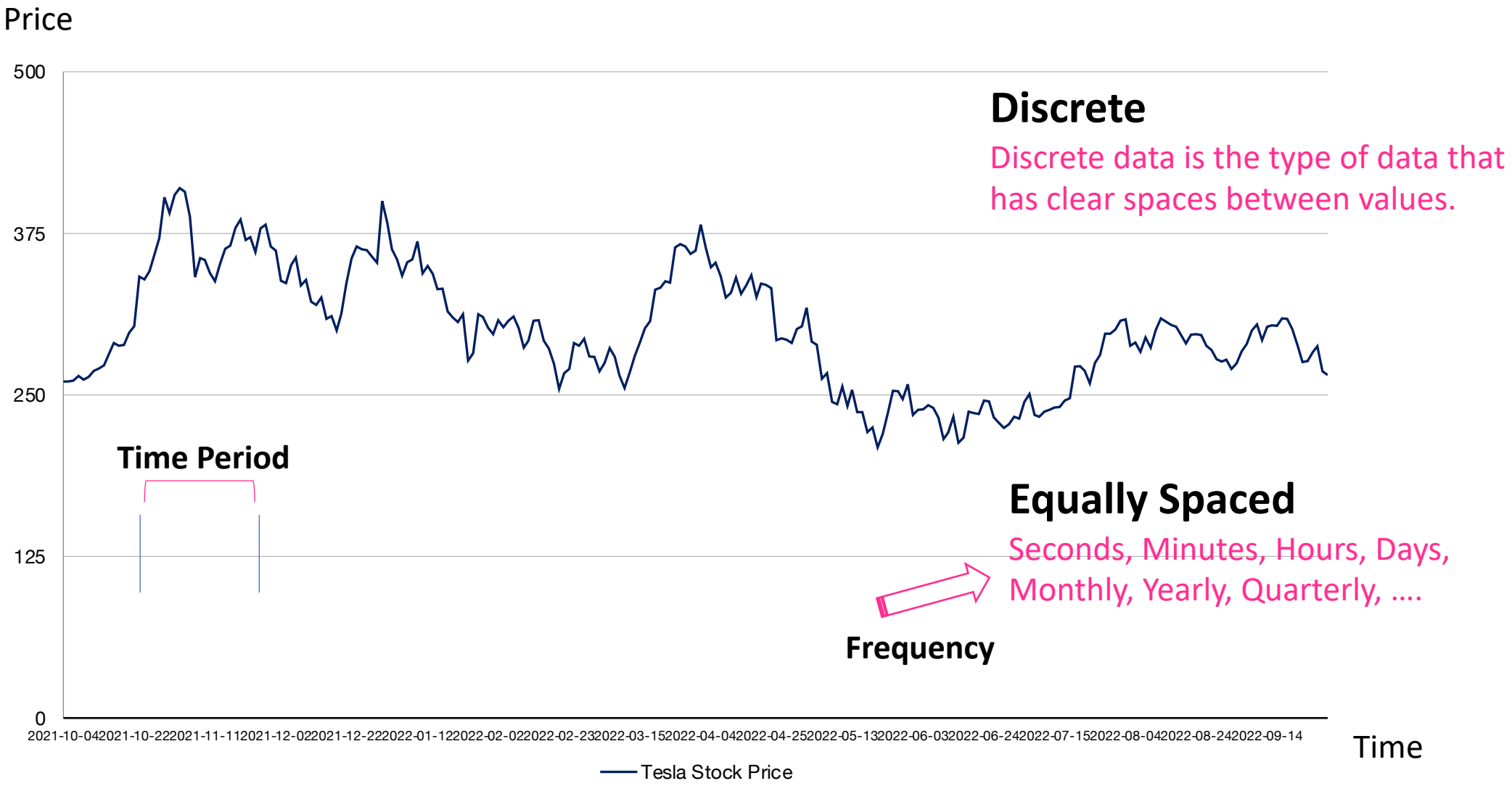
Cross Sectional, Panel and Time Series, Statistics

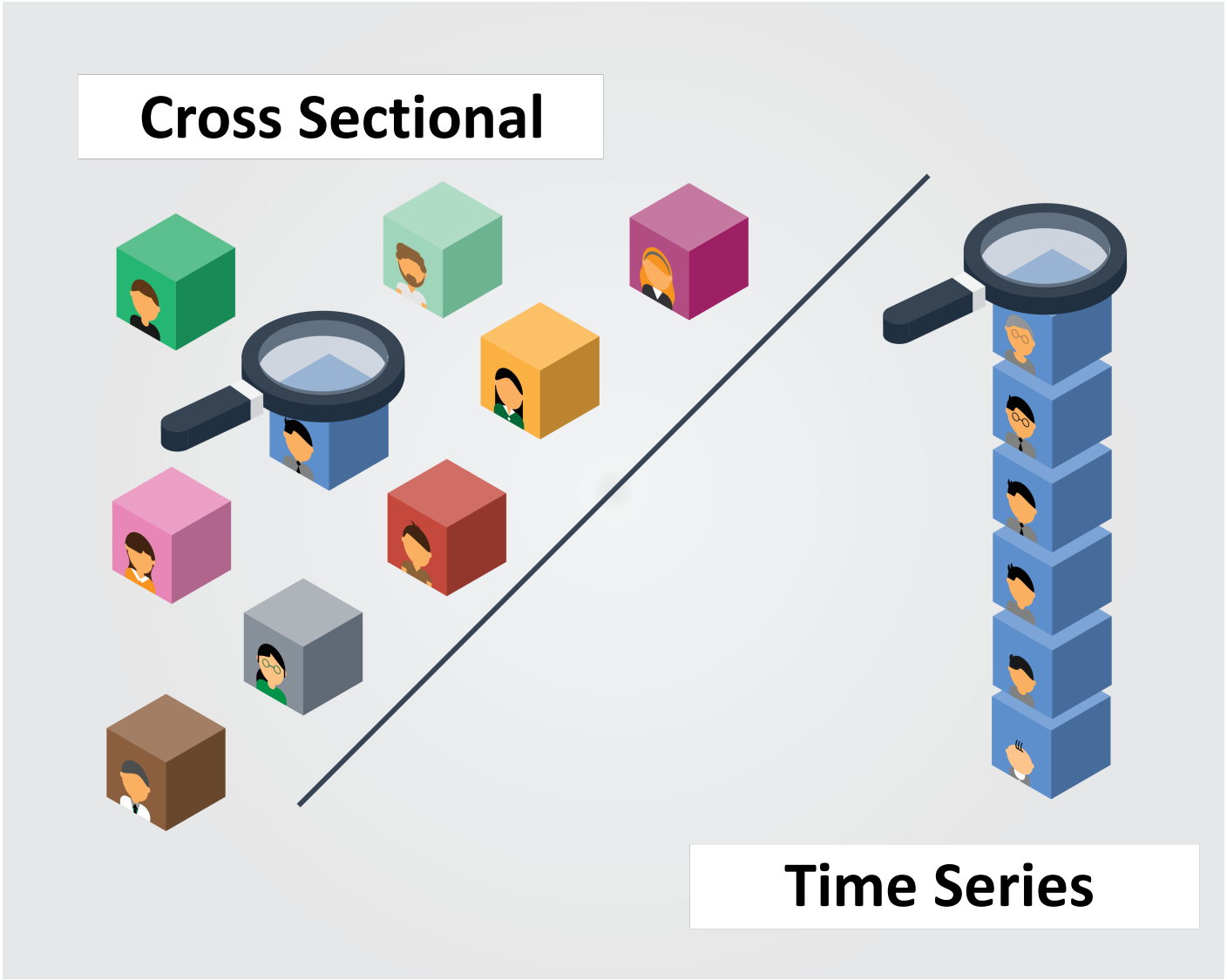
Correlation and Causation

Week 05

What is the time series and cross sectional



What is the time series and cross sectional



Panel Data

Cross-Sectional Data

Observation of the subjects is obtained at the same point in time

Date	County	Government Stability	Socioeconomic Conditions	Investment Profile
2012.12.31	Angola	8.166666667	3	7.5
2012.12.31	Bahrain	6	7	10.5
2012.12.31	DRCongo	6.875	1.5	6
2012.12.31	Egypt	5.916666667	4.666666667	6
2012.12.31	Kenya	6.416666667	1.5	7
2012.12.31	Kuwait	4.875	8.5	8.583333333
2012.12.31	Madagascar	7.666666667	3.5	6.625
2012.12.31	Malawi	6.208333333	2.5	6
2012.12.31	Mozambique	10	3.208333333	8

Time Series Data

Observation are generated over the time

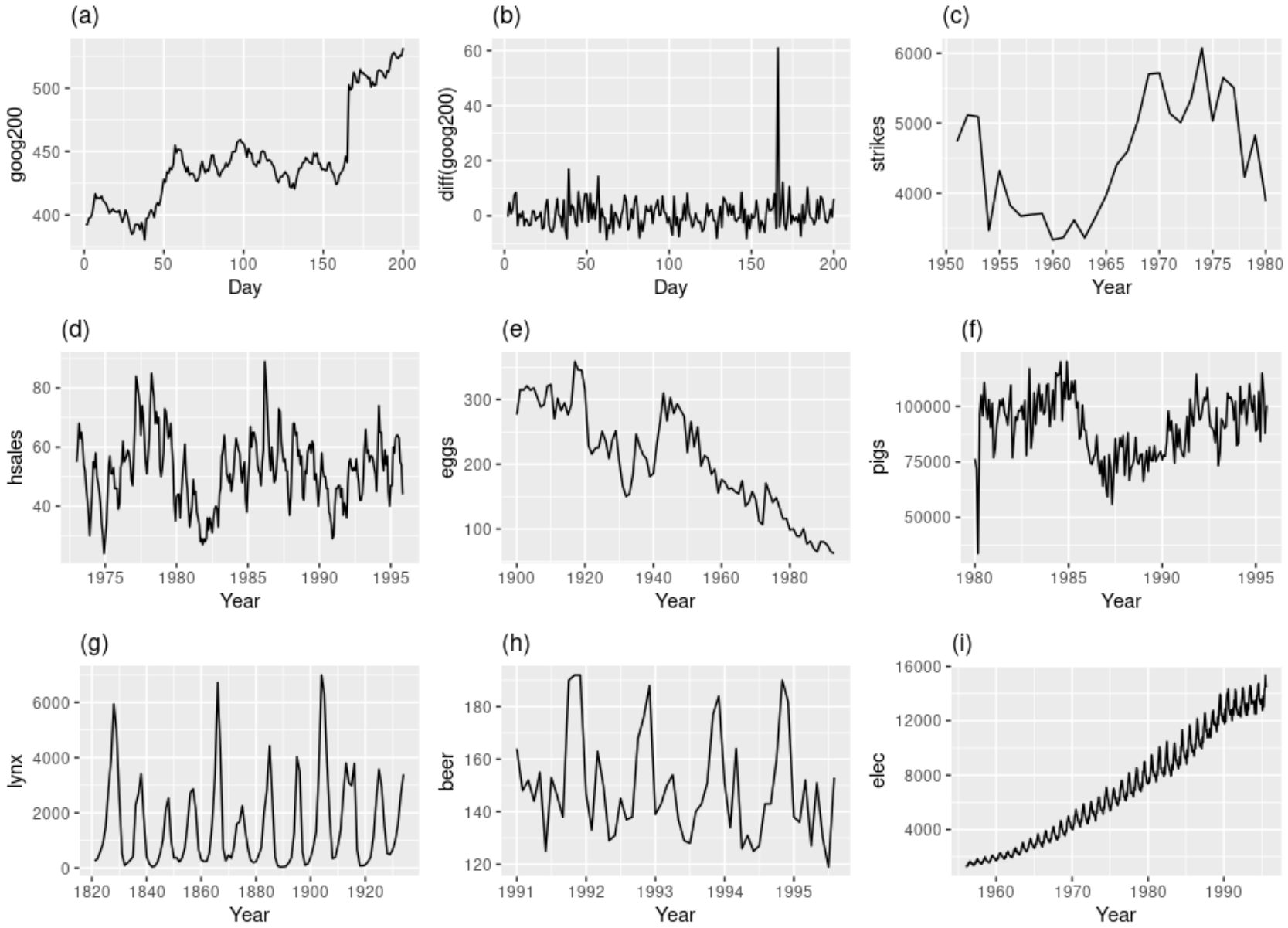
Date	County	Government Stability	Socioeconomic Conditions	Investment Profile
1984-12-31	UAE	7	5.666666667	7.666666667
1985-12-31	UAE	6.416666667	5	6.166666667
1986-12-31	UAE	5.25	5	6
1987-12-31	UAE	4.5	5	6
1988-12-31	UAE	4	5.666666667	6.333333333
1989-12-31	UAE	4	5	6
1990-12-31	UAE	4.416666667	6.833333333	6.666666667
1991-12-31	UAE	6.416666667	8	7.166666667

Panel Data

Combination of time series and cross-section observation

Date	County	Government Stability	Socioeconomic Conditions	Investment Profile
1988.12.31	Angola	6.00	6.00	7.00
1989.12.31	Angola	6.00	6.42	7.00
1990.12.31	Angola	6.00	6.17	7.00
1991.12.31	Angola	6.00	6.00	6.50
1992.12.31	Angola	4.17	4.50	4.92
1993.12.31	Angola	4.58	2.17	3.33
1994.12.31	Angola	5.00	2.42	2.42
1995.12.31	Angola	5.67	4.00	3.58
1996.12.31	Angola	7.83	3.67	4.00
1997.12.31	Angola	8.92	2.00	4.42
1998.12.31	Angola	10.25	2.00	4.00
1999.12.31	Angola	10.58	2.00	2.17
2000.12.31	Angola	11.00	2.00	2.00
2001.12.31	Angola	10.75	2.67	6.13
2002.12.31	Angola	9.63	2.58	8.08
2003.12.31	Angola	9.75	2.00	8.38
2004.12.31	Angola	10.00	3.75	8.00
2005.12.31	Angola	10.00	3.00	7.88
2006.12.31	Angola	9.63	2.00	7.88
2007.12.31	Angola	9.50	2.00	8.00
2008.12.31	Angola	9.83	2.21	8.00
2009.12.31	Angola	10.50	3.00	8.00
2010.12.31	Angola	10.25	3.00	8.00
2011.12.31	Angola	8.75	3.00	7.83
2012.12.31	Angola	8.17	3.00	7.50
2013.12.31	Angola	8.38	3.00	7.92
2014.12.31	Angola	7.08	3.00	7.83
2015.12.31	Angola	6.92	3.00	6.75
2016.12.31	Angola	6.54	3.00	6.50
2017.12.31	Angola	6.83	2.88	6.13
1984.12.31	Bahrain	5.00	6.00	6.83
1985.12.31	Bahrain	5.00	6.00	6.58
1986.12.31	Bahrain	5.00	6.00	6.00
1987.12.31	Bahrain	5.00	6.00	6.00
1988.12.31	Bahrain	5.17	5.67	5.42
1989.12.31	Bahrain	6.00	4.00	5.00

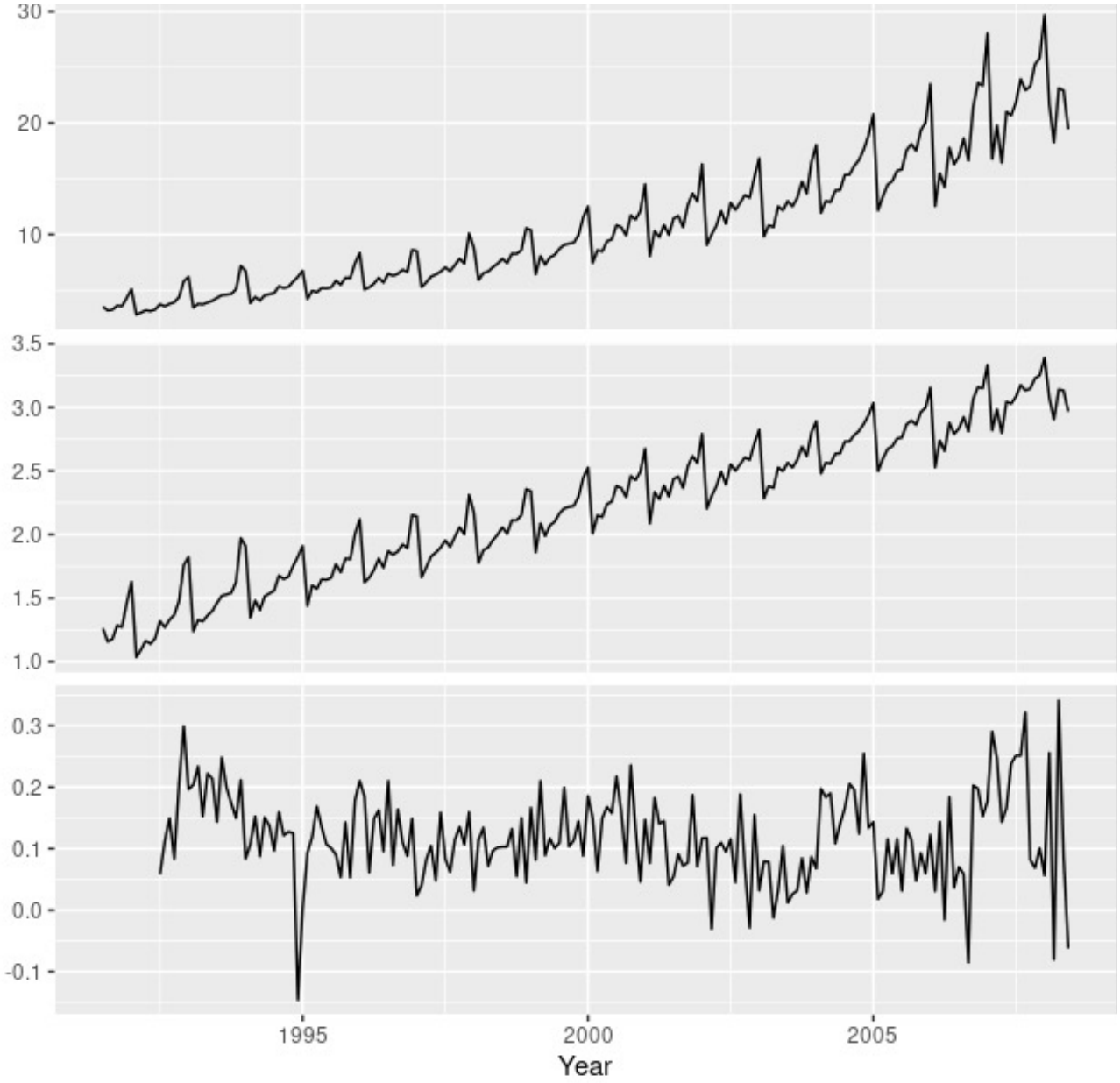
Stationarity in Time Series



Differencing

$$y_t - y_{t-1} = c + \varepsilon_t \quad \text{or} \quad y_t = c + y_{t-1} + \varepsilon_t \, .$$

The value of c is the average of the changes between consecutive observations. If c is positive, then the average change is an increase in the value of yt. Thus, yt will tend to drift upwards. However, if c is negative, yt will tend to drift downwards.



Differencing

$$y_t - y_{t-1} = c + \varepsilon_t \quad \text{or} \quad y_t = c + y_{t-1} + \varepsilon_t \,.$$

The process of subtracting one observation from other

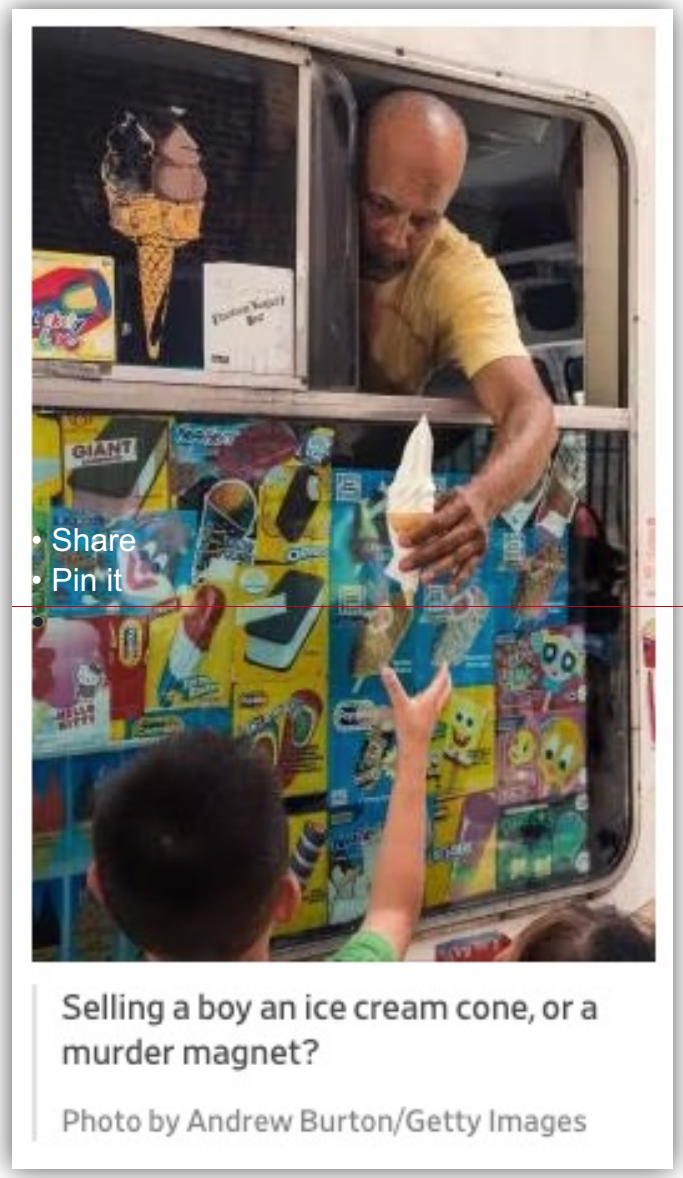
Used for transforming non-stationary data into stationary data

X = [5, 4, 6, 7, 9, 12] **So What?**

What should be the values of X after the 1 – lag differencing?

X^* = [1, -2, -1, -2, -3]

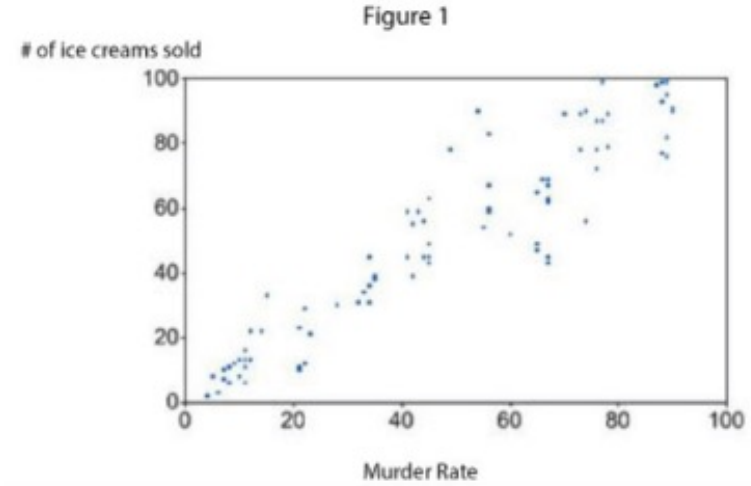
Linear Regression (OLS)

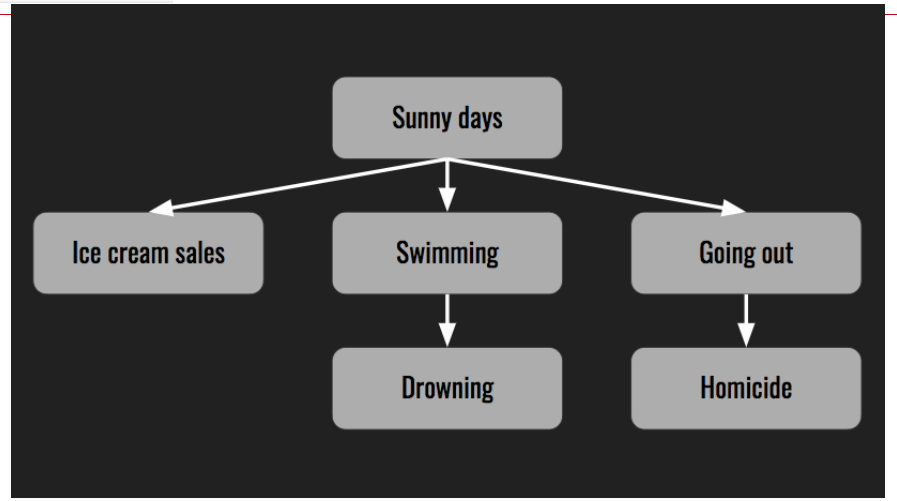


Selling a boy an ice cream cone, or a murder magnet?

Photo by Andrew Burton/Getty Images

When Ice Cream Sales Rise, So Do Homicides. Coincidence, or Will Your Next Cone Murder You? [1]

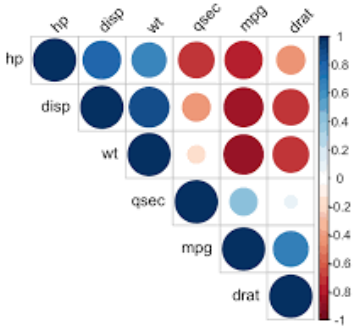




[When Ice Cream Sales Rise, So Do Homicides. Coincidence, or Will Your Next Cone Murder You?]

Linear Regression (OLS)

Create a scatter plot



Create a correlation matrix

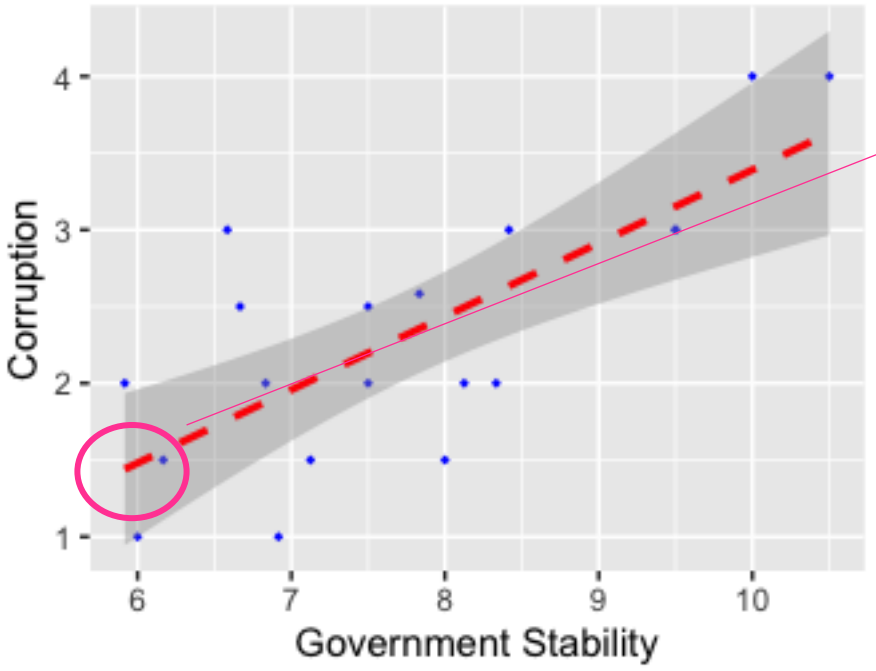
```
> describe(mtcars)
      vars  n   mean     sd median trimmed    mad   min    max  range  skew kurtosis   se
mpg      1 32  20.09   6.03  19.20   19.70   5.41 10.40  33.90  23.50  0.61    -0.37 1.07
cyl      2 32   6.19   1.79   6.00    6.23   2.97  4.00   8.00   4.00 -0.17    -1.76 0.32
disp     3 32 230.72 123.94 196.30  222.52 140.48 71.10 472.00 400.90  0.38    -1.21 21.91
hp       4 32 146.69  68.56 123.00  141.19  77.10 52.00 335.00 283.00  0.73    -0.14 12.12
drat     5 32   3.60   0.53   3.70    3.58   0.70  2.76   4.93   2.17  0.27    -0.71 0.09
wt       6 32   3.22   0.98   3.33    3.15   0.77  1.51   5.42   3.91  0.42    -0.02 0.17
qsec     7 32  17.85   1.79  17.71   17.83   1.42 14.50  22.90   8.40  0.37     0.34 0.32
vs       8 32   0.44   0.50   0.00    0.42   0.00  0.00   1.00   1.00  0.24    -2.00 0.09
am       9 32   0.41   0.50   0.00    0.38   0.00  0.00   1.00   1.00  0.36    -1.92 0.09
gear    10 32   3.69   0.74   4.00    3.62   1.48  3.00   5.00   2.00  0.53    -1.07 0.13
carb    11 32   2.81   1.62   2.00    2.65   1.48  1.00   8.00   7.00  1.05     1.26 0.29
```

Descriptive statistics

```
> summary(OLS02)

Call:
lm(formula = Corruption ~ `Bureaucracy Quality`, data = CS.data02)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2076 -0.4576  0.1490  0.4358  1.4358

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.1377     0.4531   2.511   0.0224 *
`Bureaucracy Quality`  0.7132     0.2534   2.815   0.0119 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7398 on 17 degrees of freedom
Multiple R-squared:  0.3179,    Adjusted R-squared:  0.2778
F-statistic: 7.924 on 1 and 17 DF,  p-value: 0.01193
```

Regression

Linear Regression (OLS)



Constant

Independent variable (IV)

$$Y_i = \beta_0 + \beta_1 X_i$$

Dependent Variable (DV)

Coefficient

Cording Practice: OLS

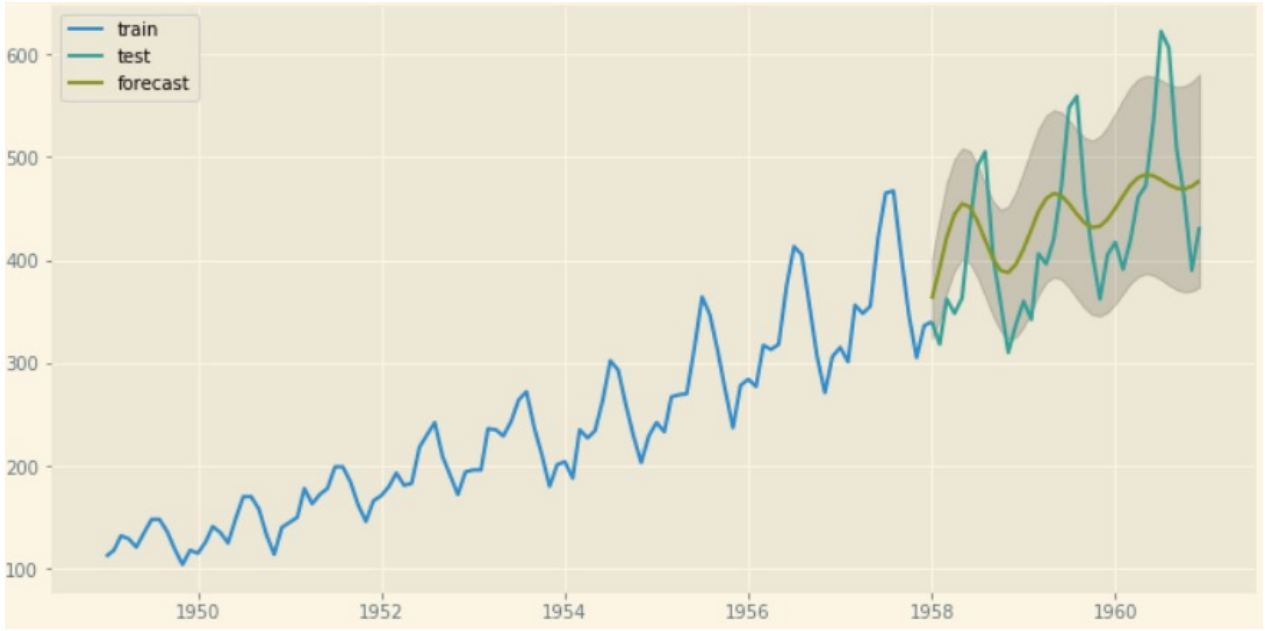
ARIMA equation

If d=0: $y_t = Y_t$

If d=1: $y_t = Y_t - Y_{t-1}$

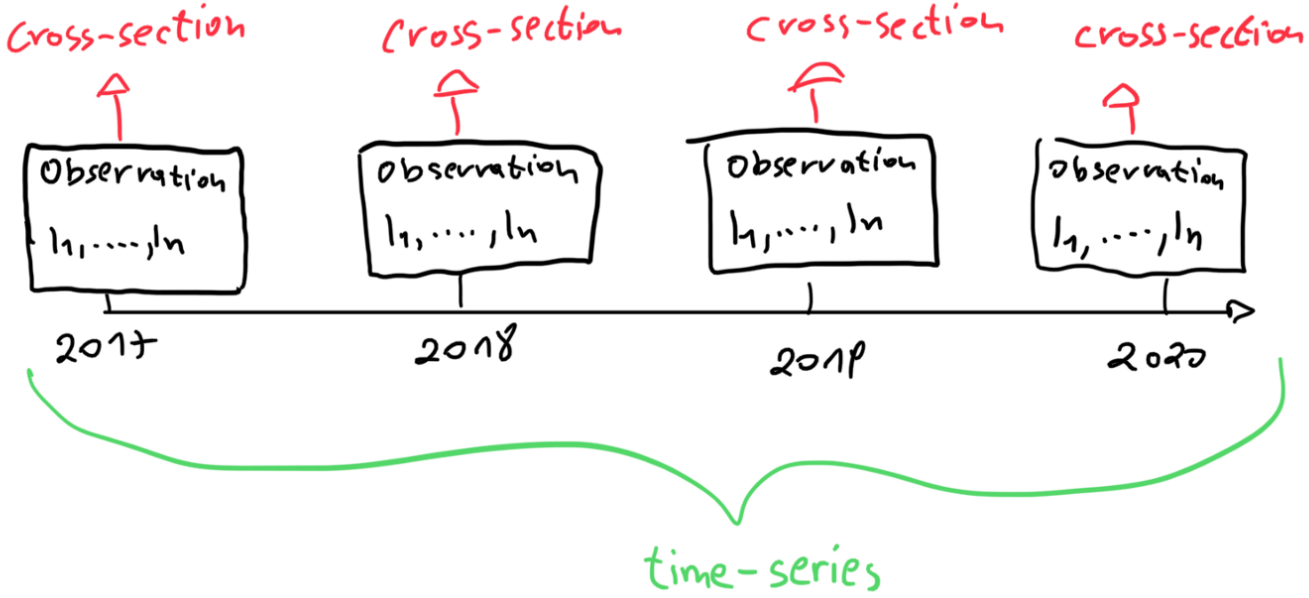
If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$



Cording Practice: Time Series

Panel Analysis



Panel Data

Combination of time series and cross-section observation

Date	County	Government Stability	Socioeconomic Conditions	Investment Profile
1988.12.31	Angola	6.00	6.00	7.00
1989.12.31	Angola	6.00	6.42	7.00
1990.12.31	Angola	6.00	6.17	7.00
1991.12.31	Angola	6.00	6.00	6.50
1992.12.31	Angola	4.17	4.50	4.92
1993.12.31	Angola	4.58	2.17	3.33
1994.12.31	Angola	5.00	2.42	2.42
1995.12.31	Angola	5.67	4.00	3.58
1996.12.31	Angola	7.83	3.67	4.00
1997.12.31	Angola	8.92	2.00	4.42
1998.12.31	Angola	10.25	2.00	4.00
1999.12.31	Angola	10.58	2.00	2.17
2000.12.31	Angola	11.00	2.00	2.00
2001.12.31	Angola	10.75	2.67	6.13
2002.12.31	Angola	9.63	2.58	8.08
2003.12.31	Angola	9.75	2.00	8.38
2004.12.31	Angola	10.00	3.75	8.00
2005.12.31	Angola	10.00	3.00	7.88
2006.12.31	Angola	9.63	2.00	7.88
2007.12.31	Angola	9.50	2.00	8.00
2008.12.31	Angola	9.83	2.21	8.00
2009.12.31	Angola	10.50	3.00	8.00
2010.12.31	Angola	10.25	3.00	8.00
2011.12.31	Angola	8.75	3.00	7.83
2012.12.31	Angola	8.17	3.00	7.50
2013.12.31	Angola	8.38	3.00	7.92
2014.12.31	Angola	7.08	3.00	7.83
2015.12.31	Angola	6.92	3.00	6.75
2016.12.31	Angola	6.54	3.00	6.50
2017.12.31	Angola	6.83	2.88	6.13
1984.12.31	Bahrain	5.00	6.00	6.83
1985.12.31	Bahrain	5.00	6.00	6.58
1986.12.31	Bahrain	5.00	6.00	6.00
1987.12.31	Bahrain	5.00	6.00	6.00
1988.12.31	Bahrain	5.17	5.67	5.42
1989.12.31	Bahrain	6.00	4.00	5.00

Cording Practice: Time Series

Next Week

Please bring your laptop.

Week 6: Web Page Data Scraping

- How do we get online data from news media and social media?
- Learning how to scrape or crawl via using the R package 'rvest', scraping (websites and tables), a nd using loops and functions