[redacted]

Name: Ben Chu
[redacted]
ben-chu@berkeley.edu

Problem 1: Golden State Warriors

a)

ttest wage, by (guard)

t-statistic: 2.0530

p-value: .0410

Since p-values are below .05, we reject that the population mean of salaries are equal. Another way of saying the same thing: t=2.053>1.96. We could not plausibly reject this p-value at .01.

$t = 2.05 > 1.96$
$p = .041 < .05$

Reject that guards in the NBA are paid more than different positions (lower average salary). This is the one-sided hypothesis.

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. ttest wage, by(guard)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	156	1529.706	87.72705	1095.71	1356.412	1703.001
1	113	1277.658	78.31673	832.5182	1122.484	1432.833
combined	269	1423.828	60.9573	999.7741	1303.811	1543.844
diff		252.048	122.7685		10.33043	493.7656

```
    diff = mean(0) - mean(1)                                      t =   2.0530
Ho: diff = 0                                     degrees of freedom =      267

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9795         Pr(|T| > |t|) = 0.0410          Pr(T > t) = 0.0205
```

b)

gen degree = coll >= 4
ttest wage, by(degree)

Greater than 4 years NBA experience
More education is correlated with a smaller salary

t-statistic: 2.3705
p-value: .0185

Reject the null hypothesis that NBA players who obtain a college degree earn the same amount as players with no college education at 10% and 5% significance; however, this does not hold up at the 1% significance level. This may happen because players who leave college early are usually better players who have the opportunity to play in the NBA and make a higher salary at a younger age.

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. gen degree = coll >= 4

. ttest wage, by(degree)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	49	1727.408	156.2117	1093.482	1413.323	2041.493
1	220	1356.212	65.22361	967.4225	1227.666	1484.758
combined	269	1423.828	60.9573	999.7741	1303.811	1543.844
diff		371.1963	156.5877		62.89262	679.5001

```
    diff = mean(0) - mean(1)                                      t =   2.3705
Ho: diff = 0                                     degrees of freedom =      267

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9908         Pr(|T| > |t|) = 0.0185          Pr(T > t) = 0.0092
```

c)
Index
ppg = “points”
total minutes = “min”
total games played = “games”

gen productivity = points/(minutes/games)
ttest productivity, by (guard)

t-statistic: -.9746
p-value: .3307

Fail to reject the null that guards are as productive as other positions.

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

There is no statistical difference that there is a difference in the overall productivity of guards in comparison to other positions.

```
. gen productivity = points/(minutes/games)

. ttest productivity, by(guard)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	156	.4038649	.0098823	.1234298	.3843435	.4233862
1	113	.4177012	.009596	.102007	.3986879	.4367145
combined	269	.4096771	.0070068	.11492	.3958818	.4234725
diff		-.0138363	.0141975		-.0417896	.0141169

```
    diff = mean(0) - mean(1)                                      t =  -0.9746
Ho: diff = 0                                     degrees of freedom =      267

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1653         Pr(|T| > |t|) = 0.3307          Pr(T > t) = 0.8347
```

d)

correlate points assists rebounds

Points, rebounds, and assists are positively correlated with each other.

The correlation between assists and rebounds is weak at .06.

Reasonably strong correlation between points and assists at .5393.

Reasonably strong between points and rebounds at .5633.

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. correlate points assists rebounds
(obs=269)

             |   points  assists rebounds
-------------+---------------------------
      points |   1.0000
     assists |   0.5393   1.0000
    rebounds |   0.5633   0.0600   1.0000


. correlate points rebounds assists rebounds, cov
(obs=269)

             |   points rebounds  assists rebounds
-------------+------------------------------------
      points |  34.6037
    rebounds |  9.58661  8.36933
     assists |  6.81438  .372593  4.61444
    rebounds |  9.58661  8.36933  .372593  8.36933
```

e)

gen index = points + rebounds + 2*assists
ttest index, by(guard)

Judging by the ttest, guards are usually more productive during games. As a result, the null hypothesis should be rejected. This means to reject the null that the performance of guards is the same as other positions' performance during games at the 10% and 5% significance level. We

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

cannot reject at the 1% significance level.

```
. gen index = points + rebounds + 2*assists

. ttest index, by(guard)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	156	18.45513	.8462608	10.56979	16.78344	20.12682
1	113	21.26549	.9520769	10.12072	19.37907	23.1519
combined	269	19.63569	.6375833	10.45714	18.38038	20.891
diff		-2.810359	1.282715		-5.335882	-.2848355

```
    diff = mean(0) - mean(1)                                      t =  -2.1909
Ho: diff = 0                                     degrees of freedom =      267

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0147         Pr(|T| > |t|) = 0.0293          Pr(T > t) = 0.9853
```

f)

gen indexper1k = index/wage
ttest indexper1k, by(guard)

Performance index per $1000 shows that guards outperform their salary in comparison to other positions, which indicates good value for what they are paid relative to their production.

Reject the null hypothesis that claims that guards are paid as much as other positions on the court relative to their performance at the 10% and 5% significance level; however, not at the 1% level, but we cannot reject at the 1% level.

p-value = .0115

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

T-statistic = -2.5462

```
. gen indexper1k = index/wage

. ttest indexper1k, by(guard)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     156    .0179347     .001336    .0166866    .0152956    .0205738
       1 |     113    .0241632    .0022052    .0234414    .0197939    .0285325
---------+--------------------------------------------------------------------
combined |     269    .0205512    .0012197    .0200042    .0181498    .0229525
---------+--------------------------------------------------------------------
    diff |           -.0062285    .0024462               -.0110449   -.0014121
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -2.5462
Ho: diff = 0                                     degrees of freedom =      267

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0057         Pr(|T| > |t|) = 0.0115          Pr(T > t) = 0.9943
```

Problem 2: Crime on Campus

a)

	Enrollment	Police	Crime
Number of observations	97	97	97
Sample mean	16076.35	20.49	394.45
Sample median	11990	16	187

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

Sample standard deviation	12298.99	15.63	460.78
Sample skewness	1.22	1.31	1.85
Mean, if public	17473.01	22.15	432.8
Mean, if private	6183.33	8.75	122.83

Summarize enrollment police crime, detail

Number of observations, sample mean, sample median (50%), sample standard deviation, sample skewness

by private, sort: summ enrollment police crime

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. summarize enrollment police crime, detail

                         enrollment
-------------------------------------------------------------
      Percentiles      Smallest
 1%         1799           1799
 5%         2446           1859
10%         3712           2123       Obs                  97
25%         6485           2167       Sum of Wgt.          97

50%        11990                      Mean           16076.35
                        Largest       Std. Dev.      12298.99
75%        21836          43030
90%        33376          49961       Variance       1.51e+08
95%        42790          54311       Skewness       1.224383
99%        56350          56350       Kurtosis       4.192367

                           police
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            5              2
10%            6              3       Obs                  97
25%            9              4       Sum of Wgt.          97

50%           16                      Mean           20.49485
                        Largest       Std. Dev.      15.63058
75%           27             60
90%           48             63       Variance       244.3151
95%           51             65       Skewness       1.310145
99%           74             74       Kurtosis       4.283173
```

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

crime

	Percentiles	Smallest		
1%	1	1		
5%	25	1		
10%	44	15	Obs	97
25%	85	17	Sum of Wgt.	97
50%	187		Mean	394.4536
		Largest	Std. Dev.	460.7839
75%	491	1712		
90%	1145	1839	Variance	212321.8
95%	1384	1962	Skewness	1.852137
99%	2052	2052	Kurtosis	6.005492

. by private, sort: summ enrollment police crime

-> private = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
enrollment	85	17473.01	12468.03	1859	56350
police	85	22.15294	15.84712	1	74
crime	85	432.8	477.9452	1	2052

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
-> private = 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
enrollment	12	6183.333	3347.161	1799	13570
police	12	8.75	6.538348	2	26
crime	12	122.8333	129.4441	15	426

```
. summ enrollment police crime if private == 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
enrollment	12	6183.333	3347.161	1799	13570
police	12	8.75	6.538348	2	26
crime	12	122.8333	129.4441	15	426

```
. summ enrollment police crime if private == 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
enrollment	85	17473.01	12468.03	1859	56350
police	85	22.15294	15.84712	1	74
crime	85	432.8	477.9452	1	2052

b)

correlate enrollment police crime, cov
correlate enrollment police crime

Each variable in this set are positively correlated, which may mean that college campuses with a higher number of police officers also have a higher number of crimes committed. Additionally,

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

larger campuses have more crime and more police officers present. This positive correlation makes sense because the values are absolute, not relative.

```
. correlate enrollment police crime, cov
(obs=97)
```

	enroll~t	police	crime
enrollment	1.5e+08		
police	137462	244.315	
crime	4.7e+06	5209.51	212322

```
. correlate enrollment police crime
(obs=97)
```

	enroll~t	police	crime
enrollment	1.0000		
police	0.7151	1.0000	
crime	0.8360	0.7233	1.0000

c)

ttest crime, by(private) unequal
ttest crime, by(private)

If unequal variance is assumed of public and private schools, then the null hypothesis claiming that crimes at the two schools are the same at the 5% and 1% significance level and rejected for both levels. There is a small p-value. Pr(|T|>t) = 0.0000

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

Assuming equal variance of public and private schools, then the null is rejected at the 5% significance level, but not at the 1% significance level. p-value= .0284 is < .05, but > .01.

Pr(|T|>t) = 0.0284
p-value: .0284

t-statistic: 2.2259

```
. ttest crime, by(private) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	85	432.8	51.84044	477.9452	329.7096	535.8904
1	12	122.8333	37.36731	129.4441	40.58844	205.0782
combined	97	394.4536	46.78551	460.7839	301.5851	487.3221
diff		309.9667	63.9042		182.2784	437.655

```
    diff = mean(0) - mean(1)                                      t =   4.8505
Ho: diff = 0                     Satterthwaite's degrees of freedom =  63.3564

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. ttest crime, by(private)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	85	432.8	51.84044	477.9452	329.7096	535.8904
1	12	122.8333	37.36731	129.4441	40.58844	205.0782
combined	97	394.4536	46.78551	460.7839	301.5851	487.3221
diff		309.9667	139.2572		33.5062	586.4271

```
    diff = mean(0) - mean(1)                                      t =   2.2259
Ho: diff = 0                                     degrees of freedom =       95

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9858         Pr(|T| > |t|) = 0.0284          Pr(T > t) = 0.0142
```

d)

gen crimerate = 1000*(crime/enrollment)

ttest crimerate, by(private) unequal

When adjusting for the greater enrollments at public vs. private schools crime is narrowly higher at the public schools than at the private schools.

22.34 public vs. 21.20 private per 1000 students.

This narrow difference is not statistically significant.

p-value: .8491
t-value: .1940

This p-value is significantly higher than .05 (or.1 or .01) significance level.

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. gen crimerate = 1000*(crime/enrollment)

. ttest crimerate, by(private) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      85    22.34158    1.861962    17.16644    18.63887     26.0443
       1 |      12    21.20101     5.57801    19.32279     8.92389    33.47812
---------+--------------------------------------------------------------------
combined |      97    22.20048    1.760902    17.34288    18.70512    25.69584
---------+--------------------------------------------------------------------
    diff |            1.140577    5.880569               -11.50996    13.79112
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   0.1940
Ho: diff = 0                     Satterthwaite's degrees of freedom =  13.5659

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.5755         Pr(|T| > |t|) = 0.8491          Pr(T > t) = 0.4245
```

Problem 3: Post Katrina Employment in New Orleans

a)

	E	U	T
Returned	139	8	147
Not	79	23	102
Total	218	31	249

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

Joint Probabilities	Employed (Y=1)	Unemployed (Y=0)	Marginal
Returned to pre-storm address (X=1)	.558	.032	.590
Have not returned (X=0)	.317	.092	.410
Total	.875	.124	

Work

P(X=1, Y=1) = 139/249 = .558
P(X=1, Y=0) = 8/249 = .032
P(X=0, Y=0) = 23/249 = .092
P(X=0, Y=1) = 79/249 = .317
P(X=1) = (139+8)/249 = .590
P(Y=1) = (139+79)/249 = .876
P(X=0) = (79+23)/249 = .410
P(Y=0) = (8+23)/249 = .124

Conditional probability table

Conditional	Employed (Y=1)	Unemployed (Y=0)
Given returned to pre-storm address (X=1)	.946	.054
Given have not returned yet (X=0)	.775	.225

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

P(Y=1 | X=1) = 139/147 = .946
P(Y=1 | X=0) = 79/102 = .775
P(Y=0 | X=1) = 8/147 = .054
P(Y=0 | X=0) = 23/102 = .225

b)

1 = returned
0 = not returened
1 = employed
0 = unemployed

E of being employed conditional on returning to pre-storm address

E[Y | X = 1] = P(Y=1 | X=1)(1) + P(Y=0 | X=1)(0) = .946

E[Y | X = 0] = P(Y=1 | X=0)(1) + P(Y=0 | X=0)(0) = .775

law of iterated expectations: E[Y] = E[E(Y|X)]

E[E(Y|X)] = P(X=1)*E(Y|X=1) + P(X=0)*E[Y|X=0] = (.590*.946)+(.410*.775) = .876

.876 matches the unconditional expectation of E[Y] = P(Y=1)(1) + P(Y=0)(0) =.876

c)

Covariance of “returned to pre-storm address” status and “employment” status

S_{XY} = 1/(n-1) ∑(Xi-$\bar{X}$)(Yi-$\bar{Y}$)   summation from i=1 to n
= 1/248∑(Xi-E[X])(Yi-E[Y])   summation from i=1 to 249
= 1/248 (139*(1-.590)(1-.875) + 8*(1-.590)*(0-.875) +79(0-.590)(1-.875) + 23(0-.590)(0-.875))
= .0415

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

d)

Since .875 ≠ .946 = (or P(Y=1) ≠ P(Y=1 | X=1)) "employment" status is not independent from "return" status, so we are able to conclude that X and Y are positively correlated with each other (relationship).

e)

2 reasons to explain the difference in employment status between returned vs. not returned:

1: People who are unemployed after the evacuation may have not have had good work prior to the storm, so they gave up on the town and cut their losses. People without money, job, or housing have no reason or resources to motivate them to return to whatever is left of their homes.

2: People who returned may have better luck finding jobs or job opportunities in town vs. people who have not returned to town.

Problem 4: Pollution

a)

by oecd, sort: summ gdppc co2pc

gdppc mean oecd: 31423.91 vs. gdppc mean non-oecd: 5585.894 greater for oecd
gdppc sd oecd:17180.88 vs. gdppc sd non-oecd: 9827.863 greater for oecd
co2pc mean oecd: 8.984034 vs. co2pc mean non-oecd: 4.074291 greater for oecd
co2pc sd oecd: 4.099933 vs. co2pc sd non-oecd: 6.430327 greater for non-oecd

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. by oecd, sort: summ gdppc co2pc
```

-> oecd = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
gdppc	150	5585.894	9827.863	150.7424	76393.83
co2pc	164	4.074291	6.430327	.0091268	40.31009

-> oecd = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
gdppc	34	31423.91	17180.88	7833.529	80276.01
co2pc	34	8.984034	4.099933	3.763572	21.36027

b)

Since the sample standard deviations of each subsample are extremely different, unequal variance should be assumed.

Reject the null hypothesis at 5% significance level.
Reject the sample means co2 emissions per capita are equal to each other.

ttest co2pc, by(oecd) unequal

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. ttest co2pc, by(oecd) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	164	4.074291	.5021242	6.430327	3.082785	5.065798
1	34	8.984034	.7031327	4.099933	7.5535	10.41457
combined	198	4.917379	.4521244	6.361954	4.025754	5.809004
diff		-4.909743	.8640164		-6.632343	-3.187143

```
    diff = mean(0) - mean(1)                                      t =  -5.6825
Ho: diff = 0                          Satterthwaite's degrees of freedom =  71.4772

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

c)

generate lgdp = log(gdp)
generate lco2 = log(co2)

View growth rates, not the absolute levels of emission and GDP since not many countries in the sample have very high levels of emissions and GDP in relation to the other countries in the sample.

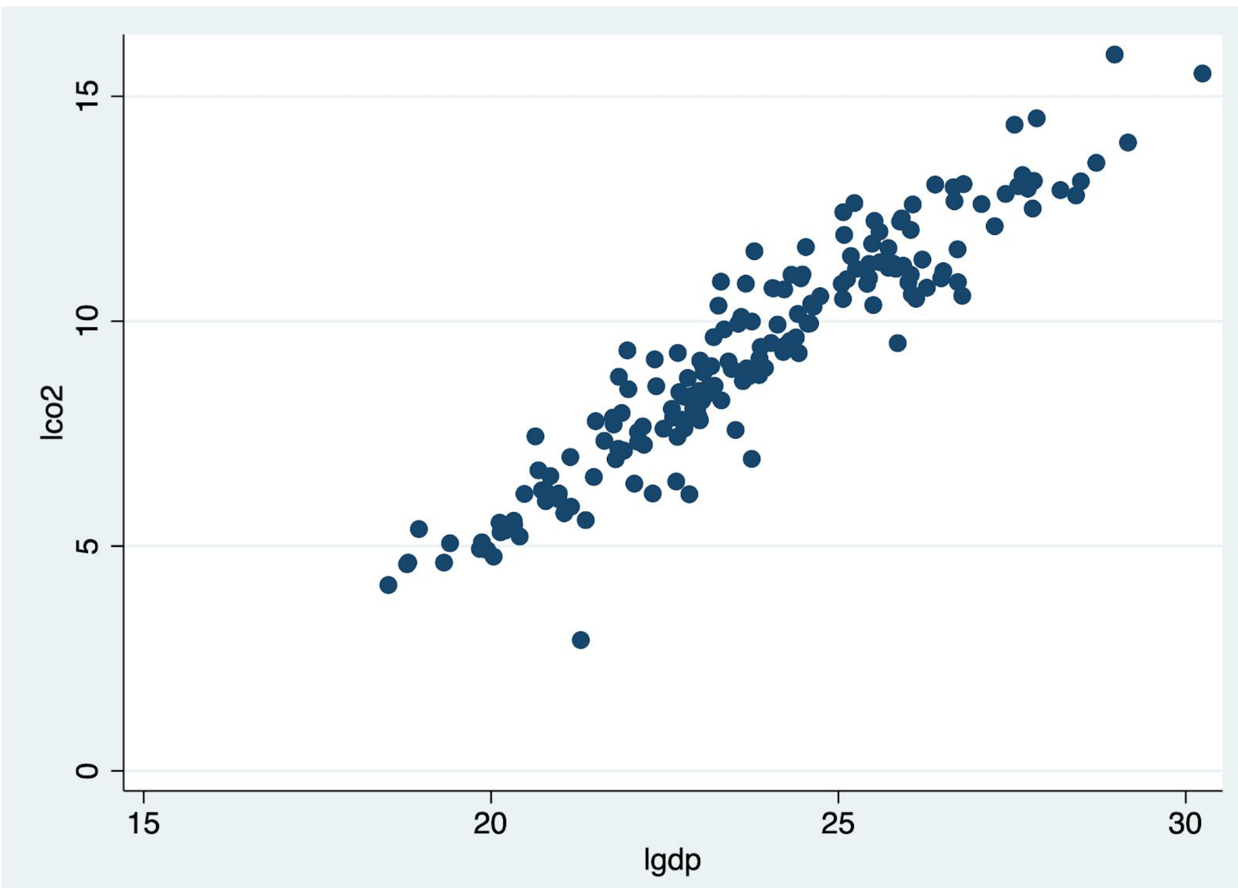
When taking the natural log of the two variables, the big picture of the data patterns become easier to visualize. The spread of values shrinks, especially for skewed data, thereby making it easier to visualize.

d)

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

scatter lco2 lgdp



GDP growth rates on x-axis
CO2 emissions on y-axis

e)

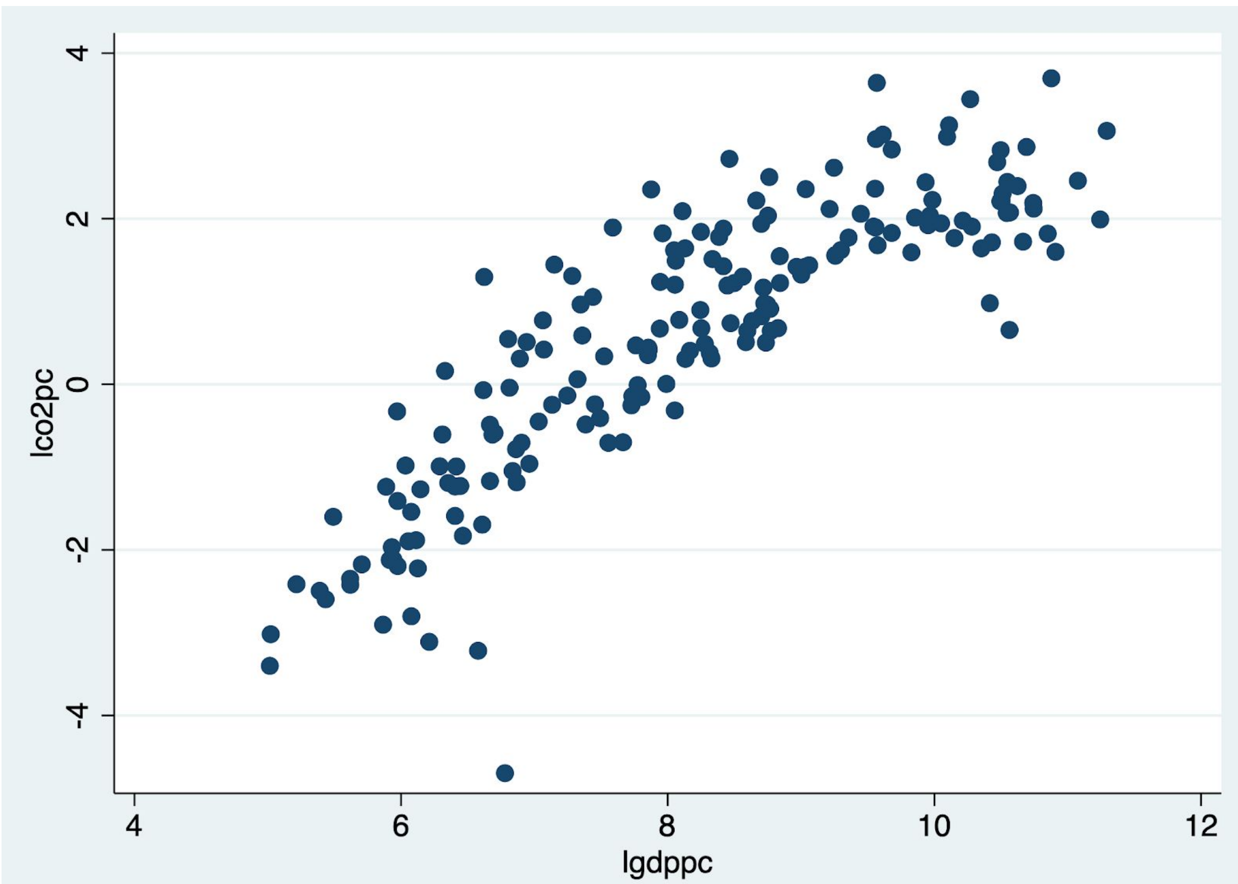
Growth rate per capita GDP on x-axis
Growth rate of per capita emissions on y-axis

generate lgdppc = log(gdppc)

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
generate lco2pc = log(co2pc)
scatter lco2pc lgdppc
```



GDP and emission rates appear to be positively correlated even after controlling for the population sizes.

The claim of high-GDP countries does not seem convincing.

The scatterplot shows a concave relationship between GDP and emissions growth rates, which begins to tilt downwards for higher GDP growth rates.

Higher lgdppc ≠ higher lco2pc after a certain point

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

Problem 5: Wages in Los Angeles

a)

summm wage, detail

Mean: 13.41386

$13.41386
IQR (interquartile range): 6.5-15.81197
25th percentile-75th percentile

$6.50-$15.81

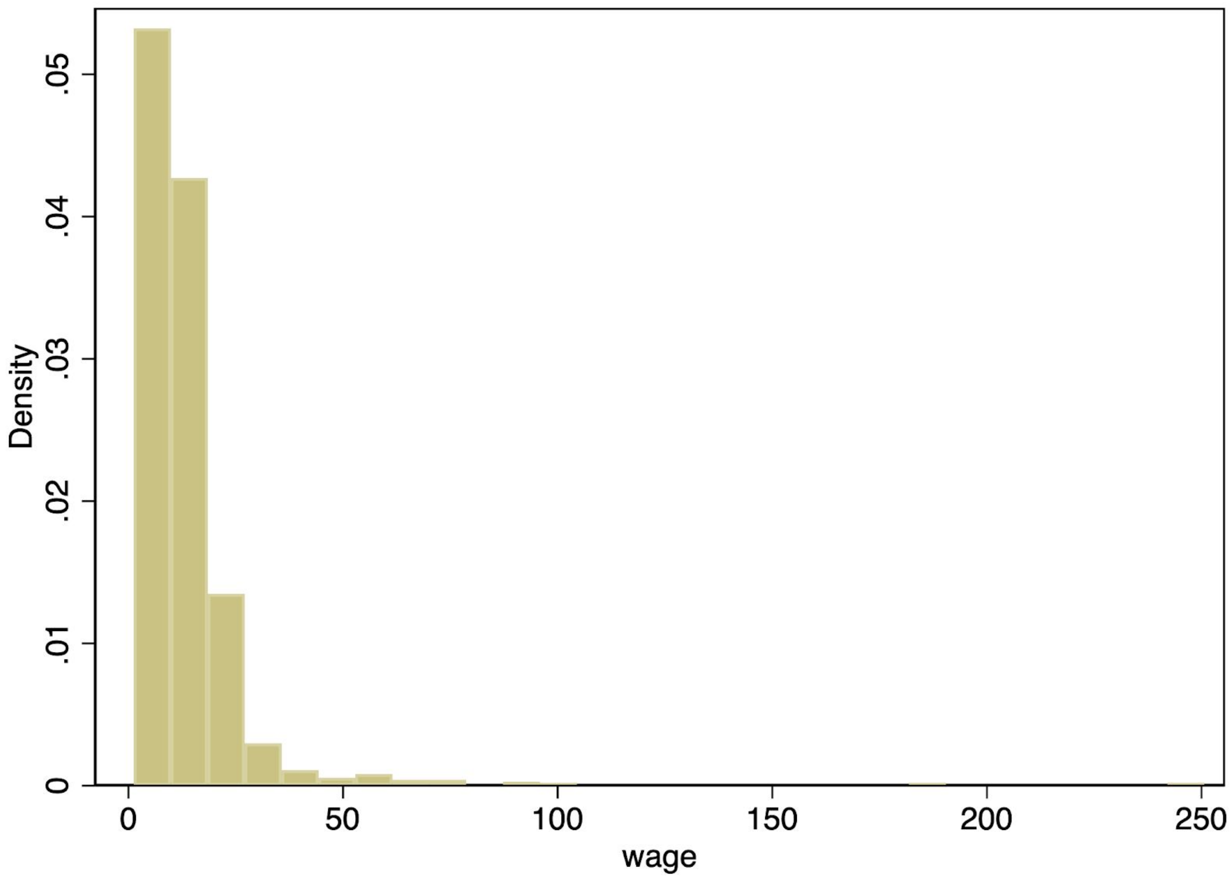
```
. summ wage, detail

                            wage
-------------------------------------------------------------
      Percentiles      Smallest
 1%        1.875           1.25
 5%     3.571429           1.25
10%     4.299754            1.5       Obs                 863
25%          6.5       1.593367       Sum of Wgt.         863

50%     10.57692                      Mean           13.41386
                        Largest       Std. Dev.      14.65928
75%     15.81197       93.99808
90%     21.92308         97.758       Variance       214.8945
95%     29.16667       187.9962       Skewness        8.17203
99%     63.33333       250.6615       Kurtosis       108.8581
```

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

b)

histogram wage

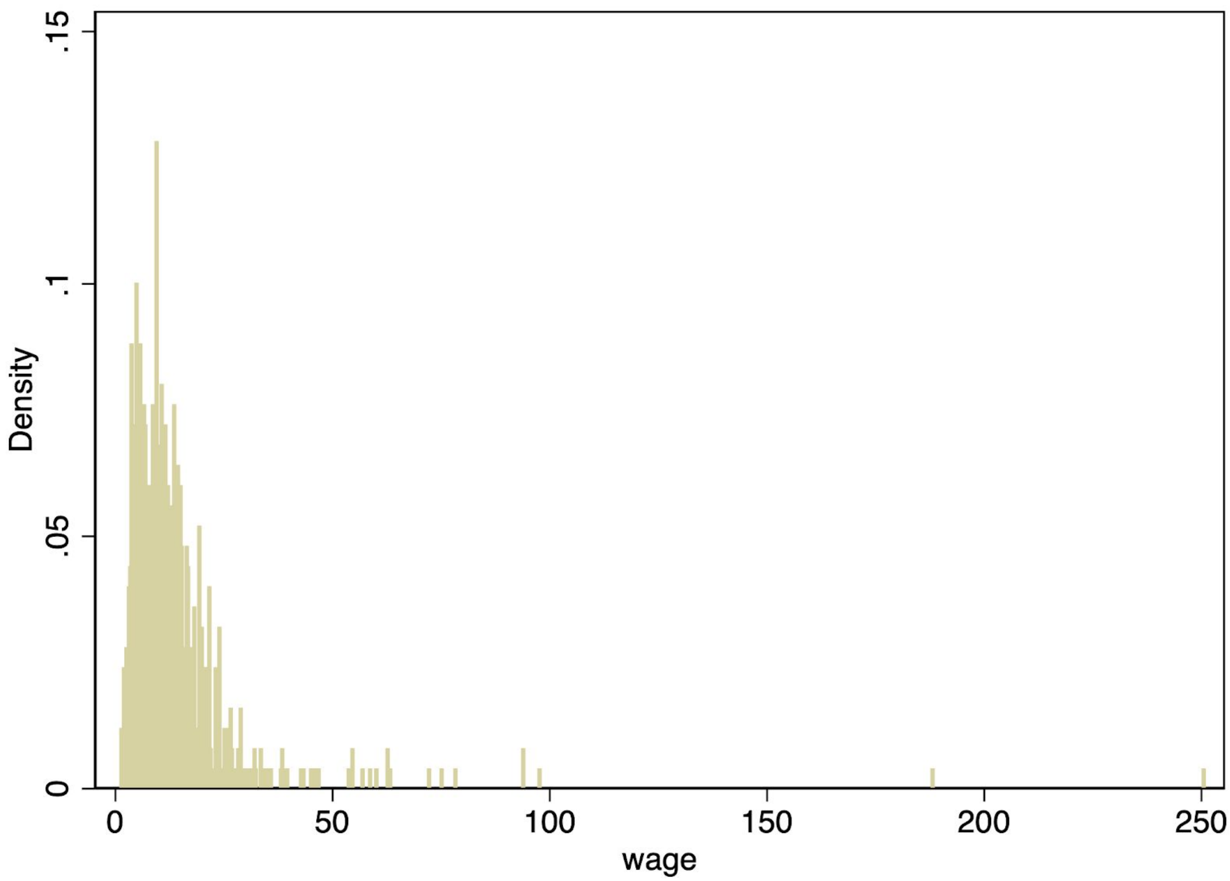


or

histogram wage, bin(863)

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu



c)

The distribution is skewed

Skewness 8.17203
summ wage, detail
Positive, asymmetrically skewed distribution
8.17203 is rightly skewed

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

d)

Take log of wage

gen lnwage = ln(wage)

summ lnwage, detail

Mean wage 2.335614
2.34

IQR (interquartile range): 1.871802-2.760767
1.87-2.76
25th-75th

```
. summ lnwage, detail

                           lnwage
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .6286086       .2231435
 5%     1.272966       .2231435
10%     1.458558       .4054651       Obs                 863
25%     1.871802       .4658496       Sum of Wgt.         863

50%     2.358675                      Mean           2.335614
                        Largest       Std. Dev.      .6847209
75%     2.760767       4.543274
90%      3.08754       4.582495       Variance       .4688427
95%     3.373027       5.236422       Skewness       .2427184
99%     4.148412       5.524104       Kurtosis       4.028374
```

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

e)

How many workers have at least an 8th grade education?

tab education

Education is discrete in this data set, not continuous, proportion ≥ 8 years of education

100-8.69=91.31

91.31% in the same sample have at least an 8th grade education

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu

```
. tab education
```

education	Freq.	Percent	Cum.
0	5	0.58	0.58
3	11	1.27	1.85
7	59	6.84	8.69
9	29	3.36	12.05
10	26	3.01	15.06
11	75	8.69	23.75
12	174	20.16	43.92
13	194	22.48	66.40
14	55	6.37	72.77
16	155	17.96	90.73
18	75	8.69	99.42
20	5	0.58	100.00
Total	863	100.00	

f)

ln(wages) on y-axis
Education on x-axis

twoway (scatter lnwage education)

Name: Julian Bertalli
SID: 3033589368
Section ID: 103
jjbertalli@berkeley.edu

Name: Ben Chu
SID: 3032040857
Section ID: 110
ben-chu@berkeley.edu