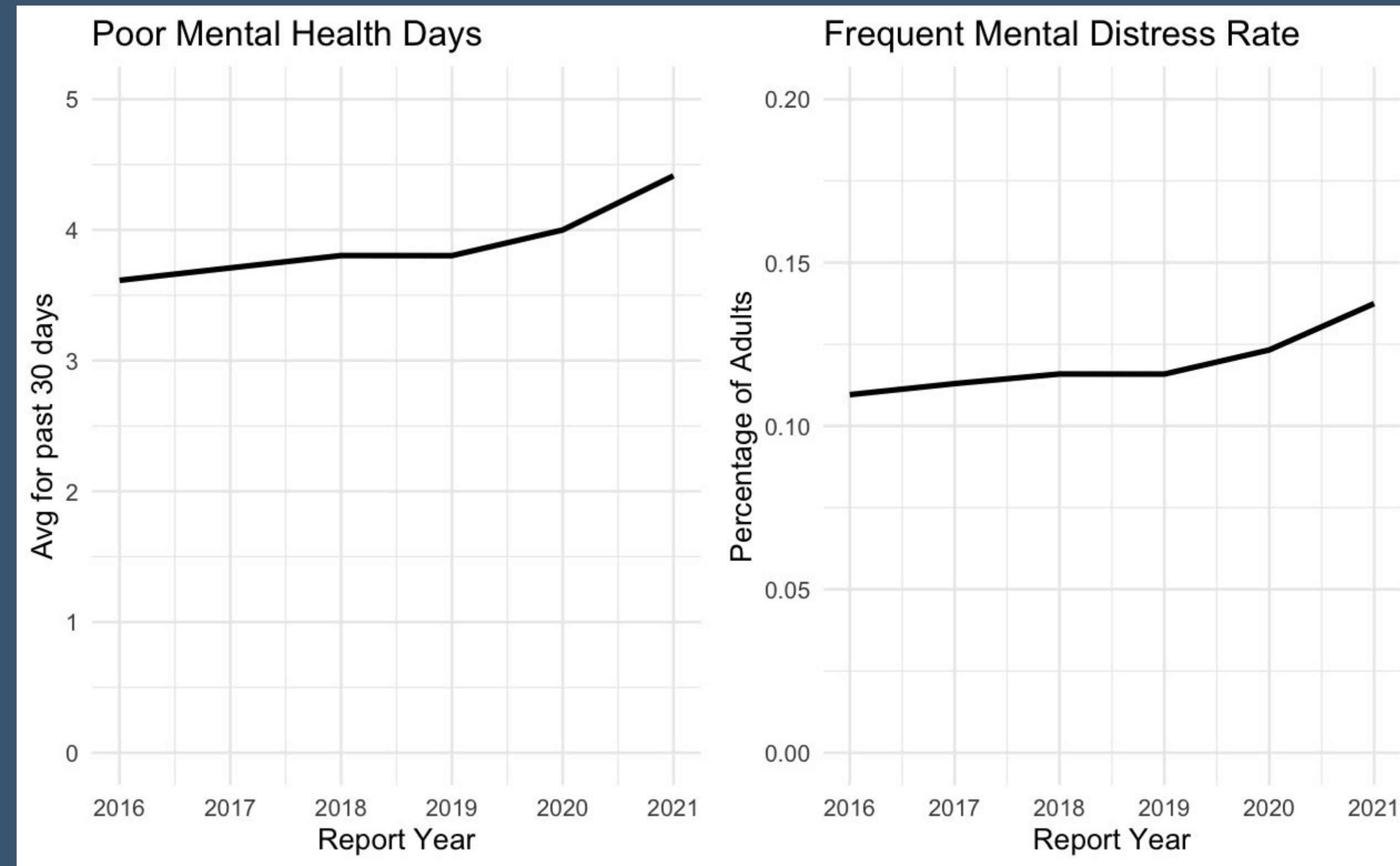
Descriptive Statistics of Dataset - Std. Dev. & Variance			
	Var Num.	Std. Dev.	Variance
mental_health_days	1	6.92e-01	4.78e-01
mental_distress_rate	2	2.40e-02	1.00e-03
inequality	3	7.33e-01	5.38e-01
median_inc	4	1.36e+04	1.85e+08
hs_grad	5	7.90e-02	6.00e-03
college	6	1.16e-01	1.30e-02
unempl	7	2.00e-02	0.00e+00
child_poverty	8	9.10e-02	8.00e-03
single_parent	9	1.06e-01	1.10e-02
severe_housing	10	4.70e-02	2.00e-03
food_index	11	1.18e+00	1.40e+00
mh_providers	12	2.00e-03	0.00e+00
pop_provider_ratio	13	2.84e+03	8.09e+06

# Annual Trends

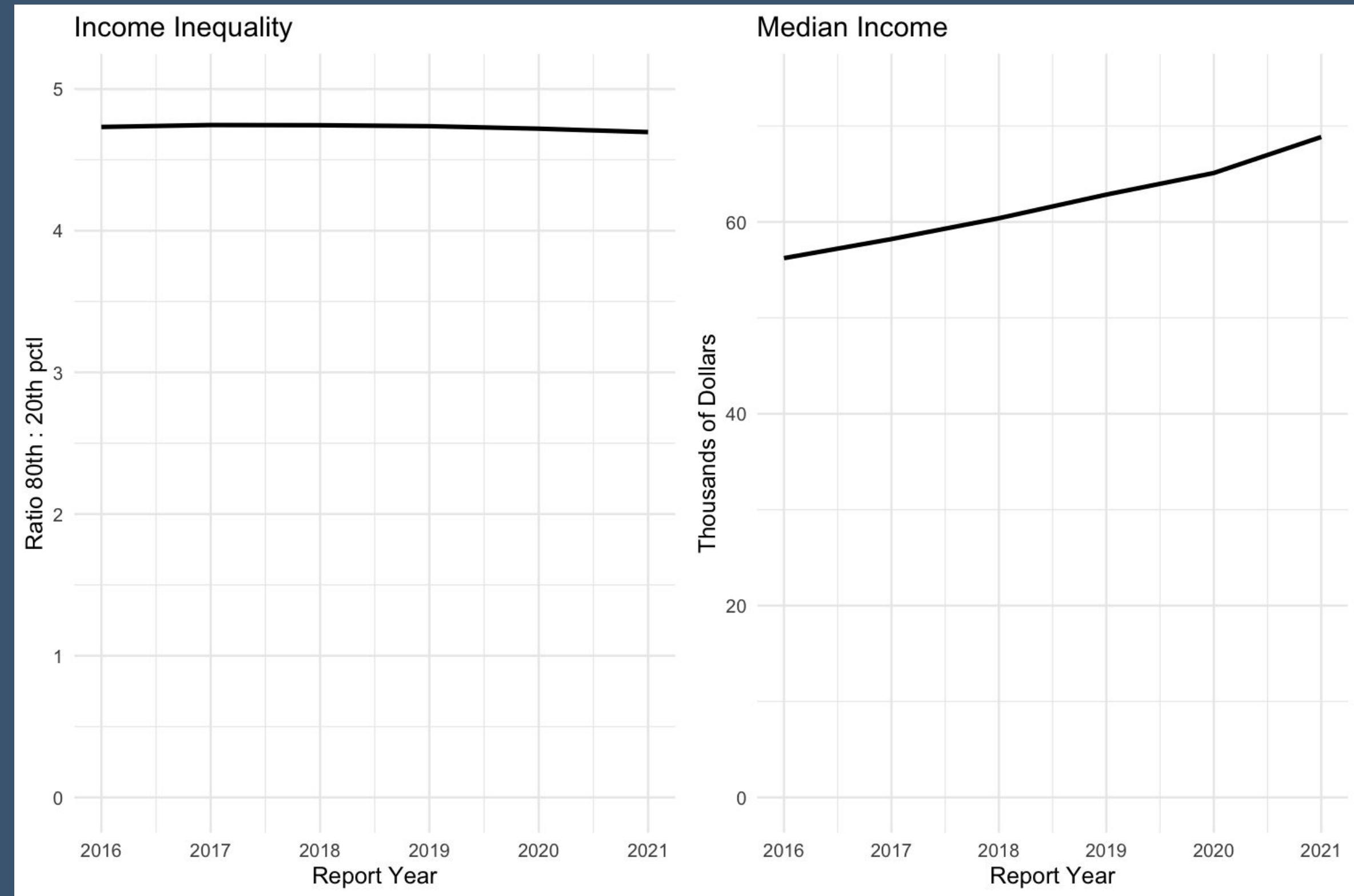
---



# Annual Trends: Mental Health Vars.



# Annual Trends: Economic Inequality Vars.

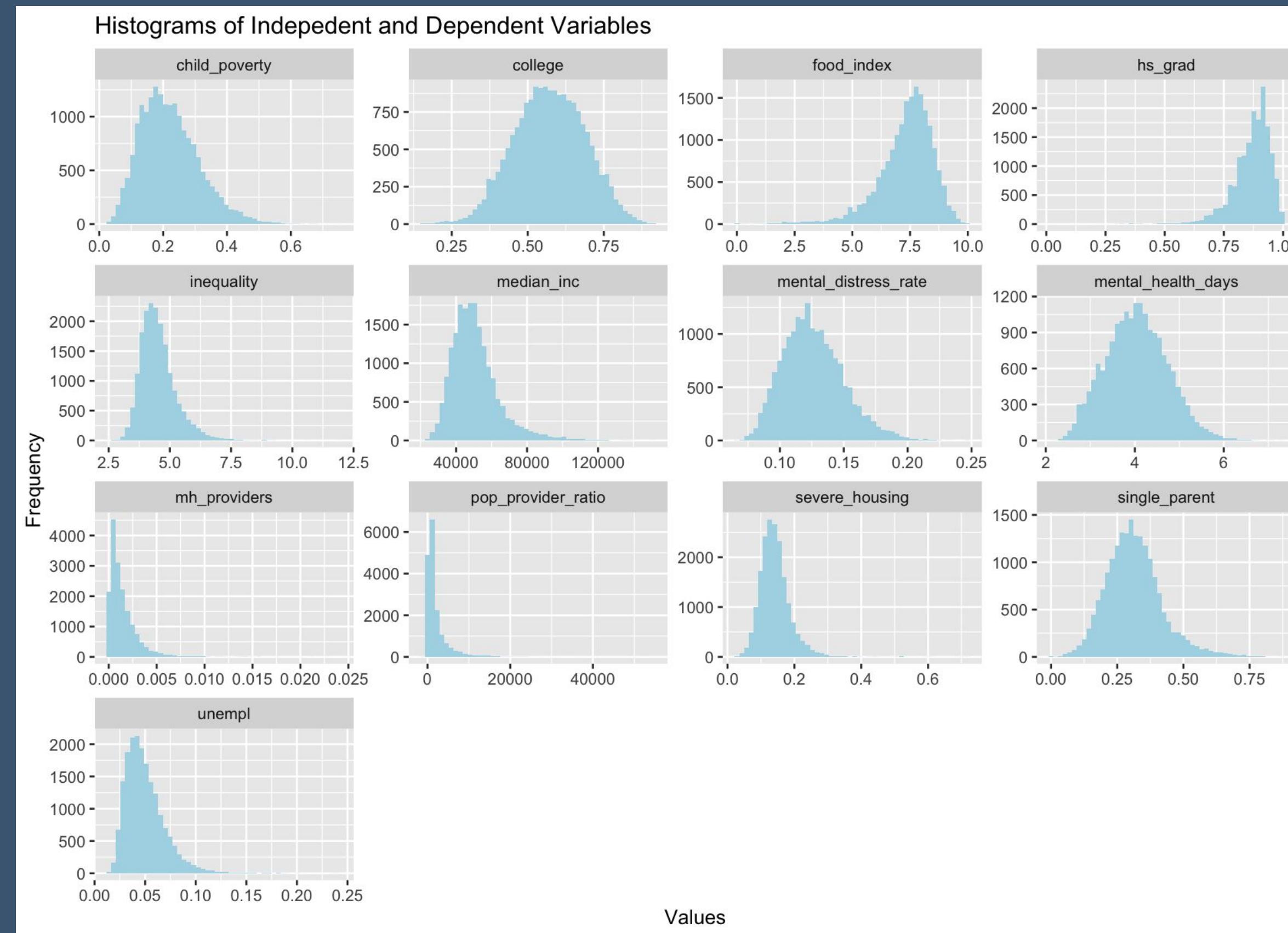
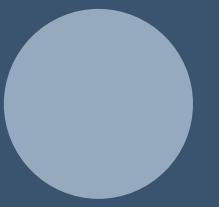


# Normality Plots & Tests

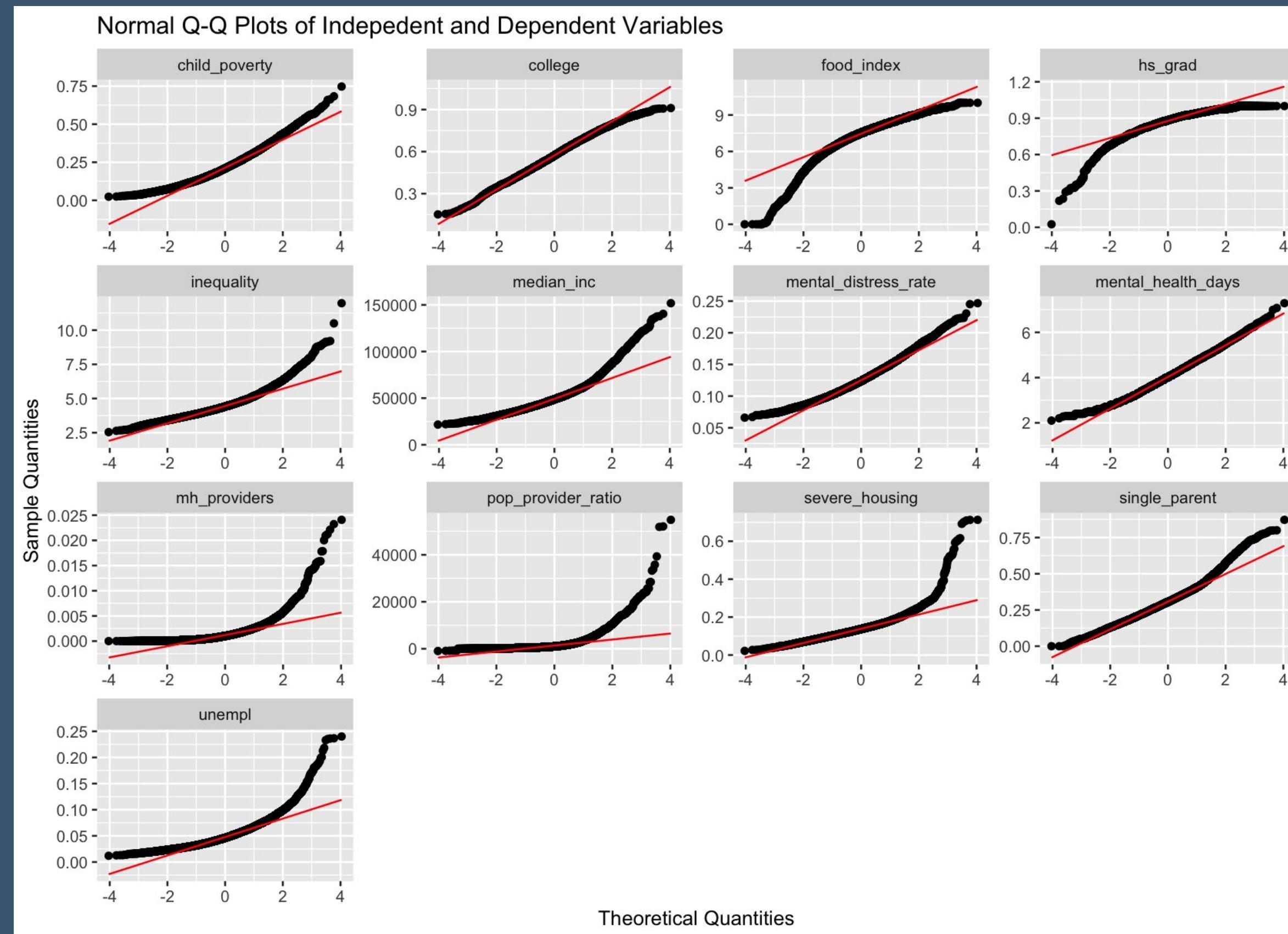
---



# Histograms: Independent and Dependent Vars.



# Normal Q-Q Plots: Independent and Dependent Vars.



# Normality Test: Shapiro-Wilks

*All variables were shortened to the first 5,000 obs. due to Shapiro-Wilks Test limitations.*

- All p-values were <0.05, indicating that the distribution of the data is significantly different from a normal distribution.

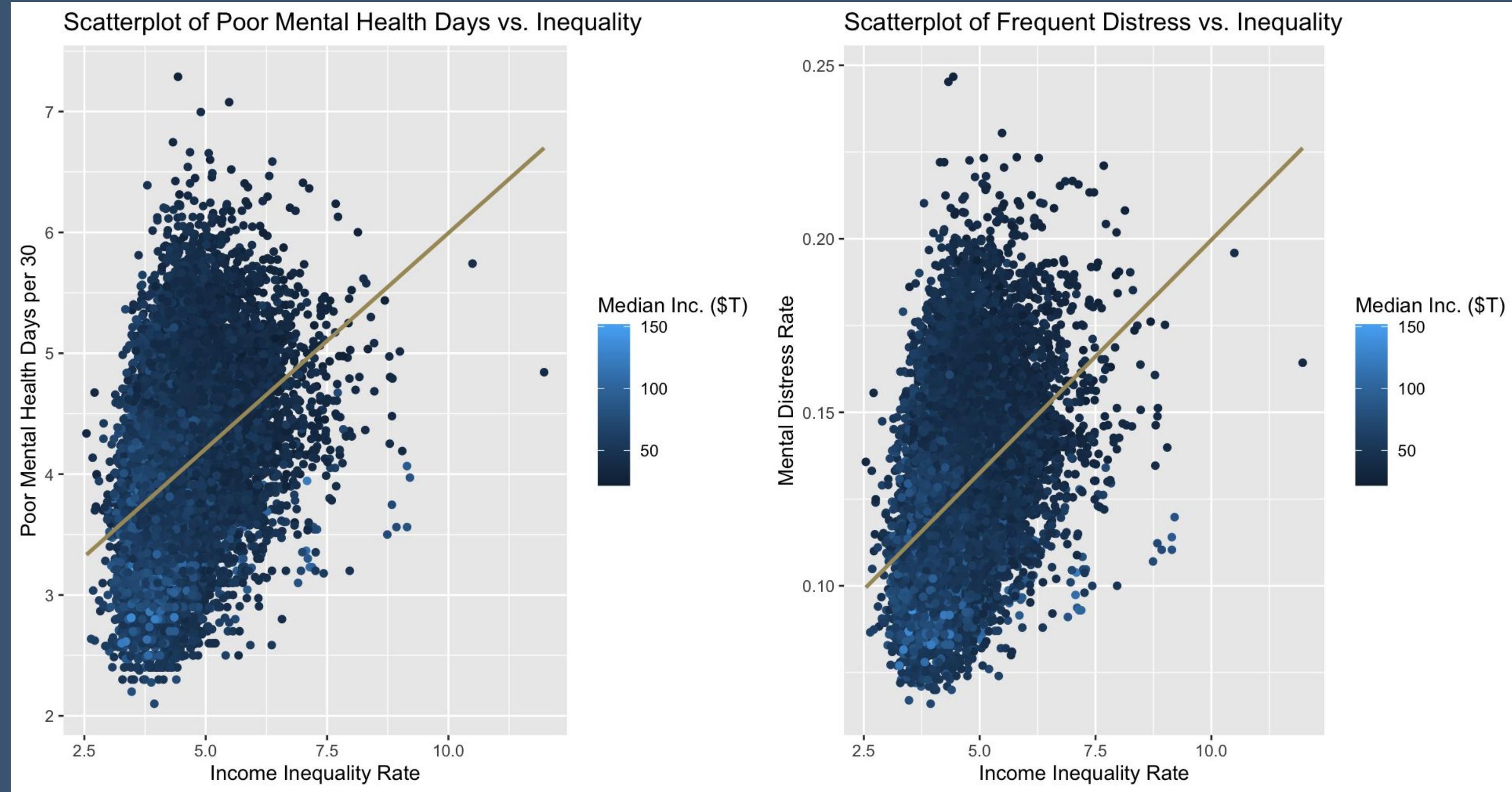
Shapiro-Wilk normality test		
Variable	W	p-value
mental_health_days	1	<2e-16
mental_distress_rate	1	<2e-16
inequality	0.9	<2e-16
median_inc	0.9	<2e-16
hs_grad	0.9	<2e-16
college	1	1e-08
unemploy	0.9	<2e-16
child_poverty	1	<2e-16
single_parent	1	<2e-16
severe_housing	0.9	<2e-16
food_index	0.9	<2e-16
mh_providers	0.7	<2e-16
pop_provider_ratio	0.6	<2e-16

# Scatterplots

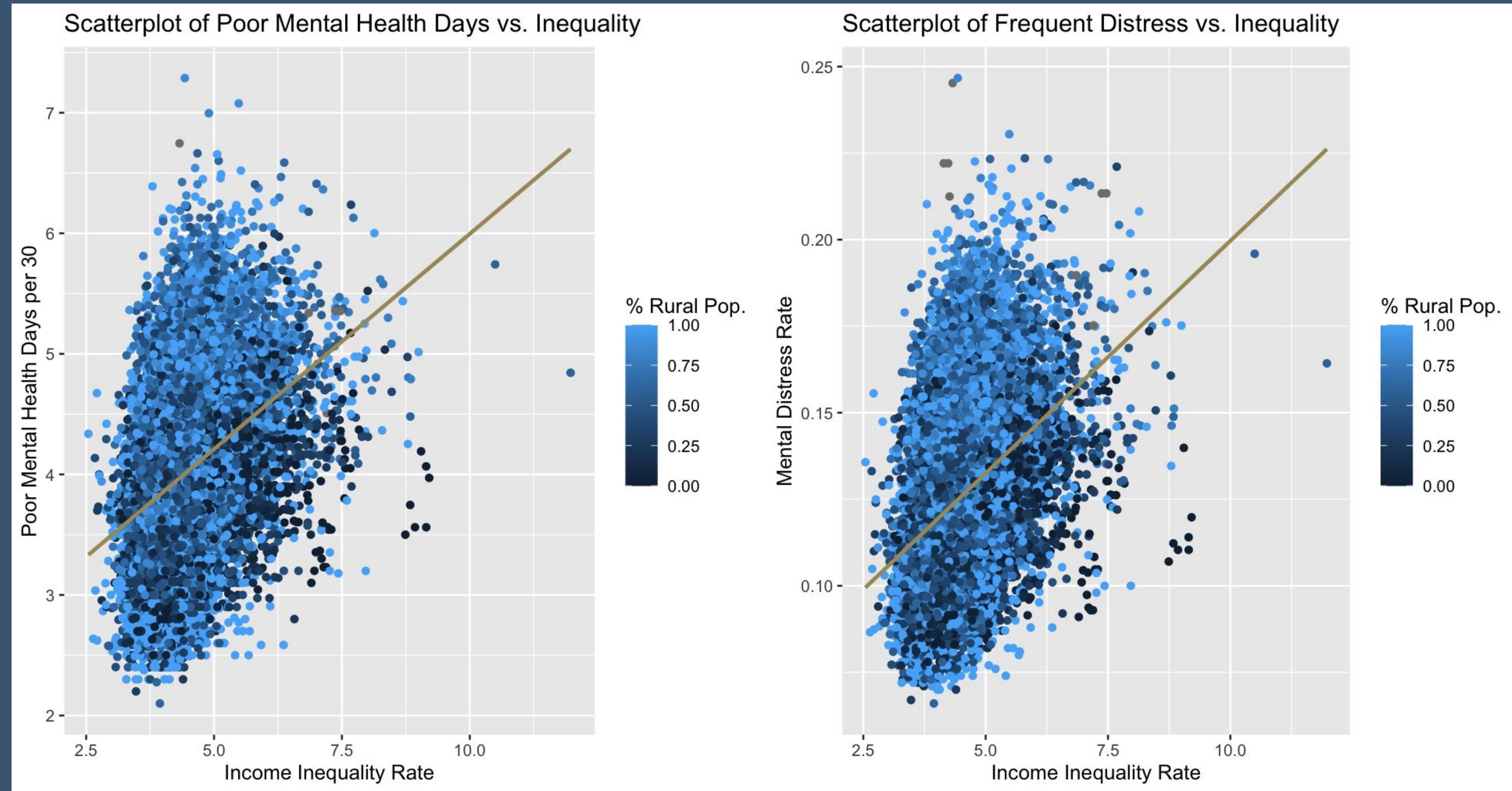
---



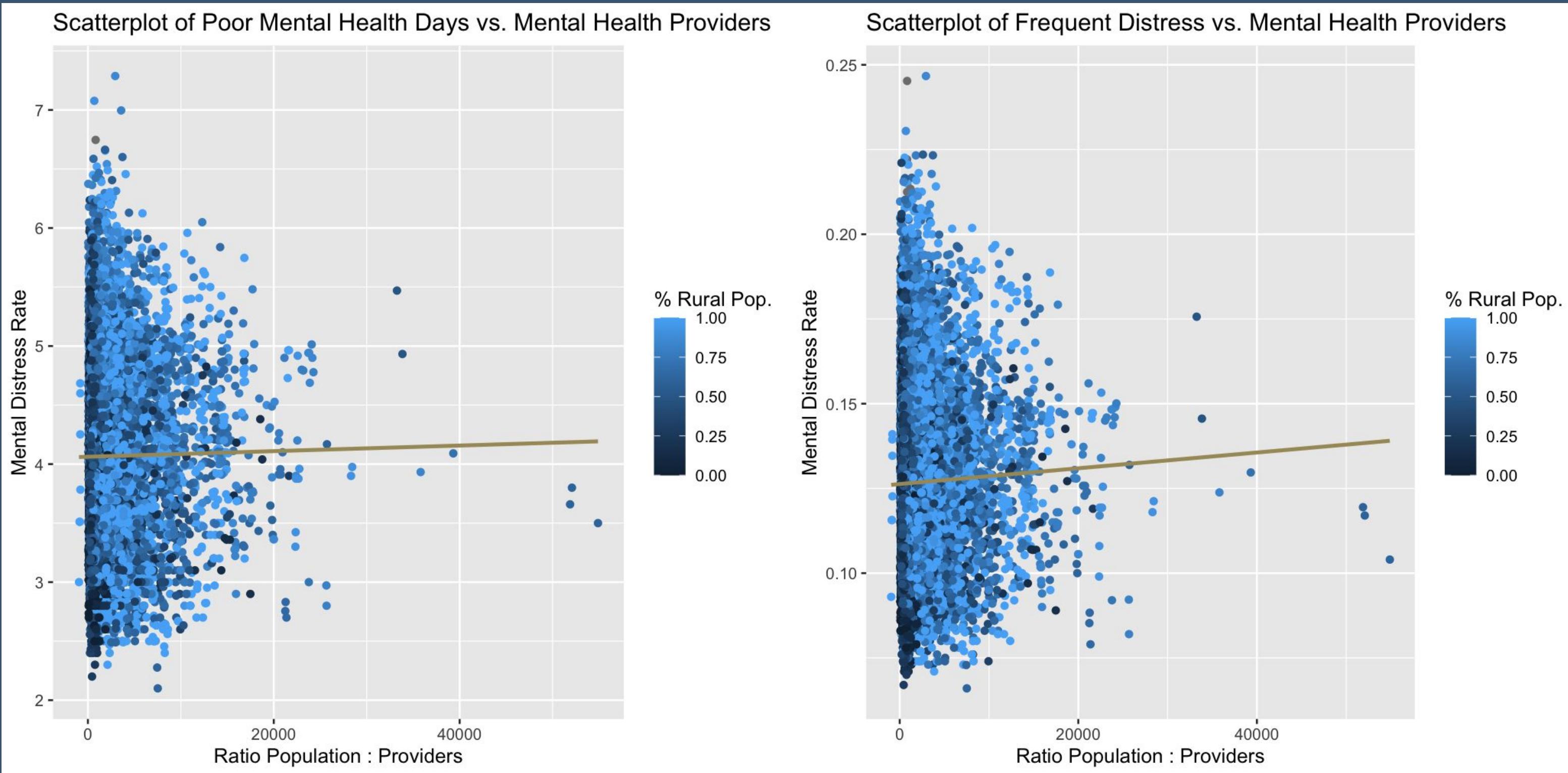
# Scatterplots: Mental Distress v. Inequality



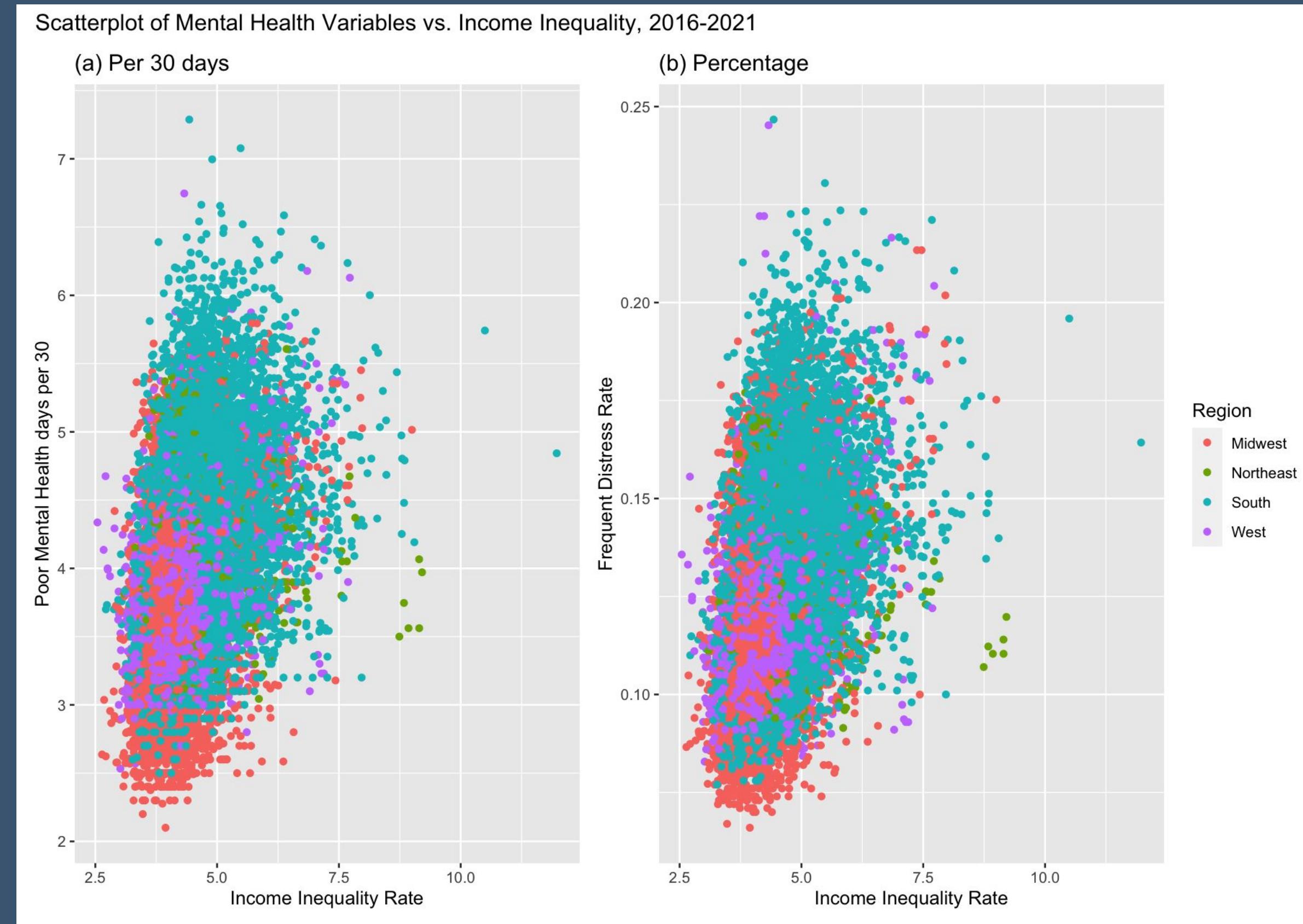
# Scatterplots: Mental Distress v. Inequality



# Scatterplots: Mental Distress v. Availability of Care



# Regional Scatterplots: Mental Health Vars. vs. Inequality



# Boxplots

---



# Description of the Boxplots

## Four Regions:

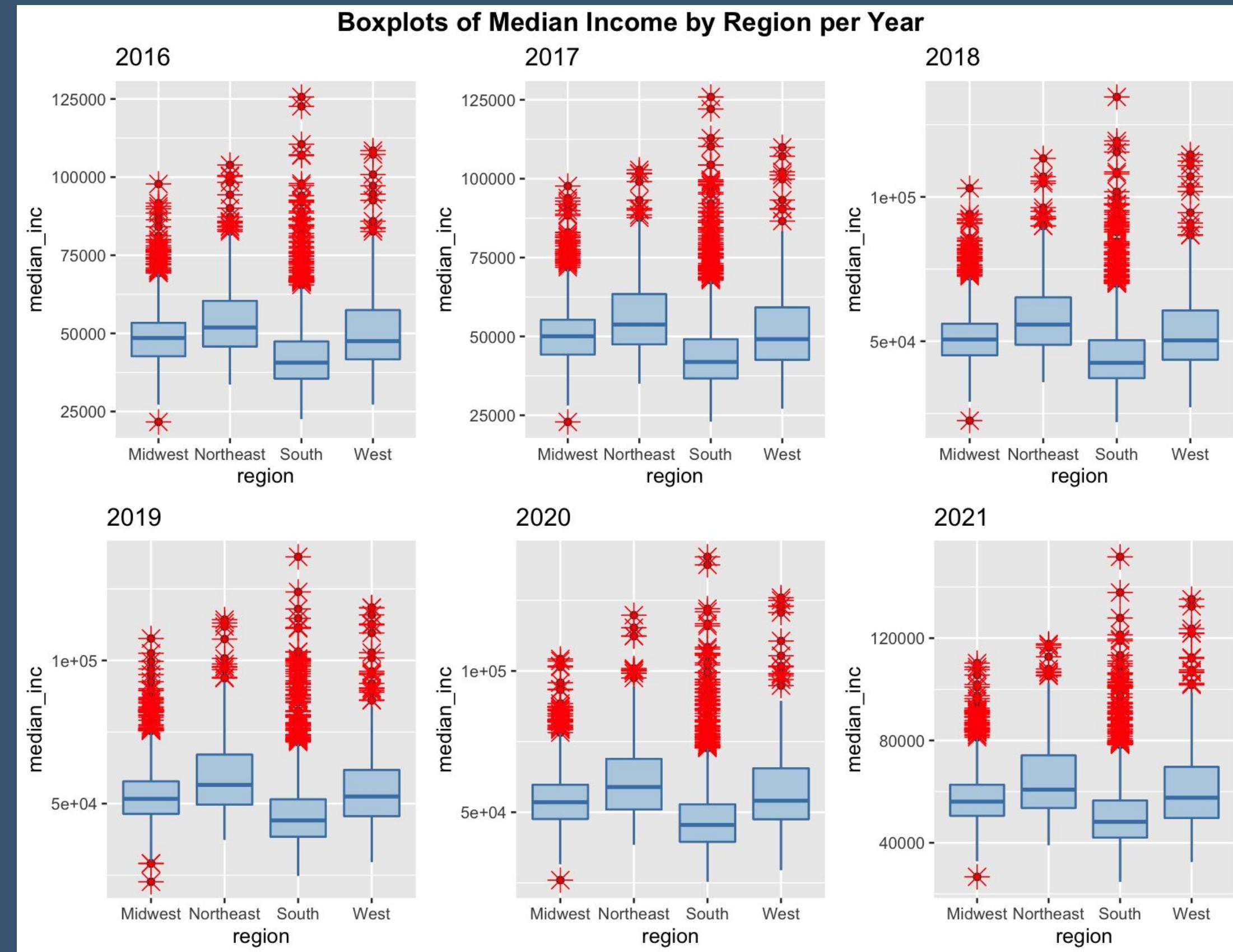
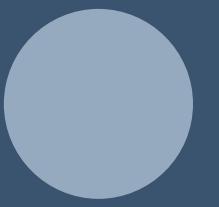
- Northeast, South, West, Midwest

## Variables:

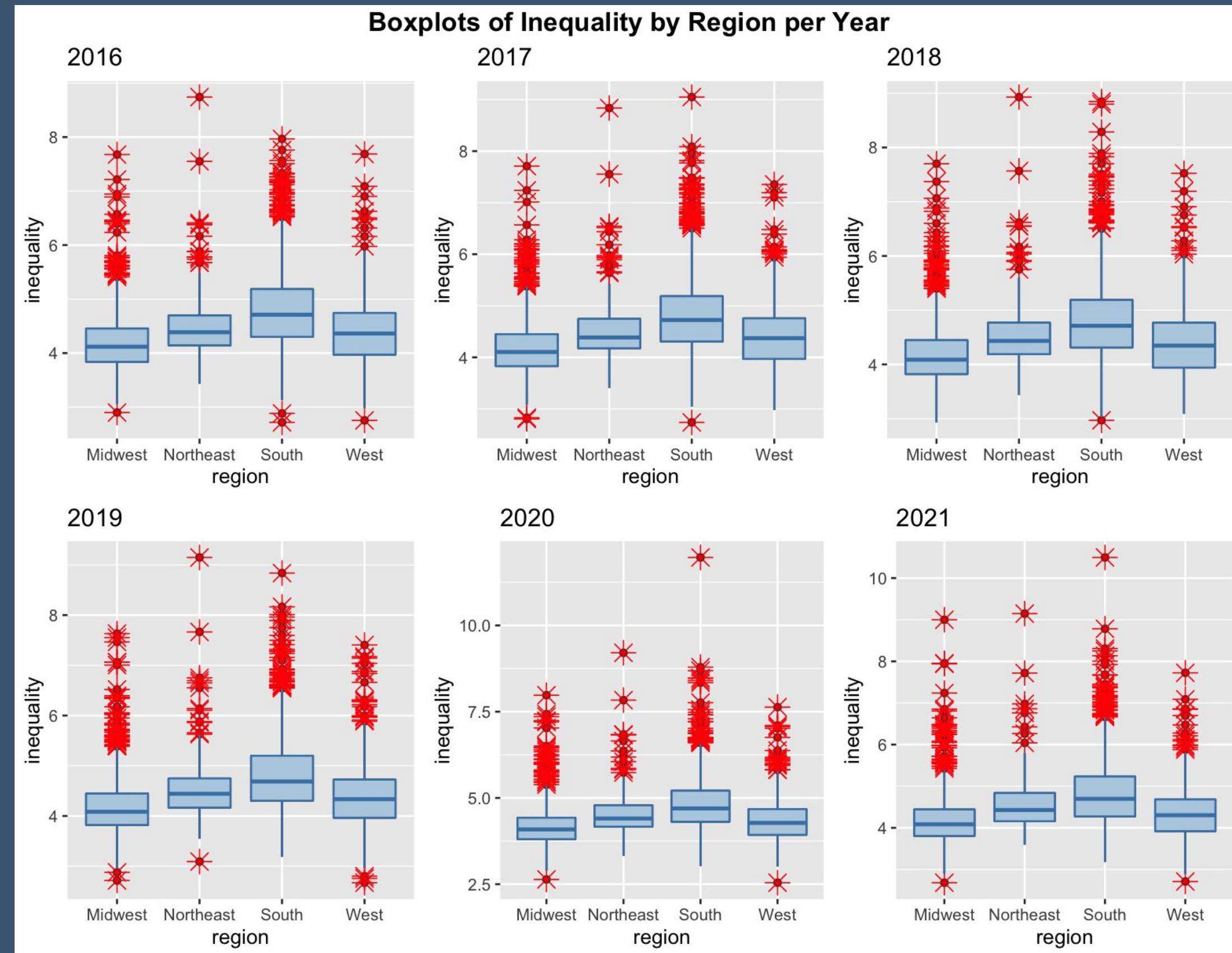
- Independent Variables:
  - **inequality** = *Household Income Ratio*
  - **median\_inc** = *Median Household Income*
- Dependent Variables:
  - **mental\_health\_days** = *Number of mentally unhealthy days*



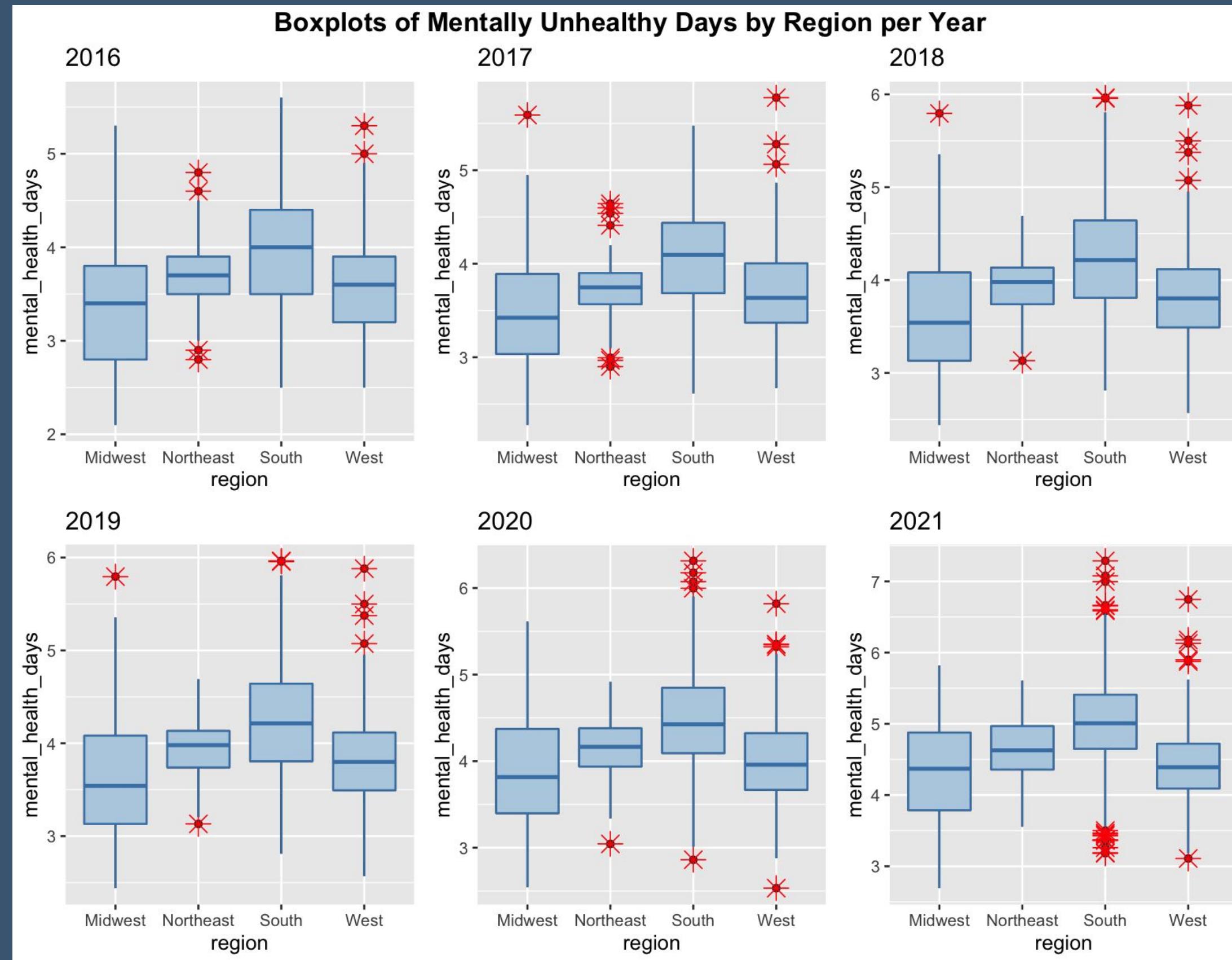
# Boxplots: Median Income by Region per Year



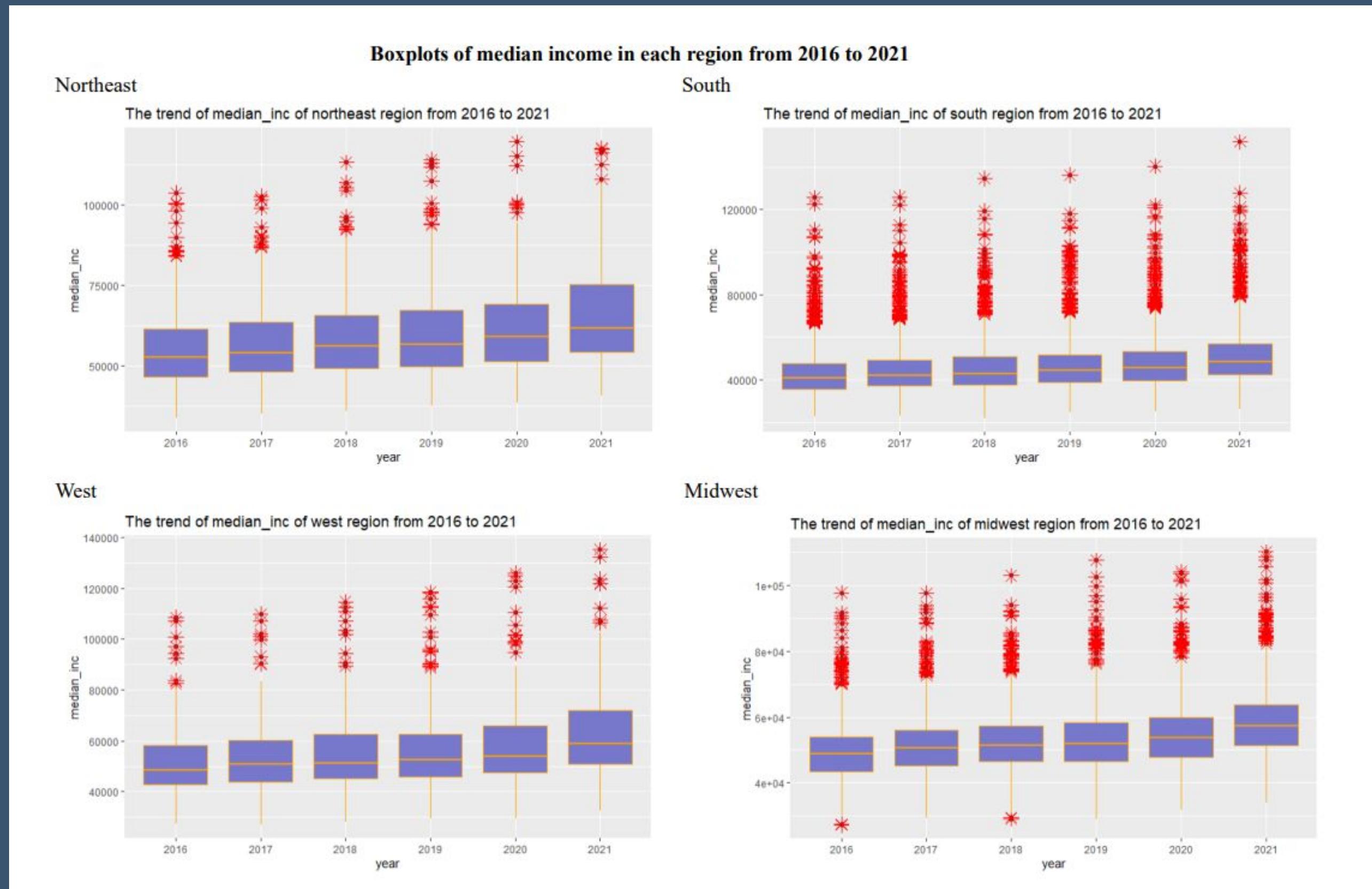
# Boxplots: Inequality by Region per Year



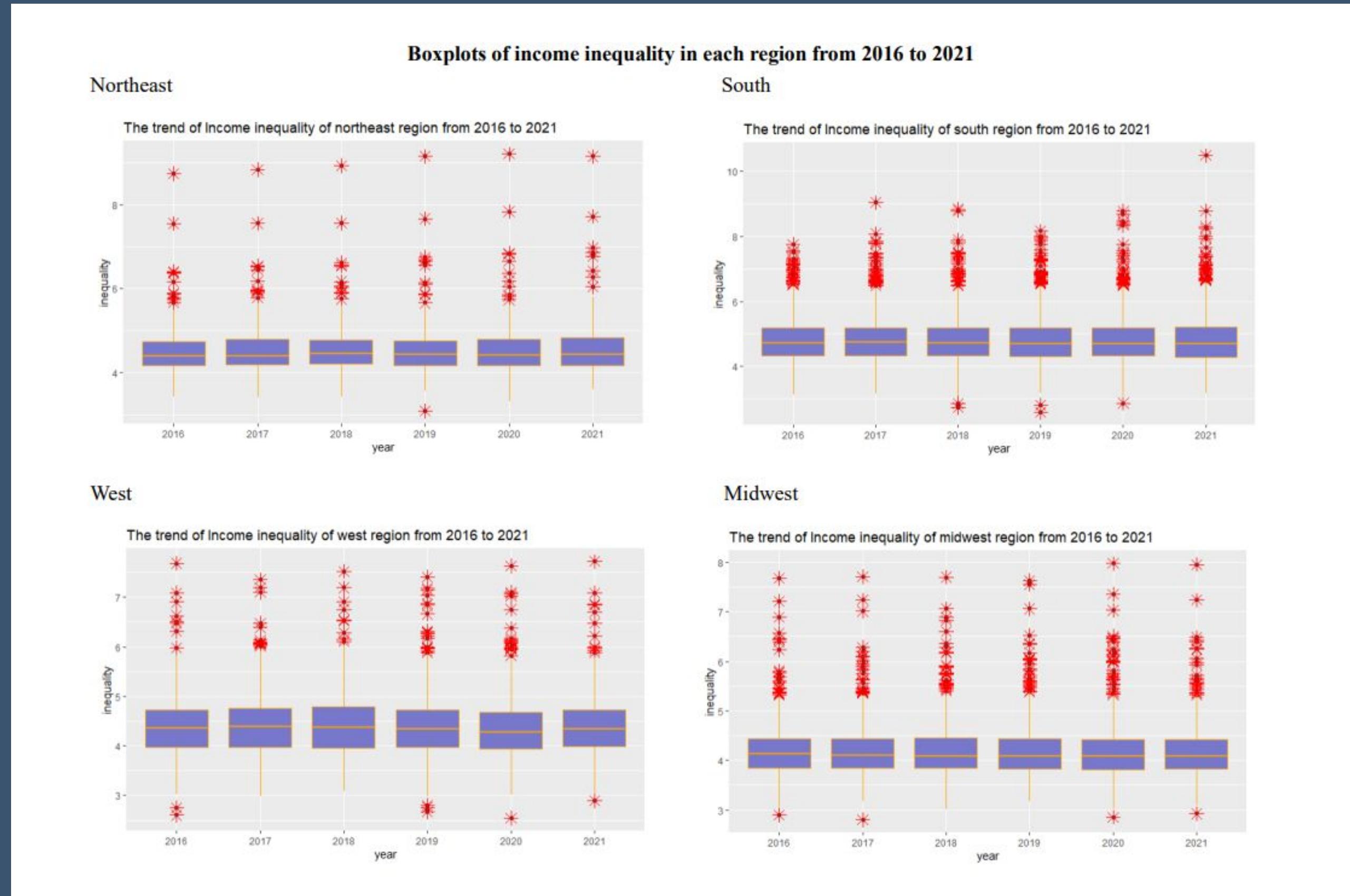
# Boxplots: Mentally Unhealthy Days by Region per Year



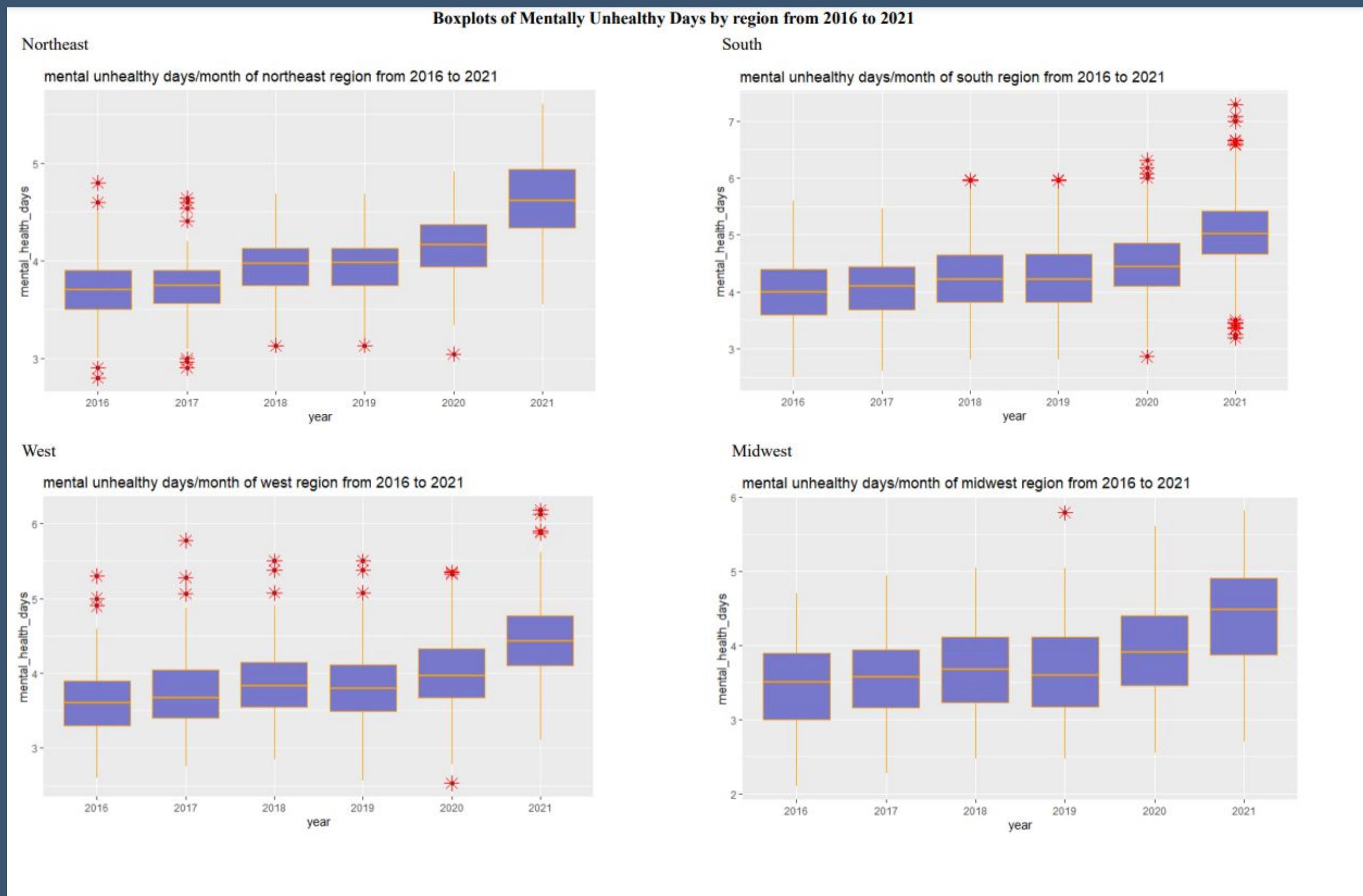
# Boxplots: Median Income per Year in each Region



# Boxplots: Inequality per Year in each Region



# Boxplots: Mentally Unhealthy Days per year in each Region



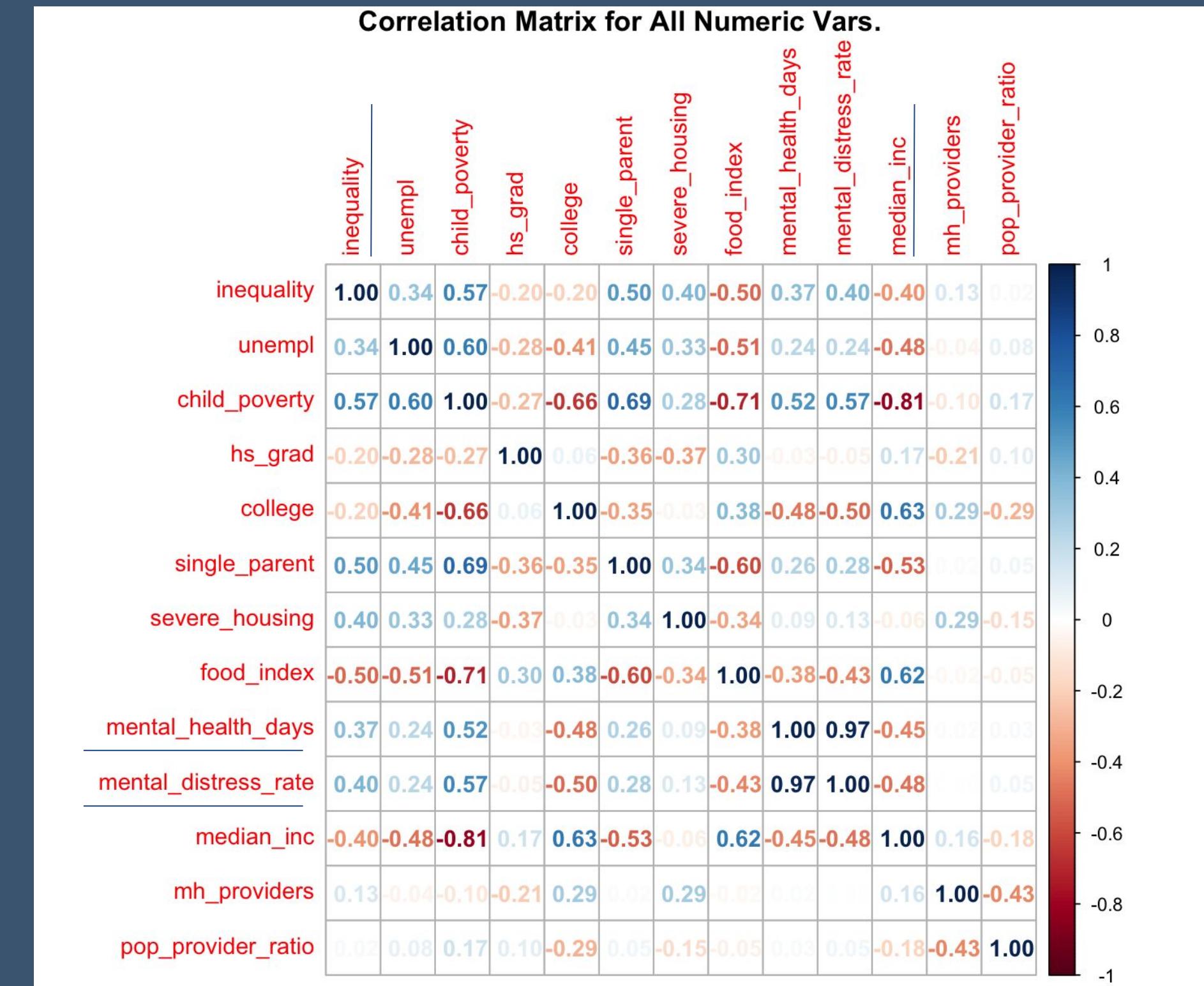
# Hypothesis Testing

---



# Correlation Matrix: All Numeric Variables

- Correlation coefficient between *mental\_health\_days* and *inequality* is 0.37.
- Correlation coefficient between *mental\_distress\_rate* and *inequality* is 0.40
- Correlation coefficient between *mental\_health\_days* and *median\_inc* is -0.45
- Correlation coefficient between *mental\_distress\_rate* and *median\_inc* is -0.48



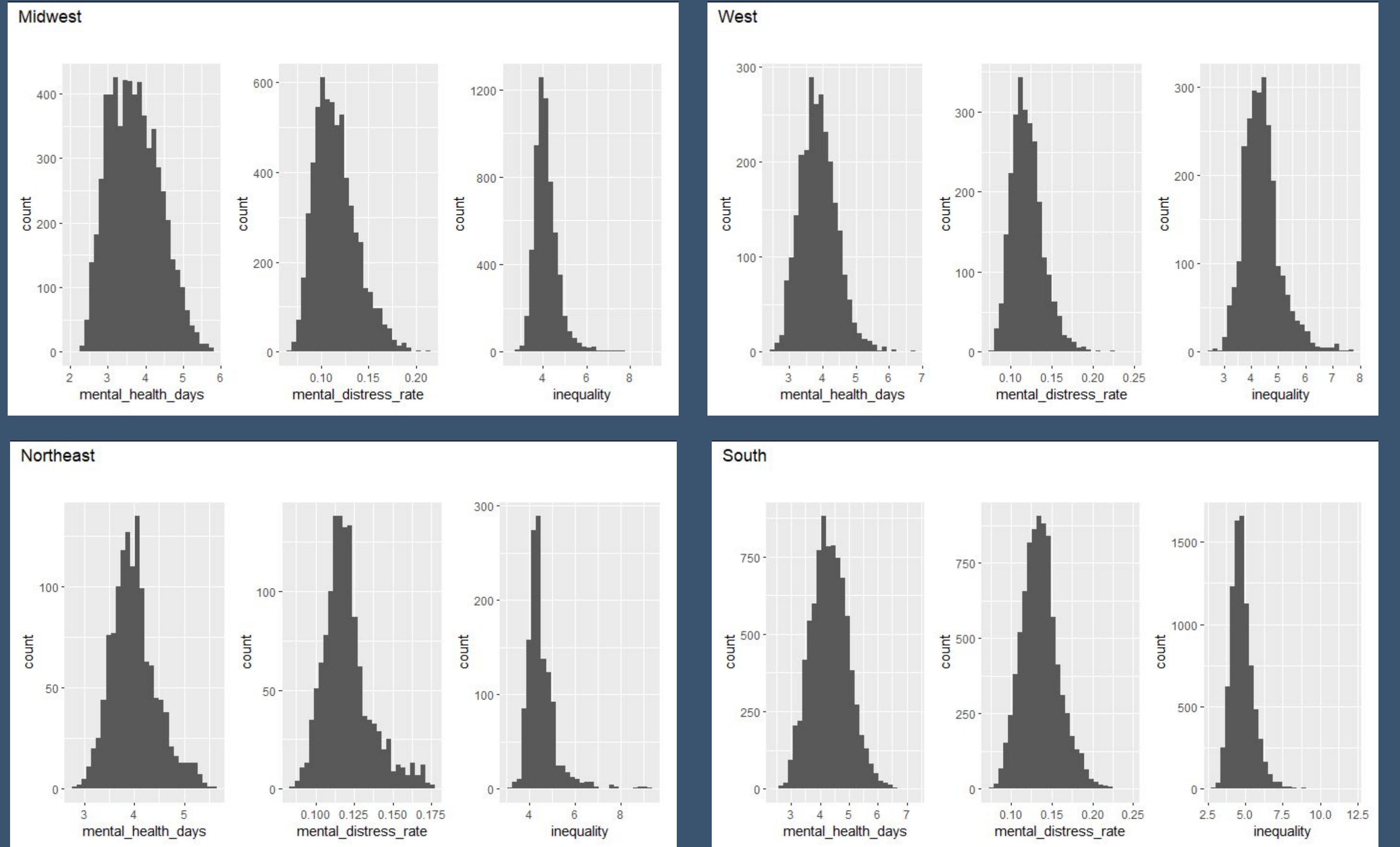
# t-interval Estimates

t-interval Estimates for Key Variables

Region	Variable	99% Conf. Int.
All	inequality	[4.5097, 4.5375]
All	mental_health_days	[4.0266, 4.0529]
All	mental_distress_rate	[0.1256, 0.1265]
Midwest	inequality	[4.1641, 4.2011]
Midwest	mental_health_days	[3.686, 3.7301]
Midwest	mental_distress_rate	[0.115, 0.1165]
Northeast	inequality	[4.4926, 4.5863]
Northeast	mental_health_days	[3.9758, 4.041]
Northeast	mental_distress_rate	[0.1193, 0.1215]
South	inequality	[4.7855, 4.8277]
South	mental_health_days	[4.315, 4.3513]
South	mental_distress_rate	[0.1355, 0.1368]
West	inequality	[4.3659, 4.4358]
West	mental_health_days	[3.8556, 3.9113]
West	mental_distress_rate	[0.1196, 0.1216]



# Regional Histogram Plots



# Regional Variance Comparison

- Comparison suggests homogeneity of variance within each group.

Regional Variance Comparison	
Variable	Variance
MW - mental_health_days	0.465
MW - mental_distress_rate	0.000501
MW - inequality	0.318
WE - mental_health_days	0.296
WE - mental_distress_rate	0.000382
WE - inequality	0.465
NE - mental_health_days	0.208
NE - mental_distress_rate	0.000244
NE - inequality	0.43
SO - mental_health_days	0.42
SO - mental_distress_rate	0.00052
SO - inequality	0.567



# ANOVA Test for Inequality

- The null hypothesis is rejected because the p-value is below a significance level of 0.05, and accept the alternative hypothesis that the means of the regions are not all the same.
- We followed up with a post-hoc test to determine which groups are significantly different from the others.
- The Tukey HSD results indicate that all regions are significantly different from each other for inequality.

ANOVA Results: Inequality by Region					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Region	3	1434	477.94	1039	0
Residuals	18461	8492	0.46	NA	NA

Tukey Multiple Comparison of Means 95% family-wise confidence level				
Fit: aov(formula = inequality ~ region, data = ranked)				
Region	diff	lower	upper	p adj
Northeast-Midwest	0.357	0.304	0.4100	0
South-Midwest	0.624	0.595	0.6532	0
West-Midwest	0.218	0.177	0.2594	0
South-Northeast	0.267	0.215	0.3190	0
West-Northeast	-0.139	-0.198	-0.0791	0
West-South	-0.406	-0.445	-0.3662	0



# ANOVA Test for Poor Mental Health Days

- The null hypothesis is rejected because the p-value is below a significance level of 0.05, and accept the alternative hypothesis that the means of the regions are not all the same.
- A post-hoc test determined which groups are significantly different from the others.
- The Tukey HSD results indicate that all regions are significantly different from each other for poor mental health days.

ANOVA Results: Poor Mental Health Days by Region					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Region	3	1471	490.347	1230	0
Residuals	18461	7363	0.399	NA	NA

Tukey Multiple Comparison of Means 95% family-wise confidence level				
Fit: aov(formula = mental_health_days ~ region, data = ranked)				
Region	diff	lower	upper	p adj
Northeast-Midwest	0.300	0.251	0.3499	0
South-Midwest	0.625	0.598	0.6532	0
West-Midwest	0.175	0.137	0.2137	0
South-Northeast	0.325	0.276	0.3731	0
West-Northeast	-0.125	-0.180	-0.0696	0
West-South	-0.450	-0.486	-0.4120	0



# ANOVA Test for Frequent Mental Distress

- The null hypothesis is rejected because the p-value is below a significance level of 0.05, and accept the alternative hypothesis that the means of the regions are not all the same.
- A post-hoc test determined which groups are significantly different from the others.
- The Tukey HSD results indicate that, with the exception of West-Northeast, all regions are significantly different from each other for mental distress rate.

ANOVA Results: Frequent Mental Distress by Region					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Region	3	1.64	0.5471	1151	0
Residuals	18465	8.78	0.399	NA	NA

Tukey Multiple Comparison of Means 95% family-wise confidence level				
Fit: aov(formula = mental_distress_rate ~ region, data = ranked)				
Region	diff	lower	upper	p adj
Northeast-Midwest	0.004660	0.00295	0.00637	0
South-Midwest	0.020434	0.01950	0.02137	0
West-Midwest	0.004824	0.00350	0.00615	0
South-Northeast	0.015774	0.01411	0.01744	0
West-Northeast	0.000164	-0.00175	0.00207	0.996
West-South	-0.015610	-0.01688	-0.01434	0



# Spearman's Correlation Tests

## Spearman's rank order correlation

Data: median\_inc | mental\_health\_days

S	rho	p-value
2e+11	-0.45	<2e-16

## Spearman's rank order correlation

Data: median\_inc | mental\_distress\_rate

S	rho	p-value
6e+11	-0.457	<2e-16

## Spearman's rank order correlation

Data: inequality | mental\_health\_days

S	rho	p-value
6e+11	0.399	<2e-16

## Spearman's rank order correlation

Data: inequality | mental\_distress\_rate

S	rho	p-value
6e+11	0.430	<2e-16

- Spearman's Correlation tests indicate that the null hypothesis can be rejected and that there is a decreasing monotonic relationship between *median\_inc* and *mental\_health\_days* and *mental\_distress\_rate*.
- Spearman's Correlation tests indicate that the null hypothesis can be rejected and that there is a increasing monotonic relationship between *inequality* and *mental\_health\_days* and *mental\_distress\_rate*.

# Conclusion

---



# Conclusion

---



- ANOVA tests indicate that means between regions are significantly different.
- Spearman's Correlation tests indicate that there are significant relationships between mental health and economic inequality.
- Initial scatterplots indicate that there is a positive relationship between inequality and poor mental health death days and mental distress.
- Boxplots indicate that there are differences in poor mental health death days and mental distress across regions.
- The correlation matrix indicates that a weak to moderate positive correlation between poor mental health death days and mental distress and inequality. The correlation matrix also indicates a slightly weak negative correlation between poor mental health death days and mental distress and median income.
- Overall, the following data analysis supports the notion that there is a relationship between economic inequality and mental health and that it differs geographically in the United States..

# References

Bechtel, Lucy, Grace Lordan, and DS Prasada Rao. "Income inequality and mental health—empirical evidence from Australia." *Health economics* 21 (2012): 4-17.

Kelley, Jonathan, and Mariah DR Evans. "Societal Inequality and individual subjective well-being: Results from 68 societies and over 200,000 individuals, 1981–2008." *Social science research* 62 (2017): 1-23.

Layte, Richard. "The association between income inequality and mental health: testing status anxiety, social capital, and neo-materialist explanations." *European Sociological Review* 28.4 (2012): 498-511.

Matthew, Pravin, and Donka Mirtcheva Brodersen. "Income inequality and health outcomes in the United States: An empirical analysis." *The Social Science Journal* 55.4 (2018): 432-442.

Pickett, Kate E., and Richard G. Wilkinson. "Income inequality and health: a causal review." *Social science & medicine* 128 (2015): 316-326.

Ribeiro, Wagner Silva, et al. "Income inequality and mental illness-related morbidity and resilience: a systematic review and meta-analysis." *The Lancet Psychiatry* 4.7 (2017): 554-562.

Robert Wood Johnson Foundation. "2021 County Health Rankings National Data." *County Health Rankings & Roadmaps*. <https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation>. Accessed: March 13, 2022.

Sommet, Nicolas, Davide Morselli, and Dario Spini. "Income inequality affects the psychological health of only the people facing scarcity." *Psychological Science* 29.12 (2018): 1911-1921.

Tibber, Marc S., et al. "The association between income inequality and adult mental health at the subnational level—a systematic review." *Social psychiatry and psychiatric epidemiology* (2021): 1-24.

U.S. Census Bureau. "Census Regions and Divisions of the United States." *Economics and Statistics Administration*.

Zimmerman, Frederick J., and Janice F. Bell. "Income inequality and physical and mental health: testing associations consistent with proposed causal pathways." *Journal of Epidemiology & Community Health* 60.6 (2006): 513-521.