## Introduction

Urbanization is the process by which a population begins to gradually shift in residency from rural to metropolitan areas. The United Nations has concluded that 55% of the world's population currently inhabits urban areas, and this proportion is expected to explode to 68% of the world's population by 2050. Sudden influxes of labor in urban areas have profound impacts on the local level economies, from residency prices to wage growth these effects are unilaterally experienced by all Americans.
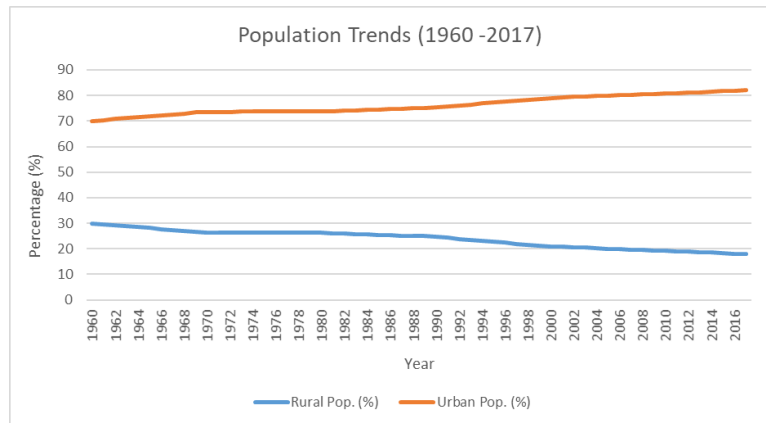
When asked, those who choose to immigrate to urban areas tend to cite job opportunities, better housing options, among other subjective incentives. However, is this truly the case? In this report the aggregate economic trends of urbanization in the United States will be explored and inferred from.
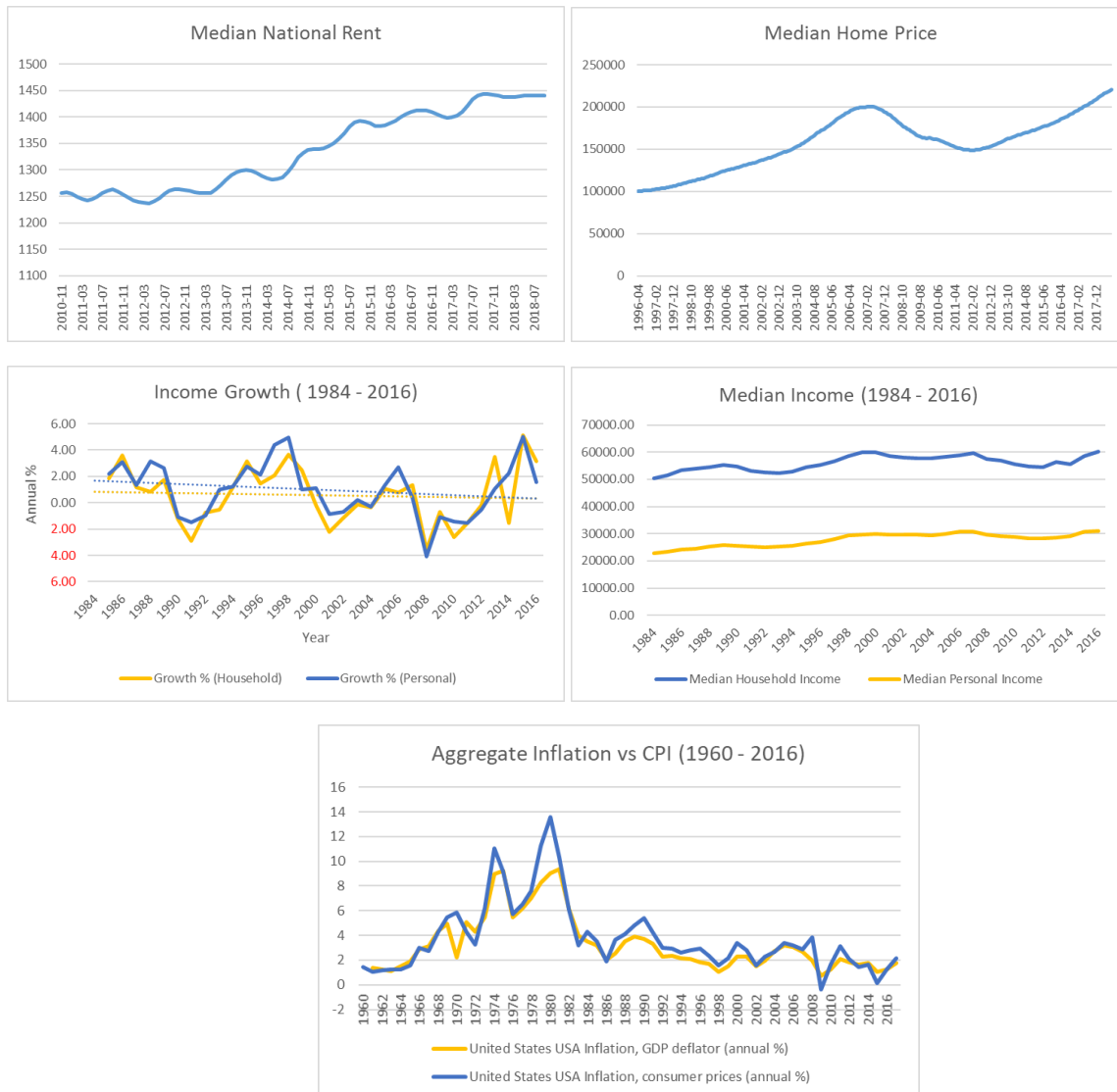
Data

The dataset will include the following variable information:

- ➢ Urban and Rural Population Proportions (1960 – 2017)
- ➢ Median Rental Rate of Top 25 Major Cities (2011 – 2018)
- ➢ GDP Deflator Inflation Rate (1947 – 2017)
- ➢ Consumer Price Index Rate (1947 -2017)
- ➢ Median Home Price Sold (1963 – 2017)
- ➢ Median USA Rental Price (1984 – 2016)
- ➢ Median Annual Household Income (1984 – 2016)
- ➢ Median Annual Personal Income (1984 – 2016)

Quick Facts

- ➢ Since 1984 the Median Personal Income has grown on average by ~1% annually
- ➢ Since 1984 the Median Household Income has grown on average by ~1% annually
- ➢ Since 1964 the average annual Median Home Price growth is ~5.60%
- ➢ Since 1984 the average annual Median Rental Rate growth is ~3.31%
- ➢ Since 1947 the aggregate inflation rate has been ~3.52%, but ~2.00% since 2000
- ➢ Since 1947 the CPI inflation rate has been ~3.18%, but ~2.00% since 2000

What does all this mean? At an average annual inflation rate of ~3.00% prices will double every 20 years. However, at the current income growth it would take at least triple this to time to double. So, from the data above price growth in general have outpaced income gains, but more specifically the standard American home has become more than unattainable to the lower 50% of the income distribution

Median National Rent

Median Home Price

Income Growth (1984 - 2016)

Median Income (1984 - 2016)

Aggregate Inflation vs CPI (1960 - 2016)

Methods

The Python Plotly package will be used to develop aggregate trend plots among the data listed above.

The t-test will be compared to examine the difference in sample means between two given samples of the same variable. The difference of pre and post 2008 Recession inflation rates will be looked at. Two separate t-tests will be utilized:

1. The common independent samples t-test which compares the mean for two groups
2. Paired sample t-test which compares means from the same group at differing times

Pearson correlation coefficients are also calculated among the various rates of change calculated for that variables listed above. The coefficients range from -1 to 1, where a value of 1 implies a perfect positive relationship between X and Y, and vise versa for a correlation value of -1. A value of 0 indicates no relationship between the variables.

Lastly, five simple linear regression models were coded to test the predictive nature of the growth in median incomes, inflation metrics, and the growth in renting as an alternative, in predicting the annual median home price sold in the USA. From these separate models, a single multivariant model will be demonstrated and the effectiveness or lack there of is addressed.

Inference

T-tests:

As mentioned earlier both a two-sample t-test and a paired t-test were used. For both we will be testing the null hypothesis that no relationship between the aggregate inflation and the consumer price index inflation, against the alternative that there is some relationship either positive or negative. If we examine the p-values of each of the standard t-tests below we infer that there is no statistically significant relationship between aggregate inflation and the consumer price index inflation pre-recession (2007/2008). However, from 2008 – 2017 we have a calculated p-value less than 0.05, from which we can infer that during this year range there was an overarching relationship between both types of inflation tested.

A paired t-test were used to factor in time elapsed through the year variable. Since we are concerned with the difference between two measures in a sample, we use the paired t-test to reduce to a one-sample t-test. From our calculated output we infer the surprising and reciprocal inference than above. On the contrary from the two-sample t-test, we infer that the p-value of the 1947-2017 range is less than 0.05 and therefore points to a relationship among inflation values. However, the inflation values from 2008-2017 we infer no relationship as the p-value is much greater than 0.05.

Pearson Correlation Coefficients:

Our dataset contained growth rates of our variables from the data section. From this statistical analysis we notice a few predictable results, for one, median household and personal income have a high positive correlation of, r = 0.82412. One would think that this relationship would be closer to 1, so what would explain this variation? Self-employment growth over time could explain this. Median personal and household income growth are positively correlated with median home price. Lastly, the inflation types are highly correlated, which seems to conflict with the t-test results and will be discussed in this section.

Simple Linear Regressions:

Our goal was to examine the variable relationships of our variables in the prediction on median home prices. From our models we conclude that both median household and personal income growth are statistically significant predictive variables, as their p-values are less than 0.05. However, and to my surprise the median monthly rental rate appears to have no predictive nature. As expected, the growth in consumer price index has no correlation with home prices, as it measures price growth in common household items such as basic groceries and energy costs. Lastly, aggregate inflation is not statistically significant.

Results

```
/* t-test for inflation rates from 1947 to 2007 */
```
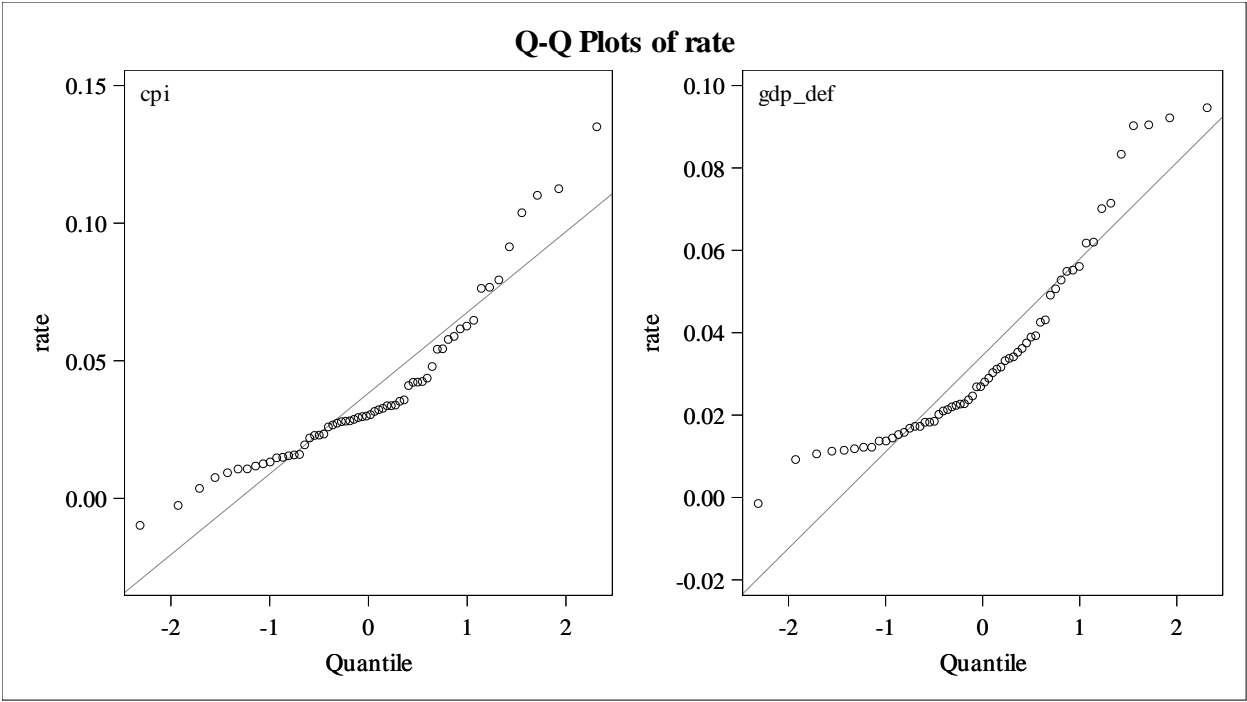
metric	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
cpi		60	0.0382	0.0293	0.00379	-0.00981	0.1350
gdp_def		60	0.0345	0.0234	0.00302	-0.00147	0.0946
Diff (1-2)	Pooled		0.00374	0.0265	0.00485		
Diff (1-2)	Satterthwaite		0.00374		0.00485		

metric	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
cpi		0.0382	0.0307	0.0458	0.0293	0.0249	0.0358
gdp_def		0.0345	0.0285	0.0405	0.0234	0.0198	0.0286
Diff (1-2)	Pooled	0.00374	-0.00586	0.0133	0.0265	0.0235	0.0304
Diff (1-2)	Satterthwaite	0.00374	-0.00587	0.0133			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	118	0.77	0.4423
Satterthwaite	Unequal	112.47	0.77	0.4423

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	59	59	1.57	0.0860

Distribution of rate

Q-Q Plots of rate
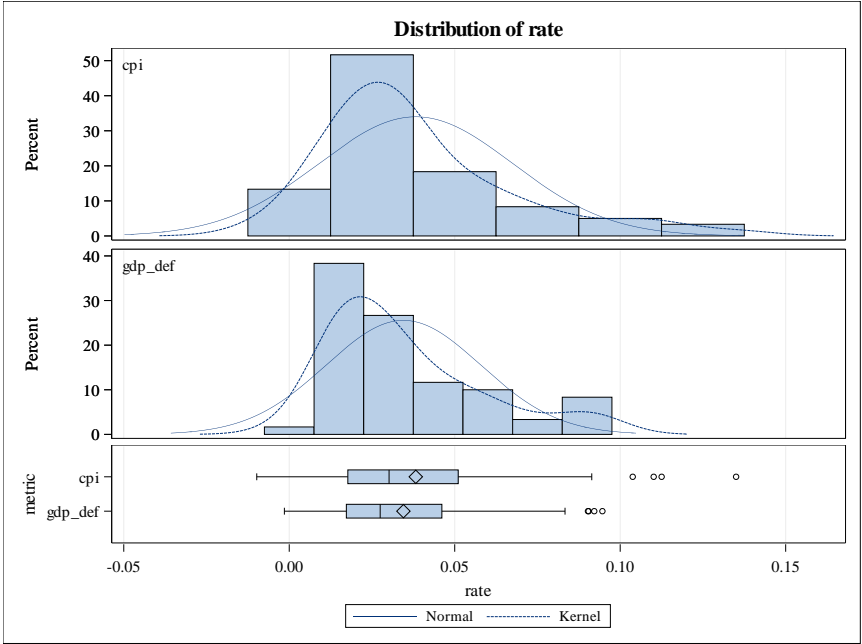

```
/* t-test for inflation rates from 2007 to 2017 */
```

metric	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
cpi		10	0.0170	0.0123	0.00390	-0.00320	0.0381
gdp_def		10	0.0156	0.00481	0.00152	0.00757	0.0209
Diff (1-2)	Pooled		0.00136	0.00936	0.00419		
Diff (1-2)	Satterthwaite		0.00136		0.00419		

metric	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
cpi		0.0170	0.00813	0.0258	0.0123	0.00849	0.0225
gdp_def		0.0156	0.0121	0.0190	0.00481	0.00331	0.00879
Diff (1-2)	Pooled	0.00136	-0.00744	0.0102	0.00936	0.00708	0.0138
Diff (1-2)	Satterthwaite	0.00136	-0.00779	0.0105			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	18	0.33	0.7485
Satterthwaite	Unequal	11.676	0.33	0.7505

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	6.57	0.0098

Distribution of rate

Q-Q Plots of rate
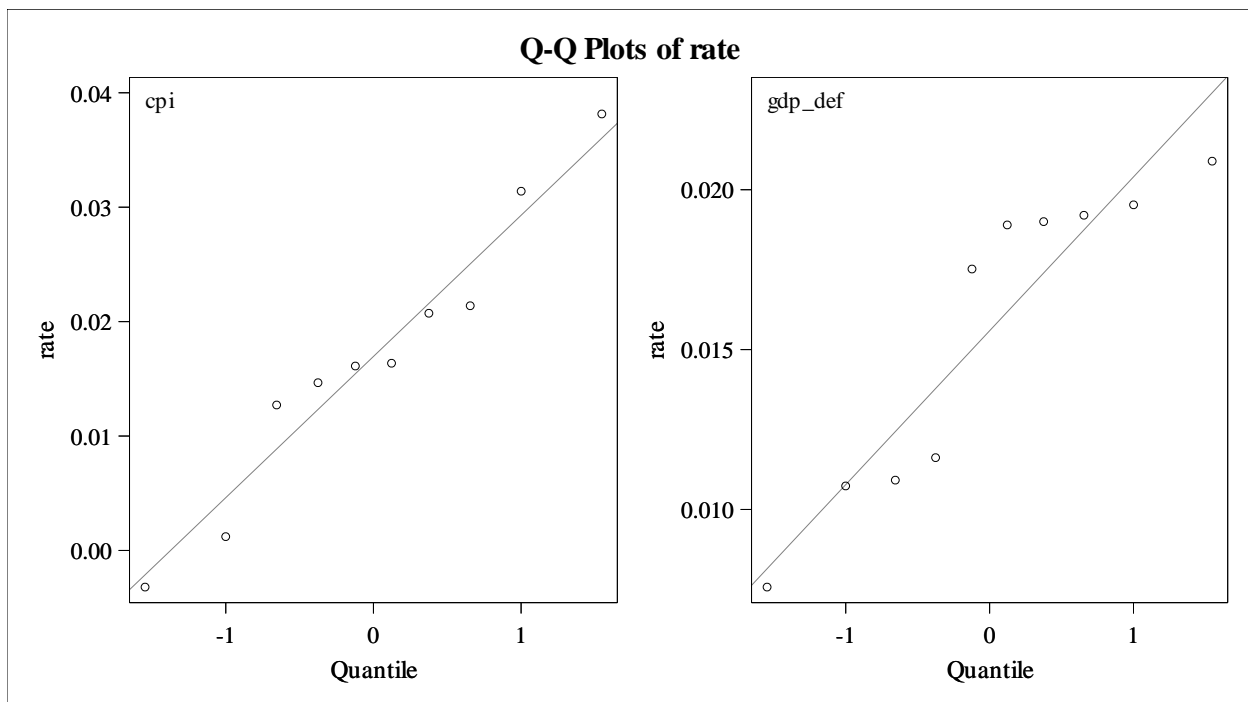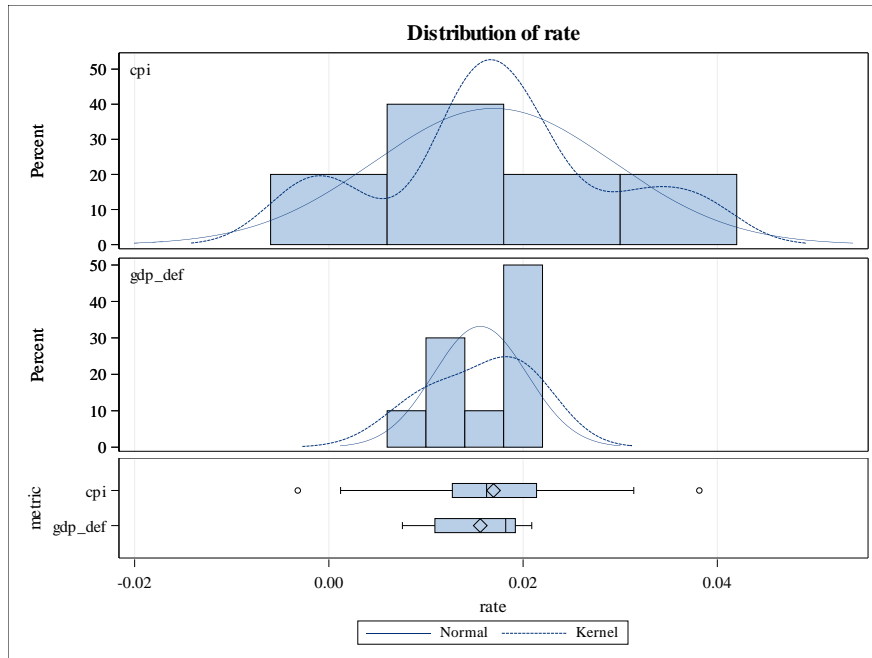
`/* paired t-test for inflation rates from 1947 to 2007 */`

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-0.00374	-0.00625	-0.00122	0.00973	0.00825	0.0119

N	Mean	Std Dev	Std Err	Minimum	Maximum
60	-0.00374	0.00973	0.00126	-0.0446	0.0194
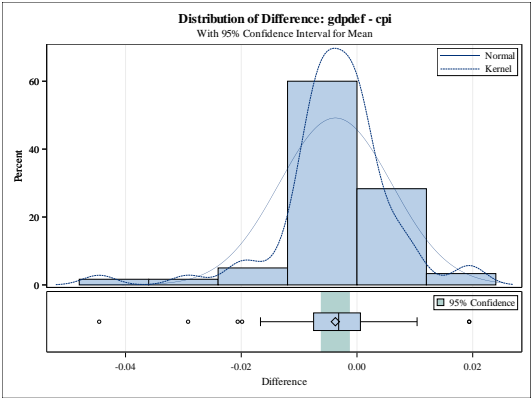
DF	t Value	Pr > |t|
59	-2.97	0.0043
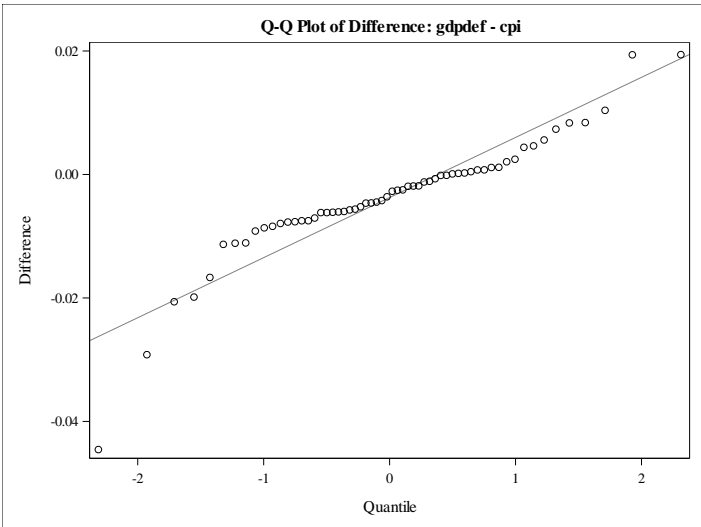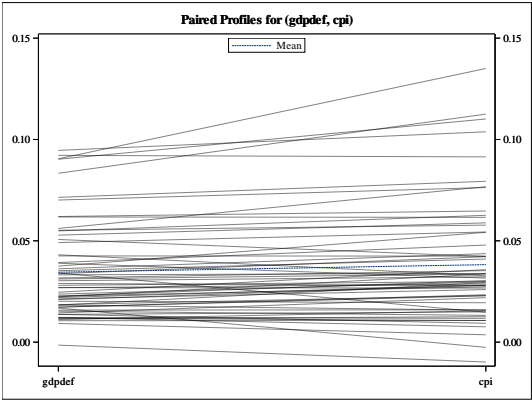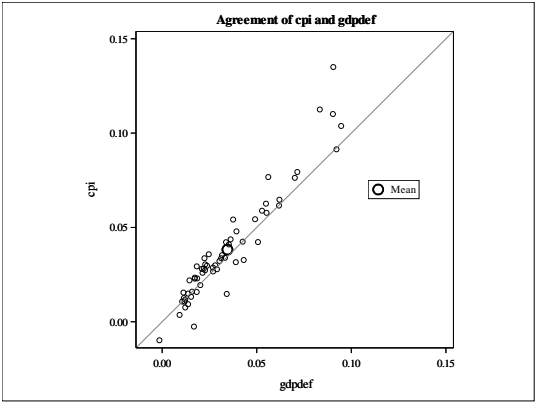
```
/* paired t-test for inflation rates from 2007 to 2017 */
```

N	Mean	Std Dev	Std Err	Minimum	Maximum
10	-0.00136	0.00879	0.00278	-0.0186	0.0108

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-0.00136	-0.00765	0.00492	0.00879	0.00604	0.0160

DF	t Value	Pr > |t|
9	-0.49	0.6353

Distribution of Difference: gdpdef - cpi
With 95% Confidence Interval for Mean

Agreement of cpi and gdpdef

Paired Profiles for (gdpdef, cpi)

Q-Q Plot of Difference: gdpdef - cpi
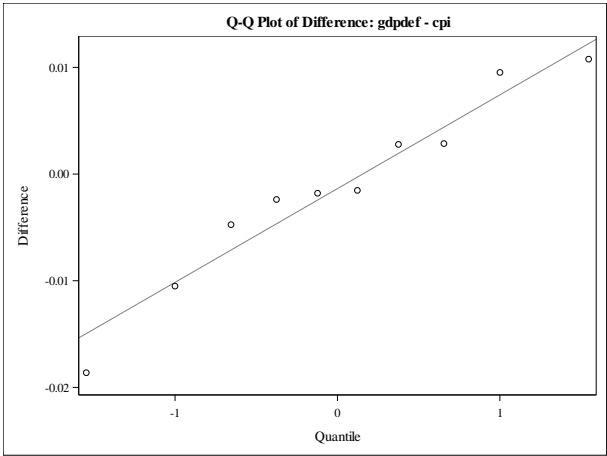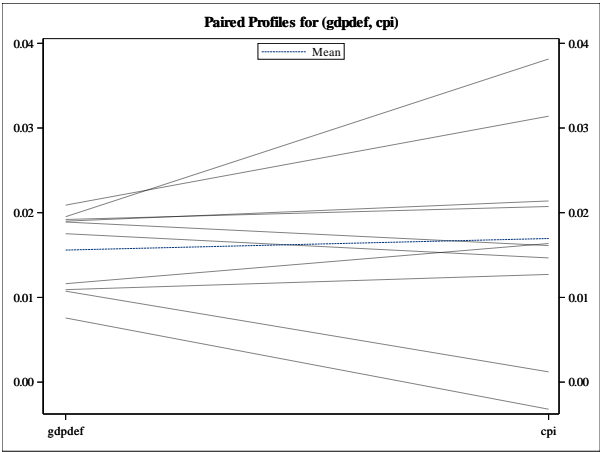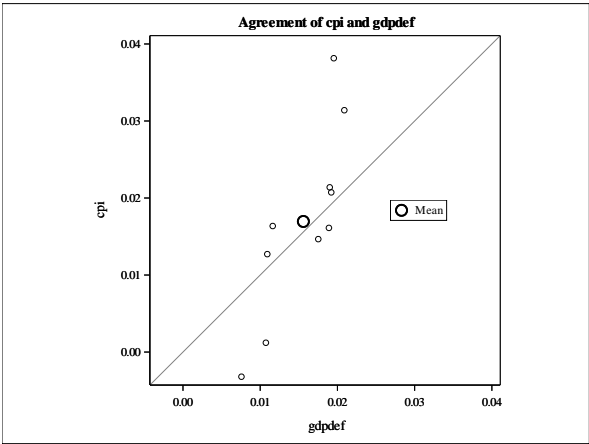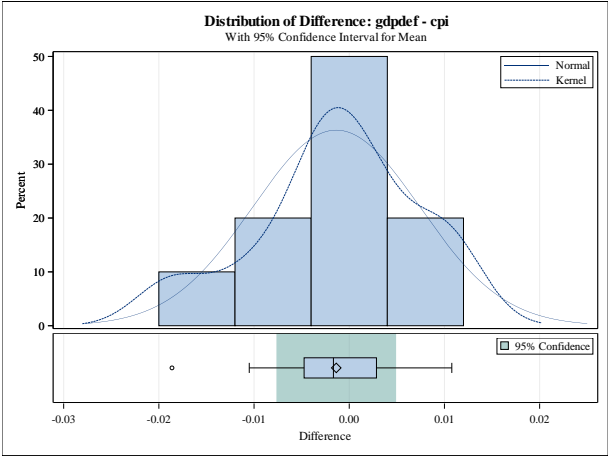
```
/* Correlations among wage growth, inflation, median income, and median */
```

7	Variables:	year	mhousehold mpersonal mhomesold mrental gdpdef cpi

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
year	32	2000	9.38083	63984	1984	2015
mhousehold	32	0.00578	0.02143	0.18485	-0.03564	0.05148
mpersonal	32	0.00978	0.02084	0.31295	-0.04104	0.05035
mhomesold	32	0.04366	0.04437	1.39728	-0.06287	0.13773
mrental	32	0.03297	0.01105	1.05511	0.00231	0.06125
gdpdef	32	0.02197	0.00810	0.70312	0.00757	0.03927
cpi	32	0.02657	0.01219	0.85020	-0.00320	0.05419

Pearson Correlation Coefficients, N = 32
### Prob > |r| under H0: Rho=0

	year	mhousehold	mpersonal	mhomesold	mrental	gdpdef	cpi
year	1.00000	-0.06896 0.7077	-0.19147 0.2938	-0.17592 0.3355	-0.43356 0.0132	-0.56695 0.0007	-0.60648 0.0002
mhousehold	-0.06896 0.7077	1.00000	0.82412 <.0001	0.35979 0.0431	0.26553 0.1419	-0.19342 0.2888	-0.32545 0.0691
mpersonal	-0.19147 0.2938	0.82412 <.0001	1.00000	0.39815 0.0240	0.24508 0.1764	-0.04169 0.8208	-0.20129 0.2693
mhomesold	-0.17592 0.3355	0.35979 0.0431	0.39815 0.0240	1.00000	0.17817 0.3293	0.17745 0.3312	0.04295 0.8155
mrental	-0.43356 0.0132	0.26553 0.1419	0.24508 0.1764	0.17817 0.3293	1.00000	0.38353 0.0302	0.27241 0.1315
gdpdef	-0.56695 0.0007	-0.19342 0.2888	-0.04169 0.8208	0.17745 0.3312	0.38353 0.0302	1.00000	0.88453 <.0001
cpi	-0.60648 0.0002	-0.32545 0.0691	-0.20129 0.2693	0.04295 0.8155	0.27241 0.1315	0.88453 <.0001	1.00000

Scatter Plot Matrix
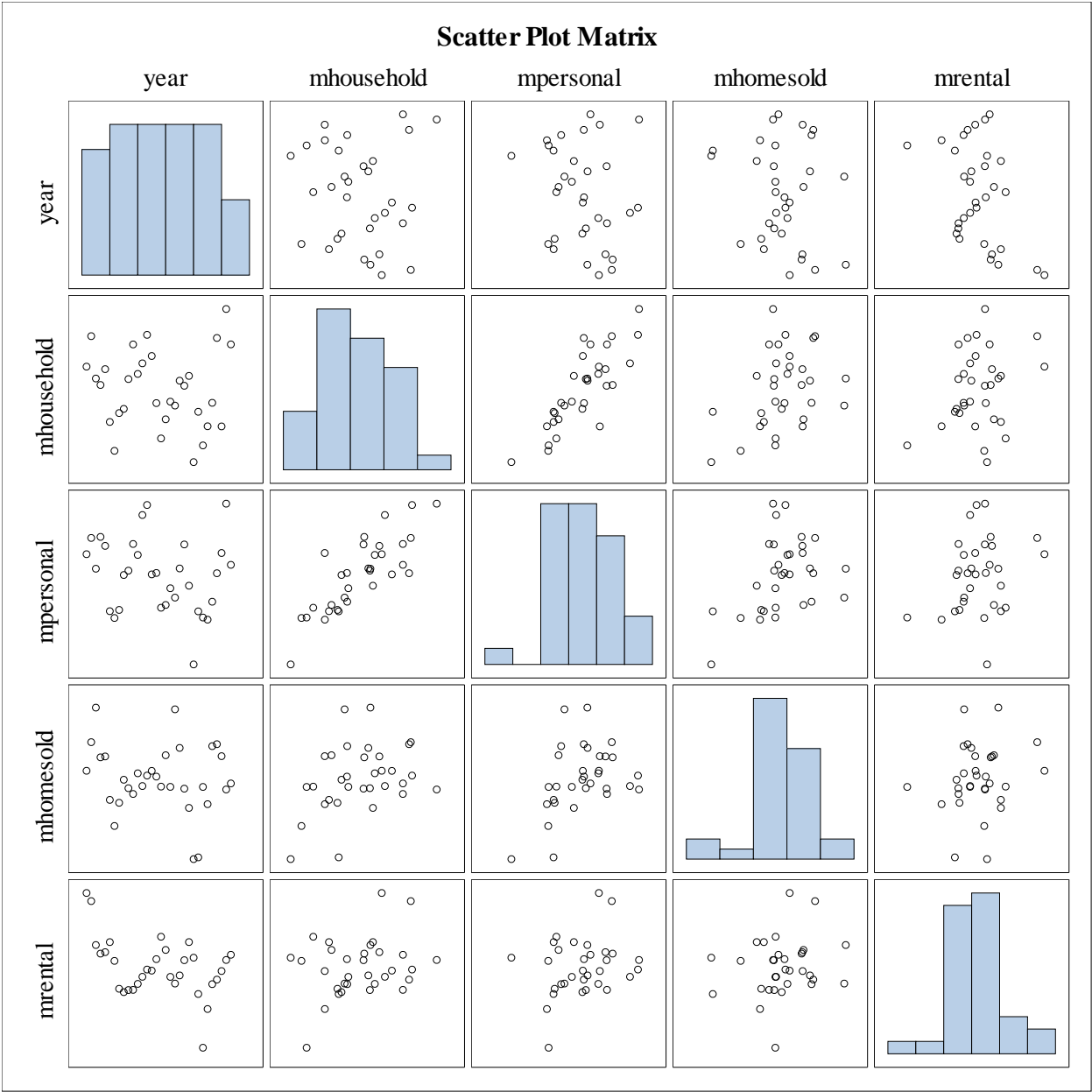
```
/* Simple Linear Regressions to predict median home price, and rent */
```

Median Home Price (Sold) = Median Household Income Growth

Number of Observations Read	32
Number of Observations Used	32

Root MSE	0.04208	R-Square	0.1294
Dependent Mean	0.04366	Adj R-Sq	0.1004
Coeff Var	96.37728		

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00790	0.00790	4.46	0.0431
Error	30	0.05313	0.00177		
Corrected Total	31	0.06103			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.03936	0.00771	5.10	<.0001
mhousehold	1	0.74503	0.35275	2.11	0.0431

Median Home Price (Sold) = Median Personal Income Growth

Number of Observations Read	32
Number of Observations Used	32

Root MSE	0.04137	R-Square	0.1585
Dependent Mean	0.04366	Adj R-Sq	0.1305
Coeff Var	94.75413		

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00967	0.00967	5.65	0.0240
Error	30	0.05136	0.00171		
Corrected Total	31	0.06103			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.03538	0.00810	4.37	0.0001
mpersonal	1	0.84761	0.35654	2.38	0.0240

Median Home Price (Sold) = Rental Rate Growth

Number of Observations Read	32
Number of Observations Used	32

Root MSE	0.04438	R-Square	0.0317
Dependent Mean	0.04366	Adj R-Sq	-0.0005
Coeff Var	101.64177		

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00194	0.00194	0.98	0.3293
Error	30	0.05909	0.00197		
Corrected Total	31	0.06103			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.02007	0.02505	0.80	0.4294
mrental	1	0.71564	0.72160	0.99	0.3293

Median Home Price (Sold) = annual CPI Growth

Number of Observations Read	32
Number of Observations Used	32

Root MSE	0.04506	R-Square	0.0018
Dependent Mean	0.04366	Adj R-Sq	-0.0314
Coeff Var	103.19919		

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00011257	0.00011257	0.06	0.8155
Error	30	0.06092	0.00203		
Corrected Total	31	0.06103			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.03951	0.01936	2.04	0.0501
cpi	1	0.15636	0.66408	0.24	0.8155

Median Home Price (Sold) = GDP Deflator Growth

Number of Observations Read	32
Number of Observations Used	32

Root MSE	0.04439	**R-Square**	0.0315
Dependent Mean	0.04366	**Adj R-Sq**	-0.0008
Coeff Var	101.65520		

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	0.00192	0.00192	0.98	0.3312
Error	30	0.05911	0.00197		
Corrected Total	31	0.06103			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.02230	0.02301	0.97	0.3402
gdpdef	1	0.97230	0.98450	0.99	0.3312

Psuedo-Code (data excluded for space)

```
/* t-test for inflation rates from 1947 to 2007 */
data pre;
      input metric $ rate @@;
      datalines;
**************;
run;
proc ttest data=pre;
      class metric;
      var rate;
      title 't-test of gdp_def vs cpi from 1947-2007';
run;

/* t-test for inflation rates from 2007 to 2017 */
data post;
      input metric $ rate @@;
      datalines;
**************;
run;
proc ttest data=post;
      class metric;
      var rate;
      title 't-test of gdp_def vs cpi from 2008-2017';

/* paired t-test for inflation rates from 1947 to 2007 */
data prepair;
      input gdpdef cpi;
      datalines;
**************;
run;
ods graphics on;

proc ttest;
```

```
        paired gdpdef*cpi;
        title 't-test (paired) of gdp_def vs cpi from 1947-2007';
run;
ods graphics off;

/* paired t-test for inflation rates from 2007 to 2017 */
data postpair;
        input gdpdef cpi;
        datalines;
***************;
run;
ods graphics on;

proc ttest;
        paired gdpdef*cpi;
        title 't-test (paired) of gdp_def vs cpi from 2008-2017';
run;
ods graphics off;


/* Correlations among wage growth, inflation, median income, and median */
data growthrates;
        input year mhousehold mpersonal mhomesold mrental gdpdef cpi;
        datalines;
***************;
run;

proc print data=growthrates;
        title 'All Rates from 1984-2015';
run;

ods graphics on;
proc corr data=growthrates plots=matrix(histogram);
        title 'Pearson Corr Among Rate Changes';
run;
ods graphics off;

/* Simple Linear Regressions to predict median home price */

ods graphics on;
proc reg;
        model mhomesold = mhousehold;
        title 'mhomesold = mhousehold';
run;
proc reg;
        model mhomesold = mpersonal;
        title 'mhomesold = mpersonal';
run;
proc reg;
        model mhomesold = mrental;
        title 'mhomesold = mrental';
run;
proc reg;
        model mhomesold = cpi;
        title 'mhomesold = cpi';
run;
proc reg;
```

```
      model mhomesold = gdpdef;
      title 'mhomesold = gdpdef';
run;
ods graphics off;
```