

Homework 3

Formal report

Logan | QMB 3200 | 9.29.2020

Table of Contents

[List of Figures and Tables 2](#_Toc52314809)

[Summary Statistics 2](#_Toc52314810)

[Box plots 2](#_Toc52314811)

[Histograms 3](#_Toc52314812)

[Weighted summary statistics 3](#_Toc52314813)

[Correlation Matrix 4](#_Toc52314814)

[Scatter plots 4](#_Toc52314815)

[Executive Summary 6](#_Toc52314816)

[Heading 2 6](#_Toc52314817)

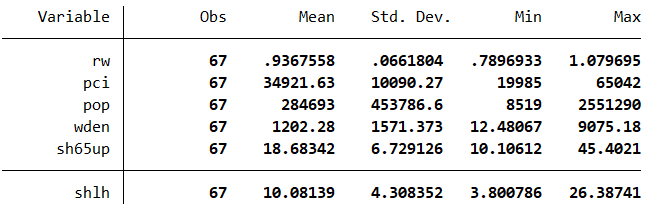
[Conclusion 7](#_Toc52314818)

[Heading 2 7](#_Toc52314819)

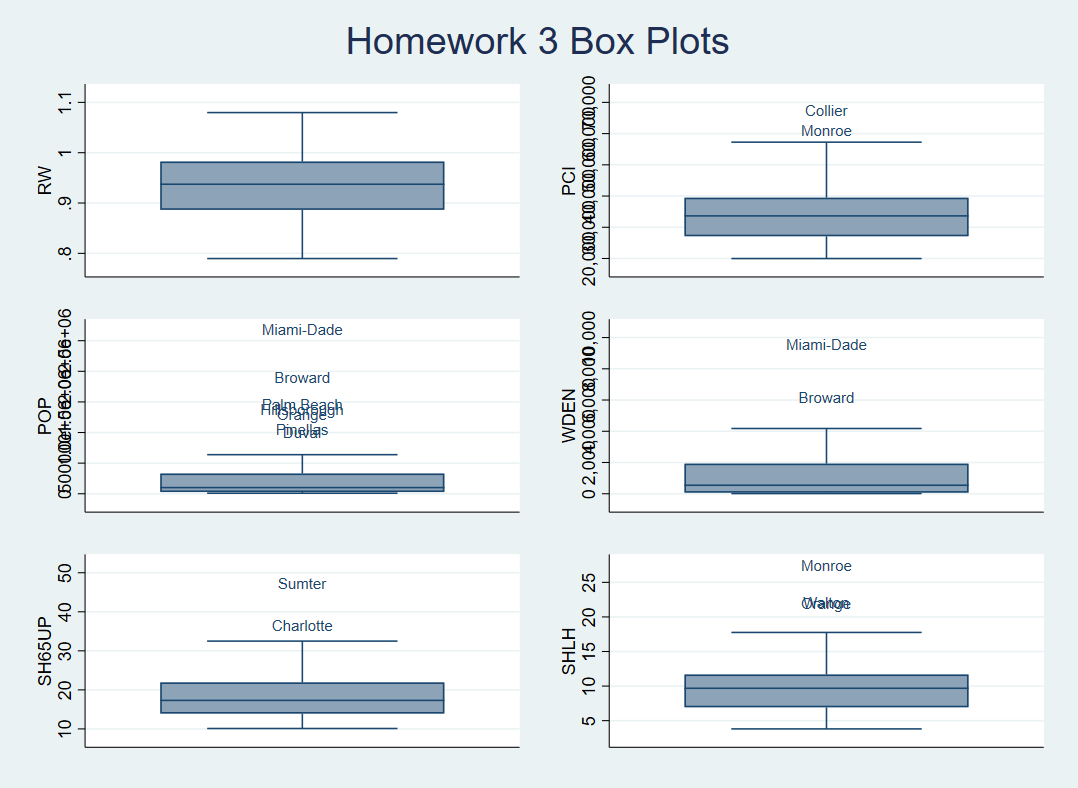
[Appendix 8](#_Toc52314820)

# List of Figures and Tables

## Summary Statistics

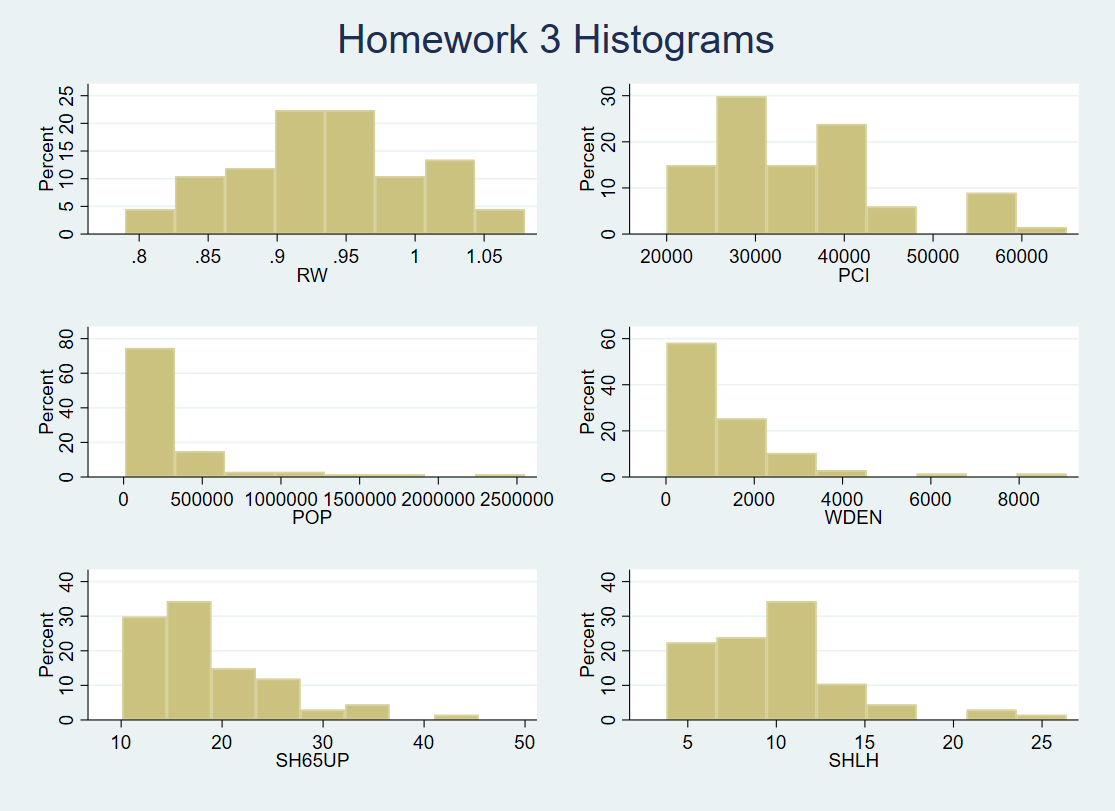
Summary Statistics Table Generated in STATA from given variables

## Box plots

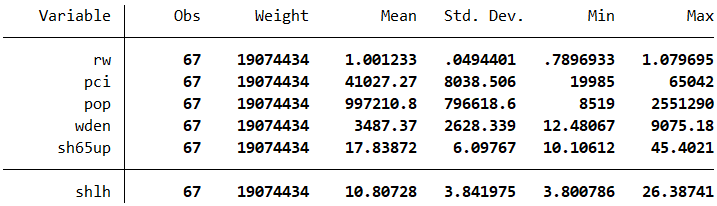


Box plots generated in STATA from given data

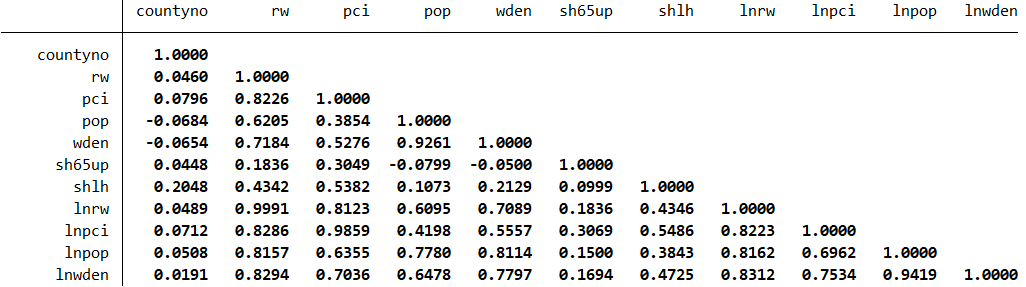
## Histograms

 Histograms generated in STATA from given data

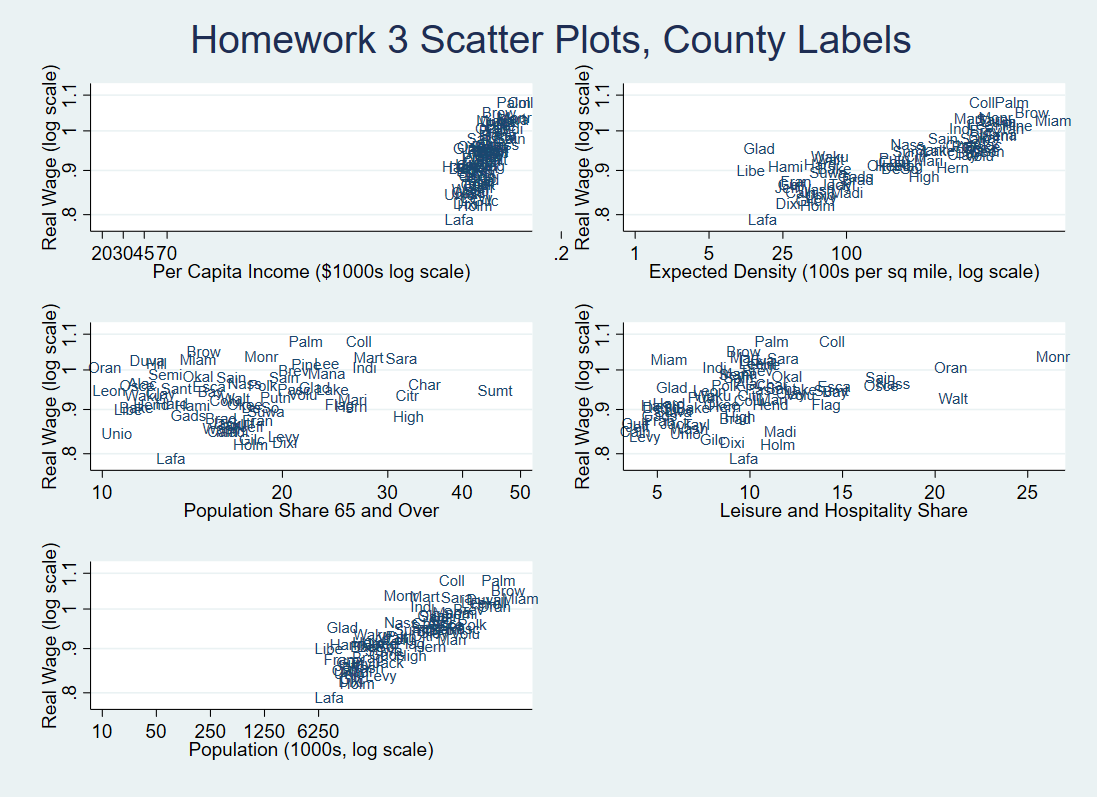
## Weighted summary statistics

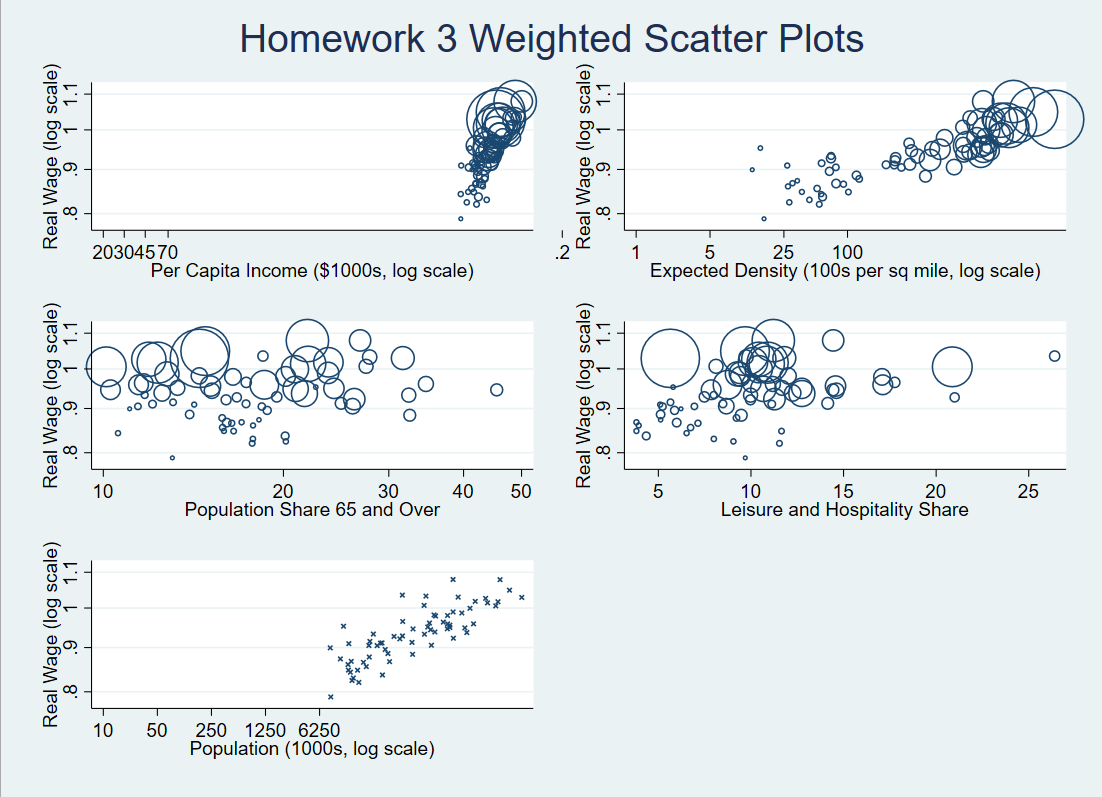
 Weighted summary statistics in STATA from given data

## Correlation Matrix

 Correlation data generated in STATA

## Scatter plots

 Scatter plots w/ county labels generated in STATA

 Weighted scatter plots generated in STATA

# Executive Summary

When looking at the summary statistics, it’s easy to recognize a trend across the entire state. There is a high variation in the spread in the variables across the state. There are very highly populated counties to smaller, sparsely populated counties. There are very high income areas to low income areas and so forth by age, tourism-funded, and rural to dense counties. Florida is different than other states that are primarily spread out or primarily higher income, and/or younger. The same information is shown graphically in Figures 1 and 2. Many of these variables are right skewed. This makes sense in two ways, first, all these variables take on only positive values. Second, many of the variables make sense in multiplicative terms. Miami-Dade and Broward (the southeastern most counties) are the most populous and dense. Monroe, the Florida Keys, has the highest Leisure and Hospitality share, reflecting high levels of tourism relative to other activities. Collier, home to Naples, is highest income. Sumter, home to the Villages senior community, has the highest retirement age share—though the Villages exists largely as its own community.

# Conclusion

In conclusion, based purely on city alone, it is easy to determine what the typical citizen could be. Florida is very diverse but specific people with reside in specific areas. Older people probably will not want to live in Miami-Dade county but rather, given they have a higher income, will live in the Florida Keys.

# Appendix

cd "C:\Users\lmm56\Documents\School\Poly 20 - 21\QMB 3200\Homework Submissions\Assignment 3"

log using "homework 3", replace

import delimited "Assignment 3 Data.csv"

summ rw pci pop wden sh65up shlh

\*\* make box plots

graph box rw, m(1, ms(i) mlabel(county) mlabpos(0)) saving("rw box", replace)

graph box pci, m(1, ms(i) mlabel(county) mlabpos(0)) saving("pci box", replace)

graph box pop, m(1, ms(i) mlabel(county) mlabpos(0)) saving("pop box", replace)

graph box wden, m(1, ms(i) mlabel(county) mlabpos(0)) saving("wden box", replace)

graph box sh65up, m(1, ms(i) mlabel(county) mlabpos(0)) saving("sh65up box", replace)

graph box shlh, m(1, ms(i) mlabel(county) mlabpos(0)) saving("shlh box", replace)

\*\*combine graphs to make them look better

graph combine "rw box.gph" "pci box.gph" "pop box.gph" "wden box.gph" ///

"sh65up box.gph" "shlh box.gph" , col(2) title("Homework 3 Box Plots") saving("box plots", replace)

\*\*make histograms

hist rw, percent saving("rw hist", replace)

hist pci, percent saving("pci hist", replace)

hist pop, percent saving("pop hist", replace)

hist wden, percent saving("wden hist", replace)

hist sh65up, percent saving("sh65up hist", replace)

hist shlh, percent saving("shlh hist", replace)

\*\*combine graphs to make them look better

graph combine "rw hist.gph" "pci hist.gph" "pop hist.gph" "wden hist.gph" ///

"sh65up hist.gph" "shlh hist.gph", col(2) title("Homework 3 Histograms") saving("histograms", replace)

\*\*create summary statistics with weighted averages for the population

summ rw pci pop wden sh65up shlh [aw=pop]

\*\*new natural log variables from existing variables

gen lnrw=ln(rw)

gen lnpci=ln(pci)

gen lnpop=ln(pop)

gen lnwden=ln(wden)

correlate

corr lnrw lnpci lnpop lnwden sh65up shlh

\*\*create scatter plots with hollow circles for markets using ms(oh)

\*\* with circle size proportional to population using aw=pop

scatter rw pci [aw=pop], ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Per Capita Income ($1000s, log scale)") xscale(log) xlab(20 30 45 70) ///

ms(oh) saving("rw v pci wght", replace)

scatter rw pop, ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Population (1000s, log scale)") xscale(log) xlab(10 50 250 1250 6250) ///

ms(x) saving("rw v pop", replace)

scatter rw wden [aw=pop], ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Expected Density (100s per sq mile, log scale)") xscale(log) xlab(.2 1 5 25 100) ///

ms(oh) saving("rw v wden wght", replace)

scatter rw sh65up [aw=pop], ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Population Share 65 and Over") xscale(log) ms(oh) saving("rw v sh65up wght", replace)

scatter rw shlh [aw=pop], ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Leisure and Hospitality Share") ms(oh) saving("rw v shlh wght", replace)

graph combine "rw v pci wght.gph" "rw v wden wght.gph" "rw v sh65up wght.gph" ///

"rw v shlh wght.gph" "rw v pop.gph" , col(2) row(3) title("Homework 3 Weighted Scatter Plots") saving("wght scatter graphs.gph", replace)

\*\*create a shorter label than the entire county name for data points

gen coab=substr(county,1,4)

\*\*create scatter plots with no marker use a label for the county name

\*\*using the 1st four letters using mlab(coab) centered on the data point using mlabpos(0)

scatter rw pci, ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Per Capita Income ($1000s log scale)") xscale(log) xlab(20 30 45 70) ///

ms(i) mlab(coab) mlabpos(0) saving("rw v pci lab", replace)

scatter rw pop, ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Population (1000s, log scale)") xscale(log) xlab(10 50 250 1250 6250) ///

ms(i) mlab(coab) mlabpos(0) saving("rw v pop lab", replace)

scatter rw wden, ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Expected Density (100s per sq mile, log scale)") xscale(log) xlab(.2 1 5 25 100) ///

ms(i) mlab(coab) mlabpos(0) saving("rw v wden lab", replace)

scatter rw sh65up, ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Population Share 65 and Over") xscale(log) ms(i) mlab(coab) mlabpos(0) saving("rw v sh65up lab", replace)

scatter rw shlh, ytitle("Real Wage (log scale)") yscale(log) ///

xtitle("Leisure and Hospitality Share") ms(i) mlab(coab) mlabpos(0) saving("rw v shlh lab.gph", replace)

graph combine "rw v pci lab.gph" "rw v wden lab.gph" "rw v sh65up lab.gph" ///

"rw v shlh lab.gph" "rw v pop lab.gph" , col(2) row(3) title("Homework 3 Scatter Plots, County Labels") saving("scatter plots with county labels.gph", replace)

log close

clear