ANALYSIS OF PANEL DATA

An Introduction

datascience@berkeley

Exploratory Panel Data Analysis: A Two-Period Panel

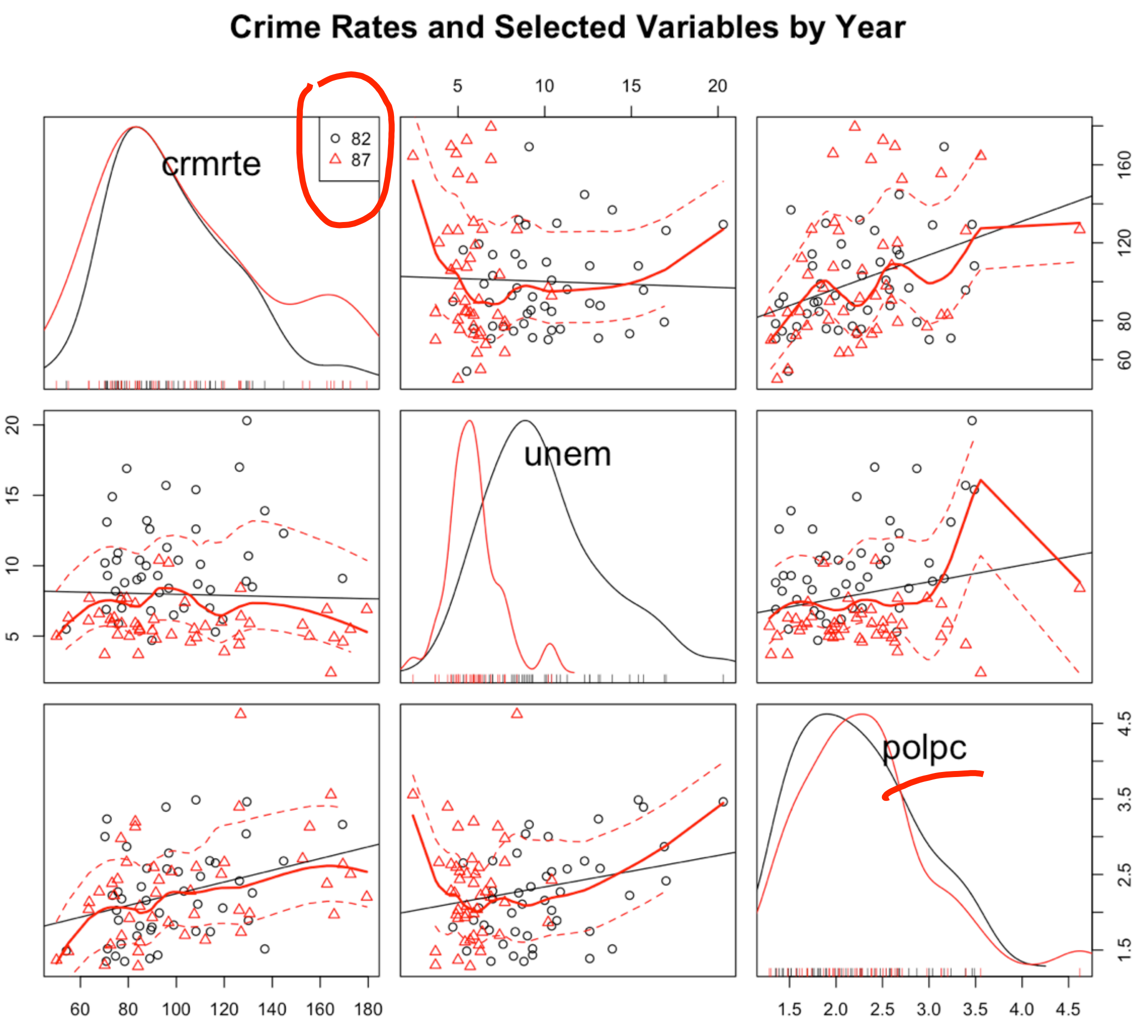
Exploring Panel Data

- The exploratory data analysis techniques studied in w203 and earlier in this course can be applied.
- However, with panel data, we first need to find out how many panels there are in the dataset.
- Then, not only do we need to explore patterns of each feature (as well as the combinations of the features) in each panel, we will have to explore the "dynamic" "temporal dependence" of the features.
- A subset of an example data (crime4) is shown here.

```
'data.frame':   630 obs. of  59 variables:
 $ county  : int  1 1 1 1 1 1 1 3 3 3 ...
 $ year    : int  81 82 83 84 85 86 87 81 82 83 ...
 $ crmrte  : num  0.0399 0.0383 0.0303 0.0347 0.0366 ...
 $ prbarr  : num  0.29 0.338 0.33 0.363 0.325 ...
 $ prbconv : num  0.402 0.433 0.526 0.605 0.579 ...
 $ prbpris : num  0.472 0.507 0.48 0.52 0.497 ...
 $ avgsen  : num  5.61 5.59 5.8 6.89 6.55 ...
 $ polpc   : num  0.00179 0.00177 0.00184 0.00189 0.00192 ...
 $ density : num  2.31 2.33 2.34 2.35 2.36 ...
 $ taxpc   : num  25.7 24.9 26.5 26.8 28.1 ...
 $ west    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ central : int  1 1 1 1 1 1 1 1 1 1 ...
 $ urban   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ pctmin80: num  20.2 20.2 20.2 20.2 20.2 ...
 $ wcon    : num  206 213 220 223 244 ...
 $ wtuc    : num  334 369 1395 399 359 ...
 $ wtrd    : num  182 190 197 201 207 ...
 $ wfir    : num  272 301 310 350 383 ...
 $ wser    : num  216 232 240 252 261 ...
 $ wmfg    : num  229 240 270 282 299 ...
 $ wfed    : num  409 420 439 459 490 ...
 $ wsta    : num  236 254 250 262 281 ...
 $ wloc    : num  231 237 249 264 289 ...
 $ mix     : num  0.0999 0.103 0.0807 0.0785 0.0932 ...
 $ pctymle : num  0.0877 0.0864 0.0851 0.0838 0.0823 ...
 $ d82     : int  0 1 0 0 0 0 0 0 1 0 ...
 $ d83     : int  0 0 1 0 0 0 0 0 0 1 ...
 $ d84     : int  0 0 0 1 0 0 0 0 0 0 ...
 $ d85     : int  0 0 0 0 1 0 0 0 0 0 ...
 $ d86     : int  0 0 0 0 0 1 0 0 0 0 ...
 $ d87     : int  0 0 0 0 0 0 1 0 0 0 ...
```

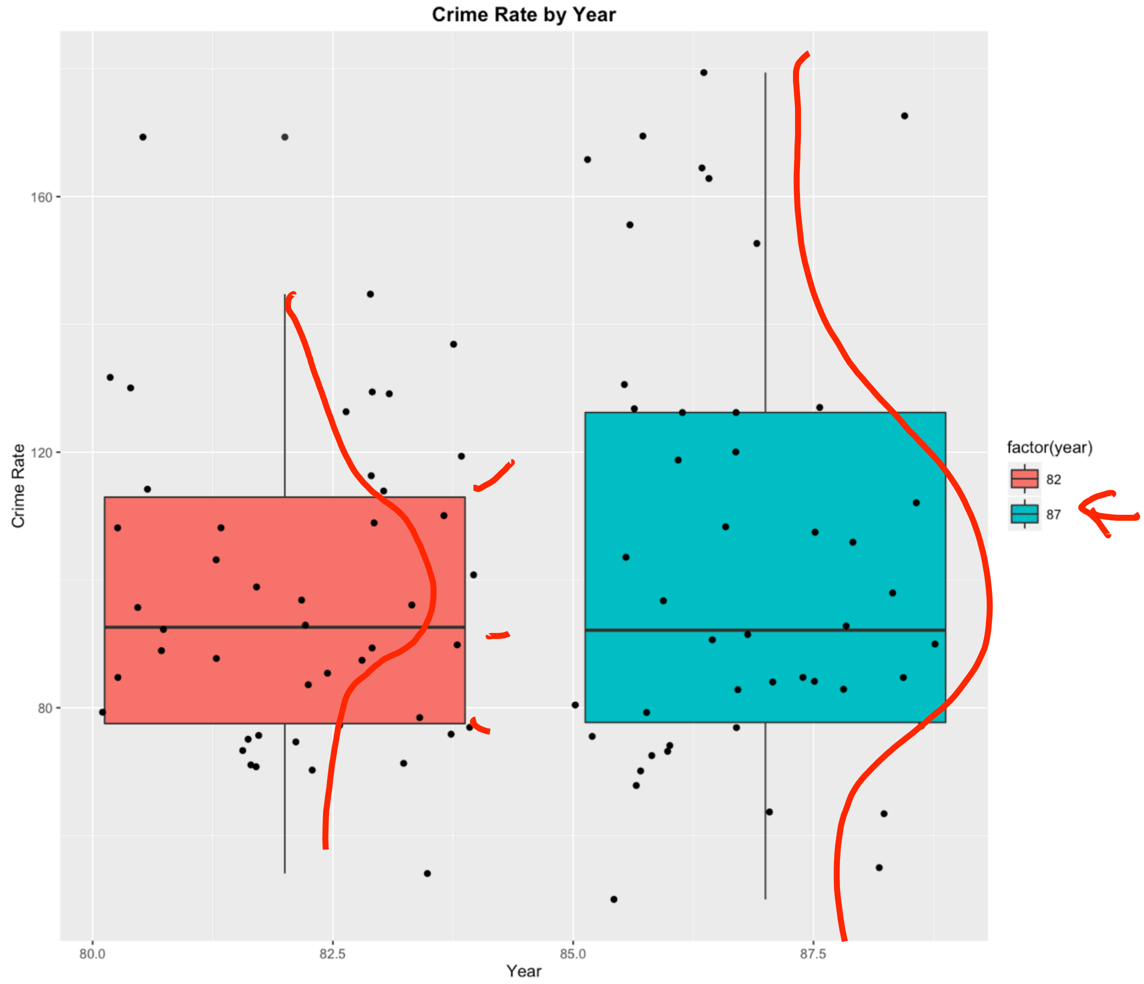
```
> table(crime4$year)

81 82 83 84 85 86 87
90 90 90 90 90 90 90
```

Distributions and Dependence Over Time



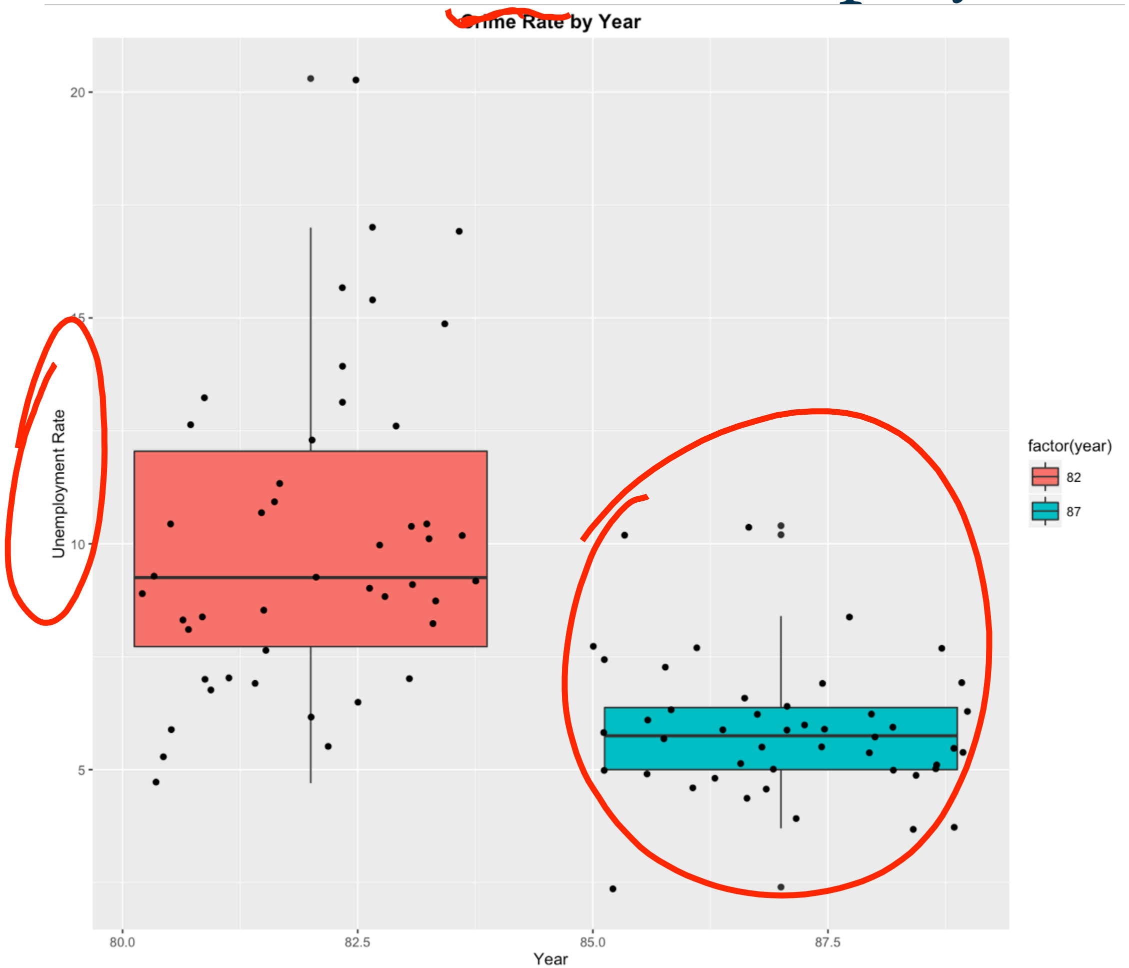
- Recognize that relationship between variable of interest and predictors may change over time.
- The distribution of each of the variables may also change over time.

Distribution of Crime Rate Over Time



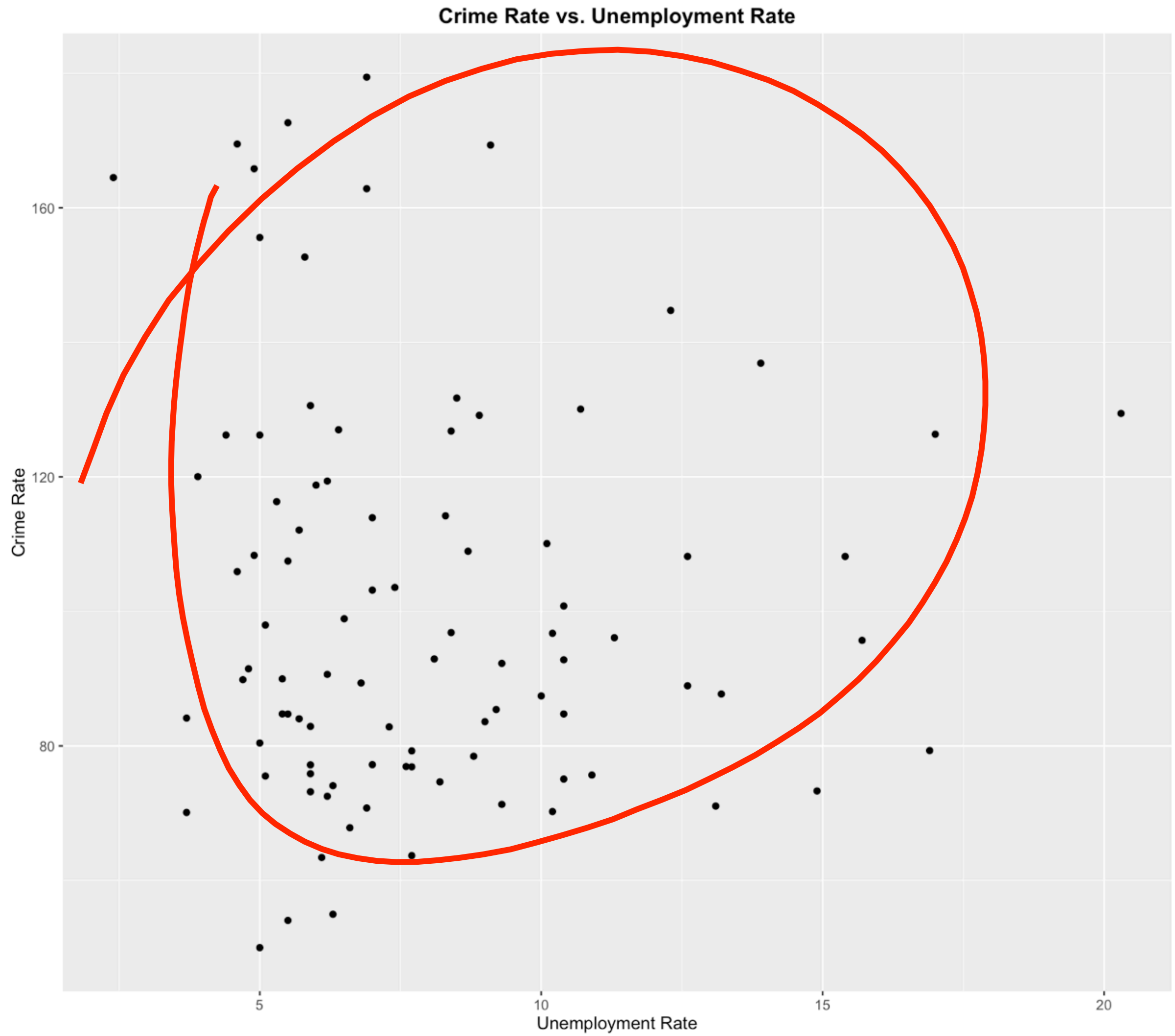

```
summary(crime2$crmrte[crime2$year==82])
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
54.06   77.54   92.61   97.71  113.00  169.30
summary(crime2$crmrte[crime2$year==87])
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
50.02   77.71   92.14  103.90  126.20  179.40
```

Distribution of Unemployment Rate Over Time




```
> summary(crime2$unem[crime2$year==82])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  4.700   7.725   9.250  10.050  12.050  20.300
> summary(crime2$unem[crime2$year==87])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.400   5.000   5.750   5.889   6.375  10.400
```

Relationship Between Crime Rate and Unemployment Rate



Relationship Between Crime Rate and Unemployment Rate Changed Over Time

Notice the change in relationship between crime rate and unemployment rate over time.

Naïve OLS Regression 1 : Using Only the 1982 Panel

```
> ols.fit1<-lm(crmrte ~ unem, data=crimes.82)
> summary(ols.fit1)

Call:
lm(formula = crmrte ~ unem, data = crimes.82)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.693 -17.292  -3.671  16.994  72.854 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   84.569     10.904   7.756 9.07e-10 ***
unem           1.307      1.027   1.272     0.21    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.74 on 44 degrees of freedom
Multiple R-squared:  0.03549,	Adjusted R-squared:  0.01357 
F-statistic: 1.619 on 1 and 44 DF,  p-value: 0.2099
```

Berkeley SCHOOL OF INFORMATION