ECE1724H S2: Empirical Software Engineering

Time-Series Design and Analysis

The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

about final report

The page number is a soft requirement.
Story need to be cohesive and complete

Your feedback is important to me!

UNIVERSITY OF TORONTO

APSC Winter 2021 Grad for Sp Topics in Software Engineer ECE1724H-S-LEC9101

Medium Online

Timing Scheduled

- Start Date 2021-03-22 00:00
- End Date 2021-04-04 23:59

Response Rate

	Responded	Invited	% Rate
Students	1	10	10.00%

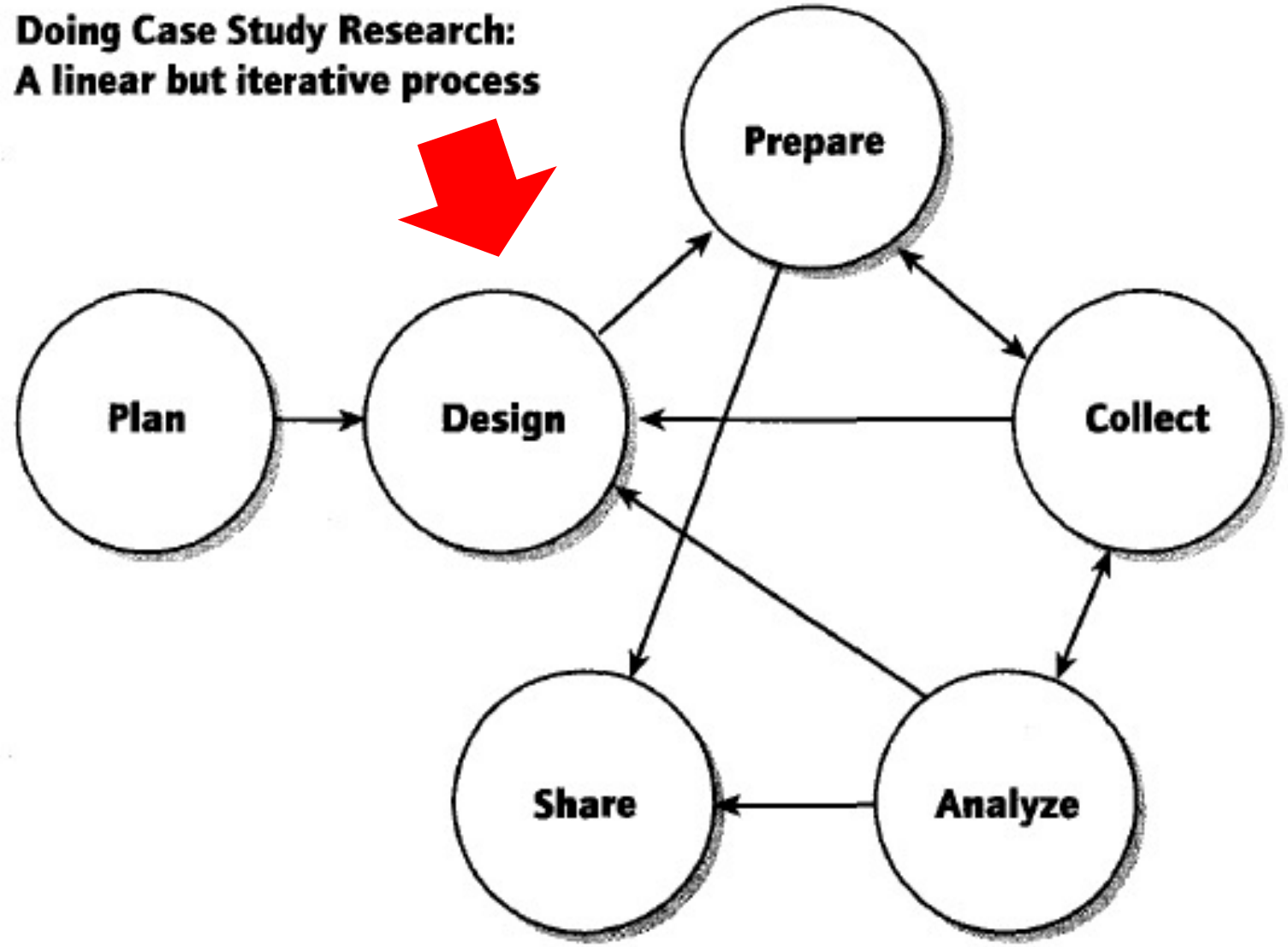

CASE STUDY

	(1)	(2)	(3)
METHOD	Form of Research Question	Requires Control of Behavioral Events?	Focuses on Contemporary Events?
Experiment	how, why?	yes	yes
Survey	who, what, where, how many, how much?	no	yes
Archival Analysis	who, what, where, how many, how much?	no	yes/no
History	how, why?	no	no
Case Study	how, why?	no	yes

Figure 1.1 Relevant Situations for Different Research Methods SOURCE: COSMOS Corporation (1983)

When should you use a case study?

- When you can't control the variables
- When there are many more variables than data points
- When you cannot separate phenomena from context
  - Phenomena that don't occur in a lab setting
  - E.g. large scale, complex software projects
  - Effects can be wide-ranging.
  - Effects can take a long time to appear (weeks, months, years!)
- When the context is important
  - E.g. When you need to know how context affects the phenomena
- When you need to know whether your theory applies to a specific real world setting


Doing Case Study Research:
A linear but iterative process
Prepare
Plan
Design
Collect
Share
Analyze

How can I evaluate a case study?

Same criteria as for other empirical research:

- Construct Validity
  - Concepts being studied are operationalized and measured correctly
- Internal Validity
  - Establish a causal relationship and distinguish spurious relationships
- External Validity
  - Establish the domain to which a study's findings can be generalized
- Empirical Reliability
  - Demonstrate that the study can be repeated with the same results

TESTS	Case Study Tactic	Phase of research in which tactic occurs
Construct validity	♦ use multiple sources of evidence	data collection
	♦ establish chain of evidence	data collection
	♦ have key informants review draft case study report	composition
Internal validity	♦ do pattern matching	data analysis
	♦ do explanation building	data analysis
	♦ address rival explanations	data analysis
	♦ use logic models	data analysis
External validity	♦ use theory in single-case studies	research design
	♦ use replication logic in multiple-case studies	research design
Reliability	♦ use case study protocol	data collection
	♦ develop case study database	data collection

Figure 2.3 Case Study Tactics for Four Design Tests

Research Ethics

- Reasons to take ethics seriously:
  - Funding depends on it
  - Relationship with research subjects/organisations depends on it
  - Legal issues (e.g. liability for harm to subjects/organisations)
  - Compliance with privacy and data protection laws
  - …and it's the right thing to do!
- Institutional Review Boards (IRB)
  - Approval usually needed for all studies involving human subjects
  - Every IRB has it's own rules…
    - A study approved at one university may be disallowed at another!
    - Design of the study might have to be altered
  - Institutional research funding may depend on this process!
  - Note: guidelines from other fields may not apply to Software Engineering
    - E.g. use/ownership of source code
    - E.g. effect of process improvement on participants

What is the IRB

Institutional review boards (IRBs)

Also called research ethics committees, IRB's provide protection for human research participants through review of the ethical acceptability of proposals for human research.

What does it consist of?

5+ sufficiently qualified members

Members with diverse experience and expertise to safeguard subjects' rights and welfare and to evaluate research acceptability.

At least one member knowledgeable about any regularly researched vulnerable groups.

Can include specialized experts in relation to particular cases.

According to the NIH

StudyFind

Why do we need the IRB?

StudyFind

1 Functions and Operations

Write procedures for researchers to report issues

Review and amend study procedures

Approval of procedures by majority vote

2 Review of studies

Require informed consent and documentation

Approve, Modify, Disapprove Research

Review research annually

3 Authority

IRB must approve research to move forward. IRB can suspend or terminate research for serious harm or noncompliance.

4

Keep records of research proposals, meetings ,etc.

According to the NIH

Agenda for Today

- Paper reading presentation
- Time-series analysis

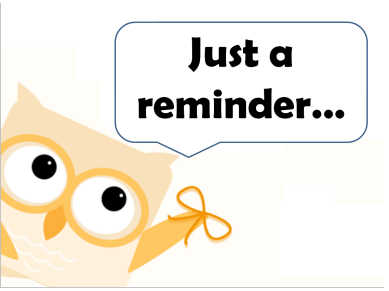
- - Murphy-Hill, Emerson, et al. "Do developers discover new tools on the toilet? download." *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 2019.
- - Amershi, Saleema, et al. "Software engineering for machine learning: A case study. download" *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 2019.

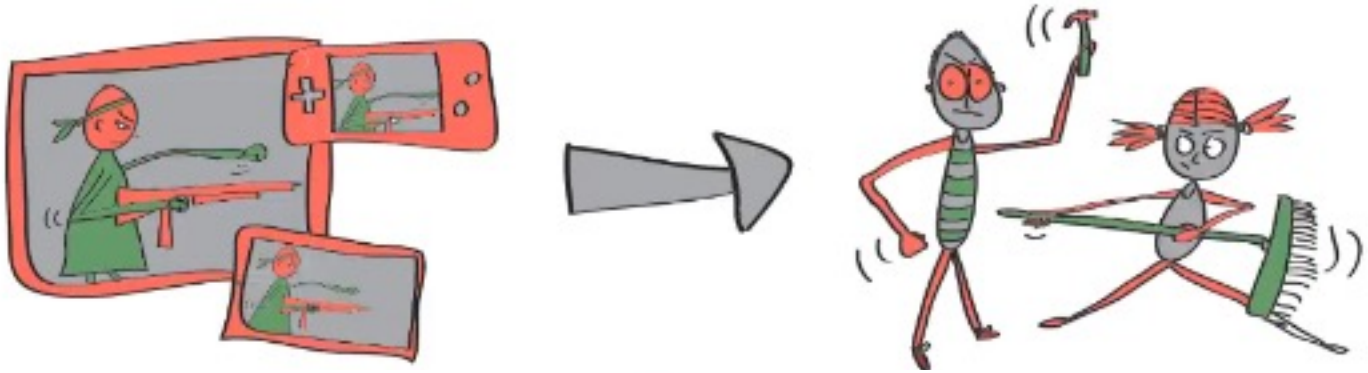
Agenda for Today

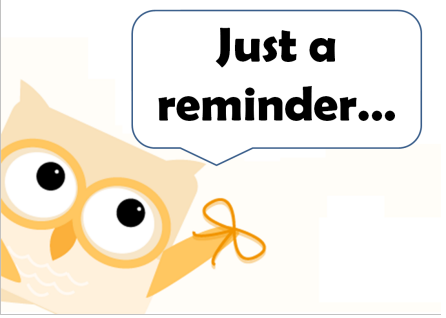
- Paper reading presentation
- Time-series analysis

The Vocabulary of Experiments

- **Experiment**: A study in which an intervention is deliberately introduced to observe its effects.
- **Randomized Experiment**: An experiment in which units are assigned to receive the treatment or an alternative condition by a random process such as the toss of a coin or a table of random numbers.
- **Quasi-Experiment**: An experiment in which units are not assigned to conditions randomly.

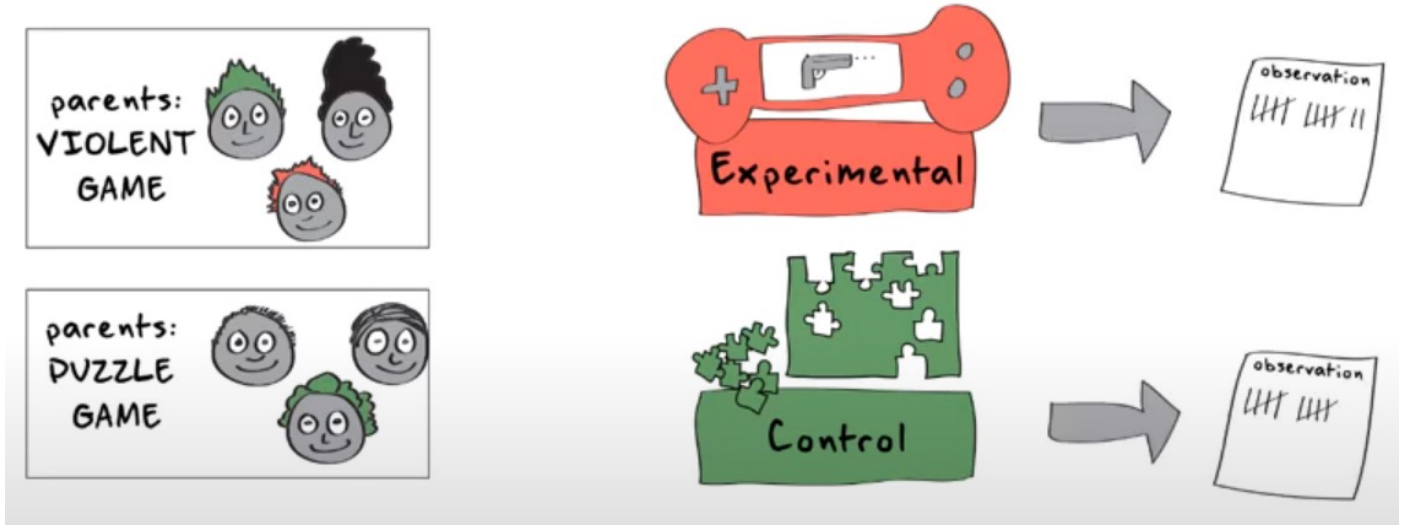





Just a
reminder...

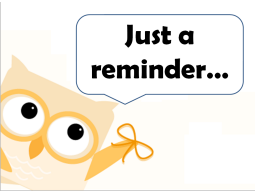
Quasi-experiments Design

- Static group comparison design



Group 1 $X_1 \rightarrow O_1$

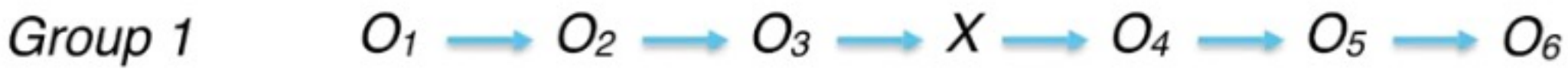
Group 2 $X_2 \rightarrow O_2$

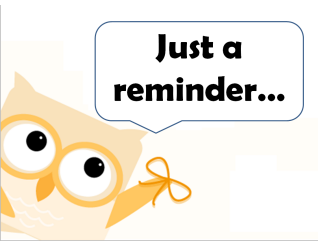


Quasi-experiments Design

- One-group Pretest-Posttest design

- Interrupted time series design

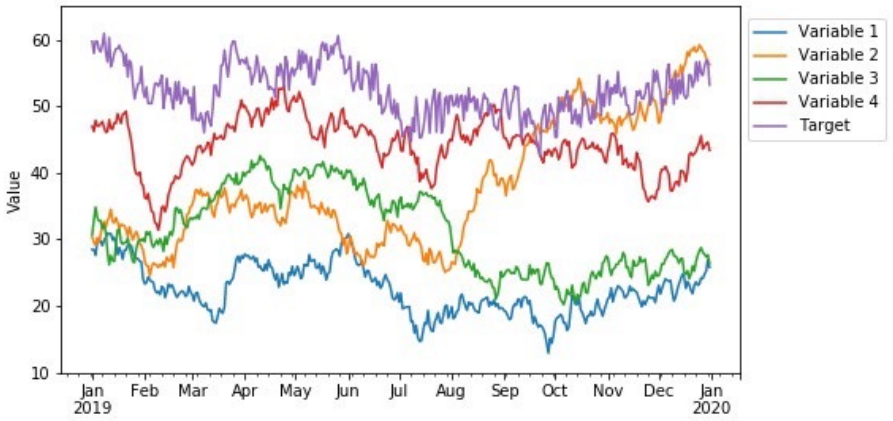


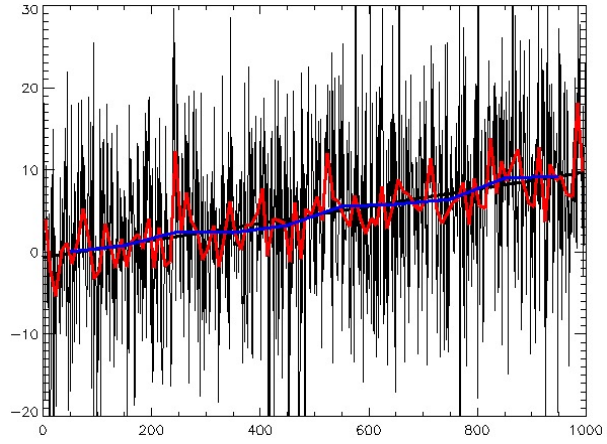


1982, the Arizona state legislature mandated severe penalties for driving while intoxicated

1976 – 1982 vs 1982 – 1984 → decrease in traffic fatalities

What is a time series?






I'VE FINALLY FOUND IT... AFTER 15 YEARS
THE SCROLL OF TRUTH!
RobotatertotComics
It's impossible to predict stock prices
NYEHHH
imgflip.com

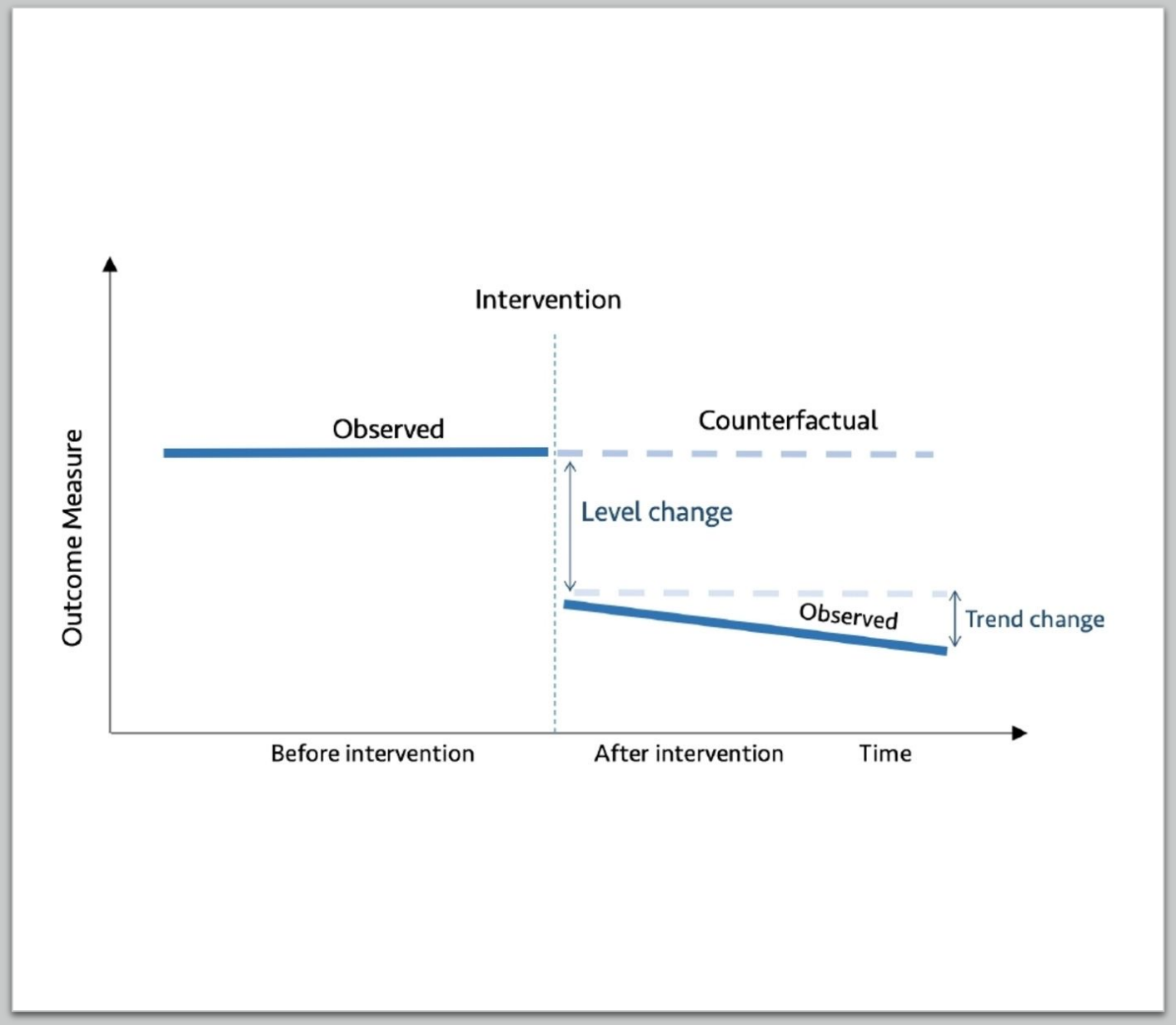
What is a time series?

- Time series refers to a large series of observations made on the same variable consecutively over time.
- The observations can be on the same units, or on different but similar units

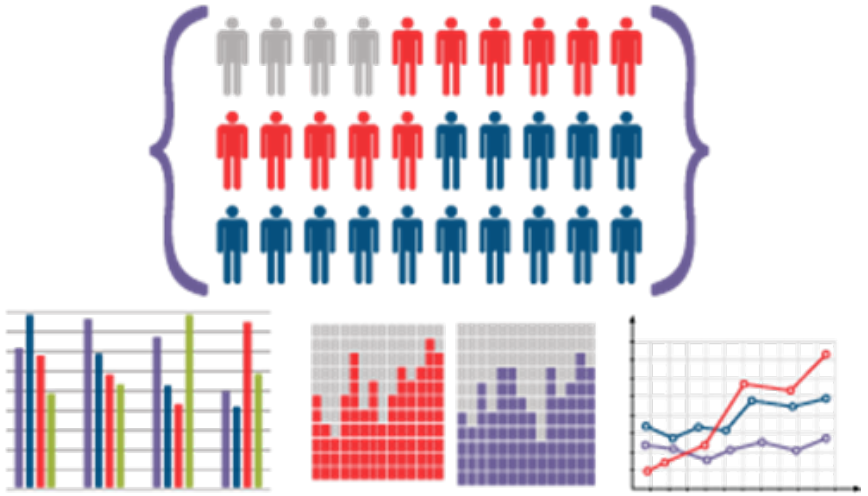
Interrupted Time Series

Causal hypothesis –

the observations after treatment will have a different slope or level from those before treatment.



Interrupted Time Series - Application scenarios

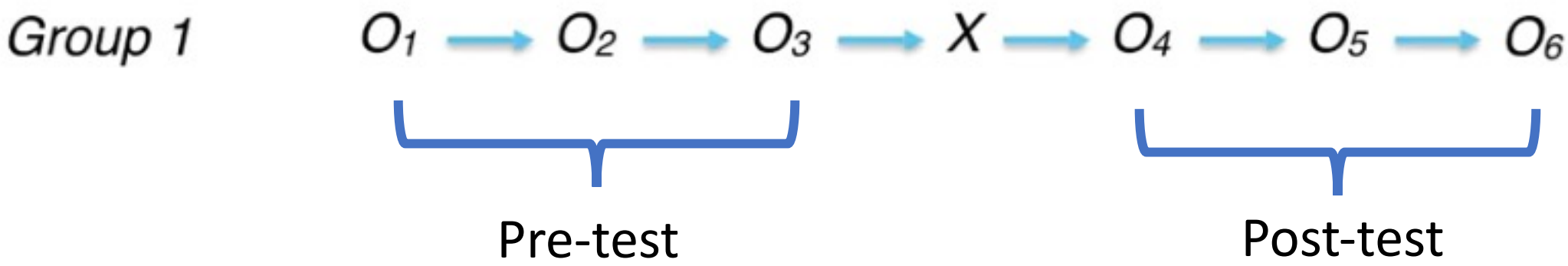


United States
CONSUMER PRODUCT
SAFETY COMMISSION

Interrupted Time Series

- The strongest quasi-experimental design to evaluate longitudinal effects of time-delimited interventions.
- How much did an intervention change an outcome of interest?
  - immediately and over time;
  - instantly or with delay;
  - transiently or long-term;
- Could factors other than the intervention explain the change?

Types of Effects



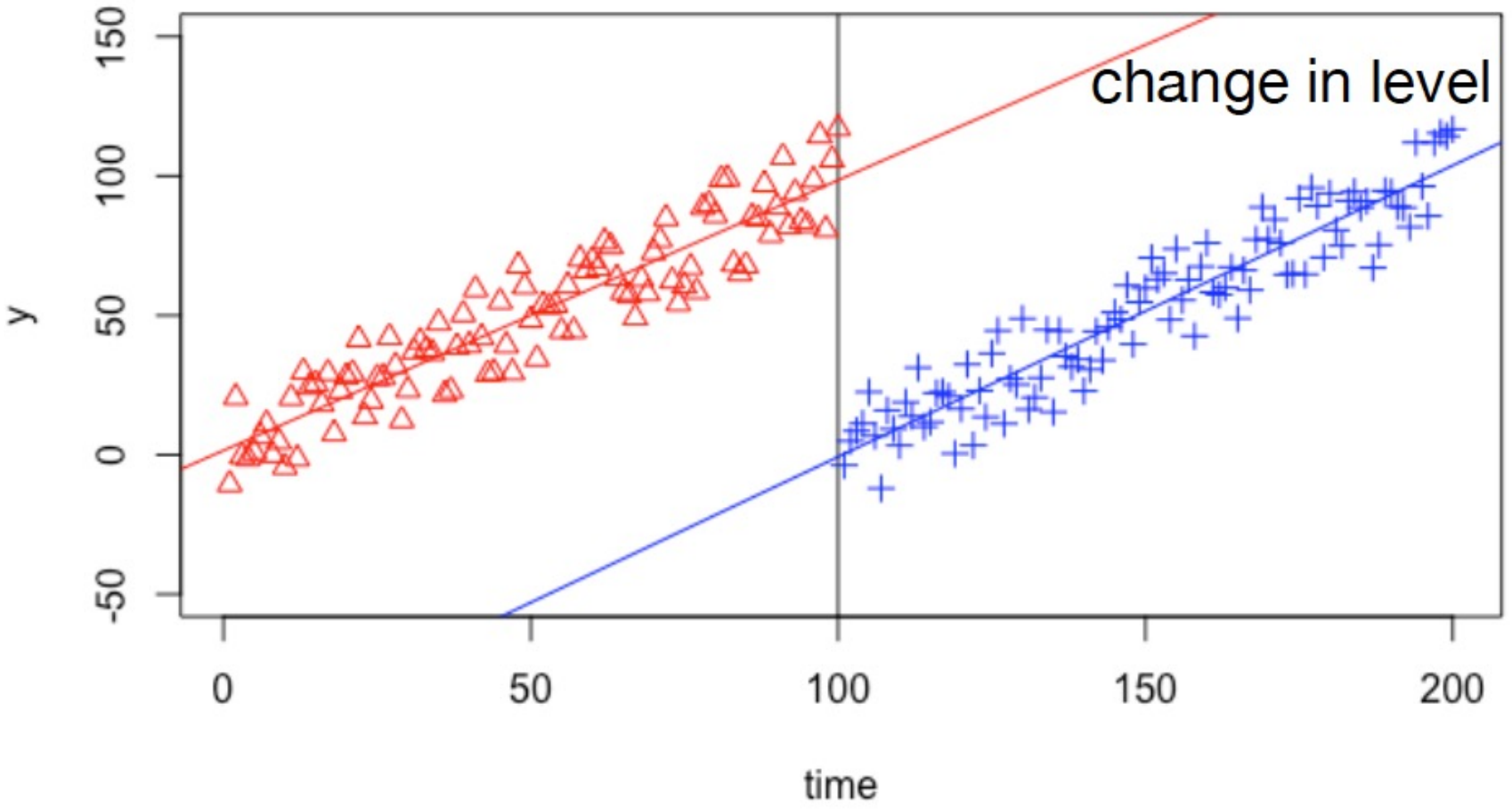
Causal hypothesis –

the observations after treatment will have a different slope or level from those before treatment.

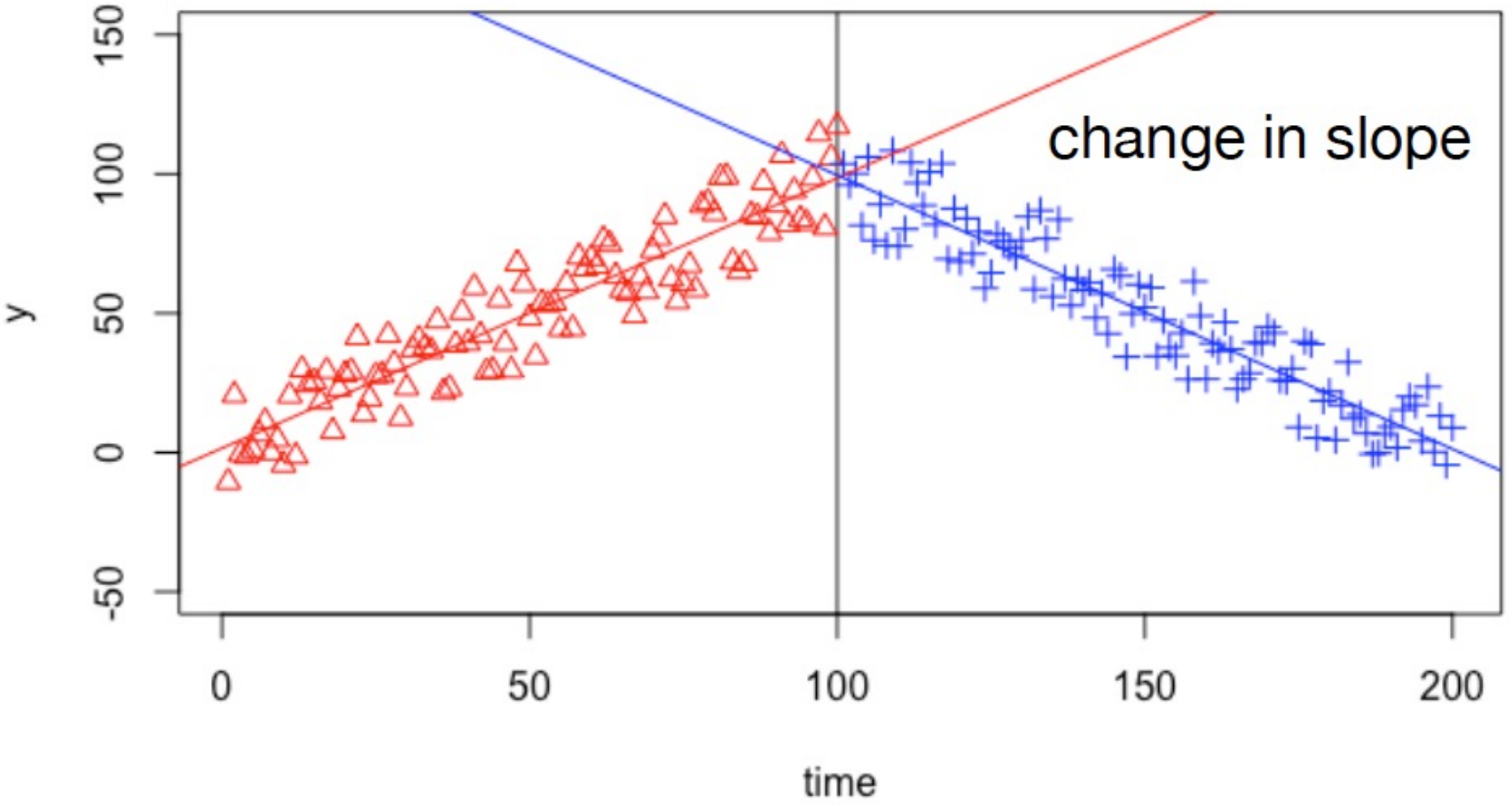
Types of Effects

- Three dimensions
  - form (level, slope, variance, cyclicity)
  - permanence (continuous or discontinuous)
  - immediacy (immediate or delayed)

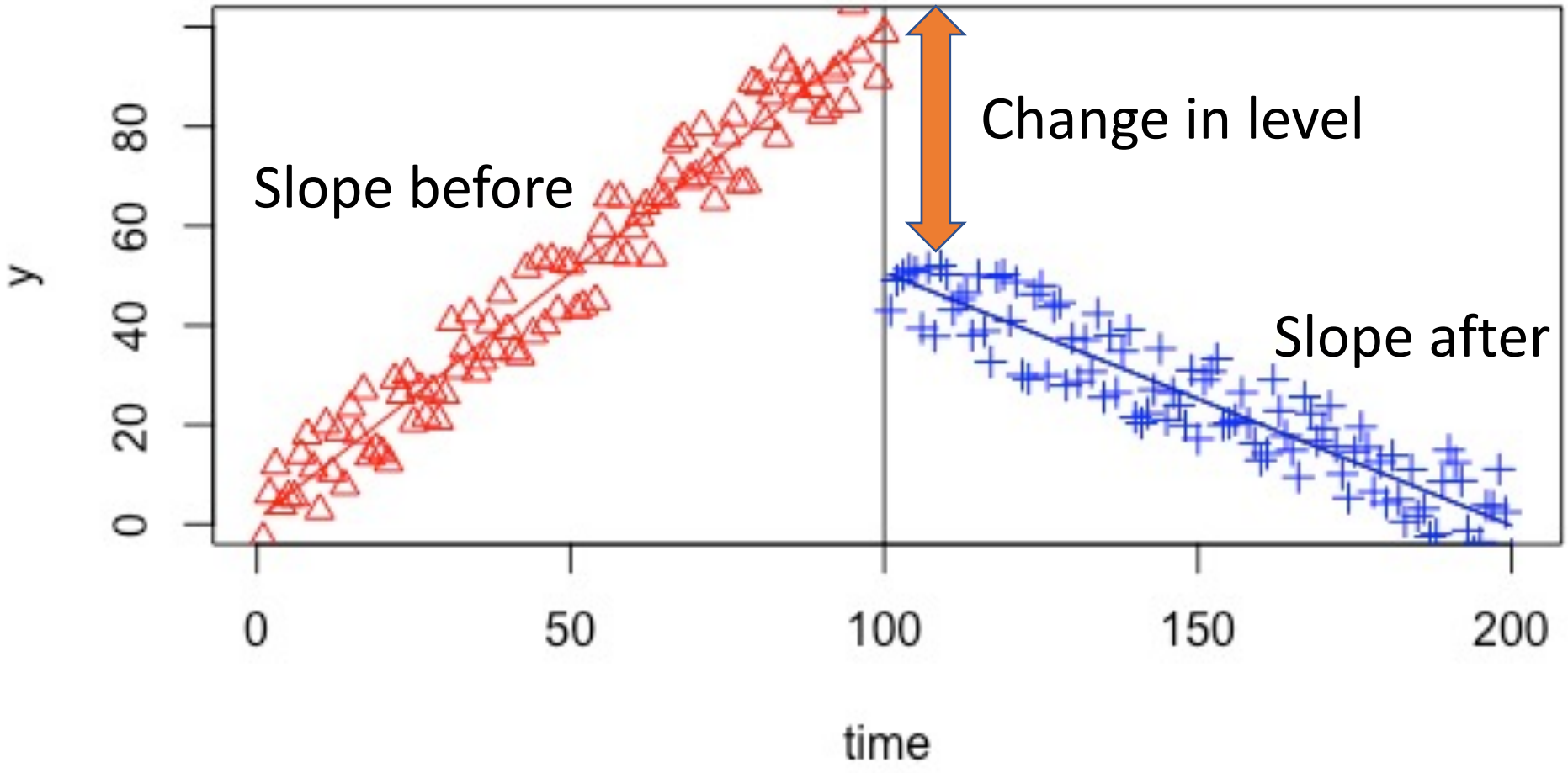
a change in level/intercept



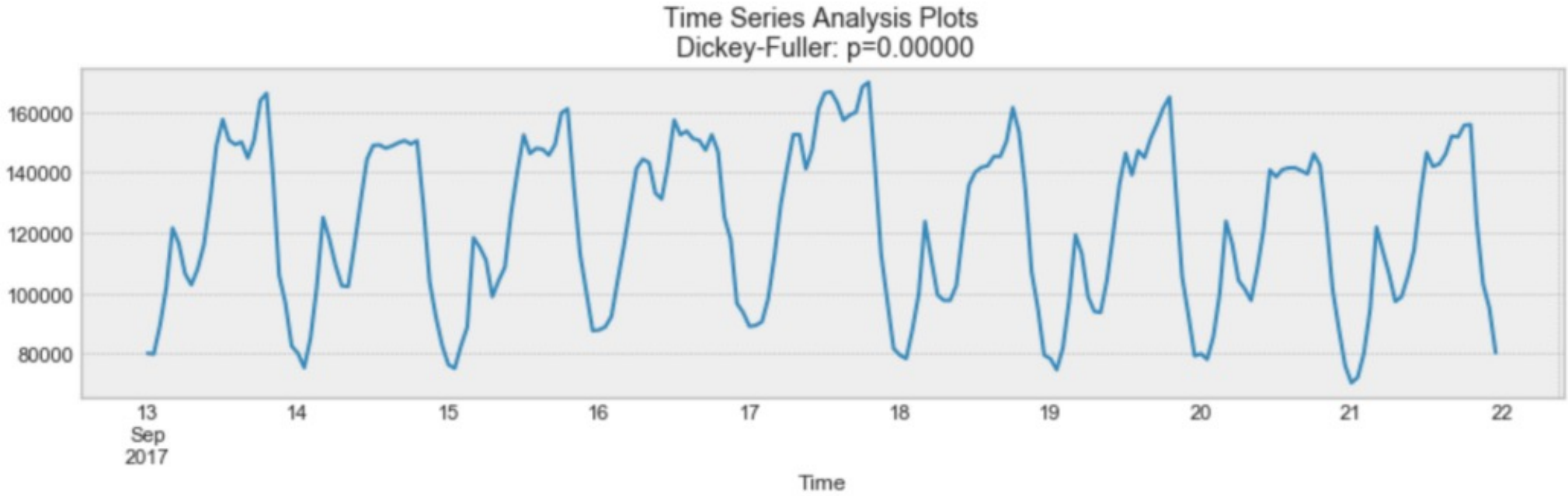
a change in slope/drift/trend



Types of Effects - form



Seasonality



Cyclicity

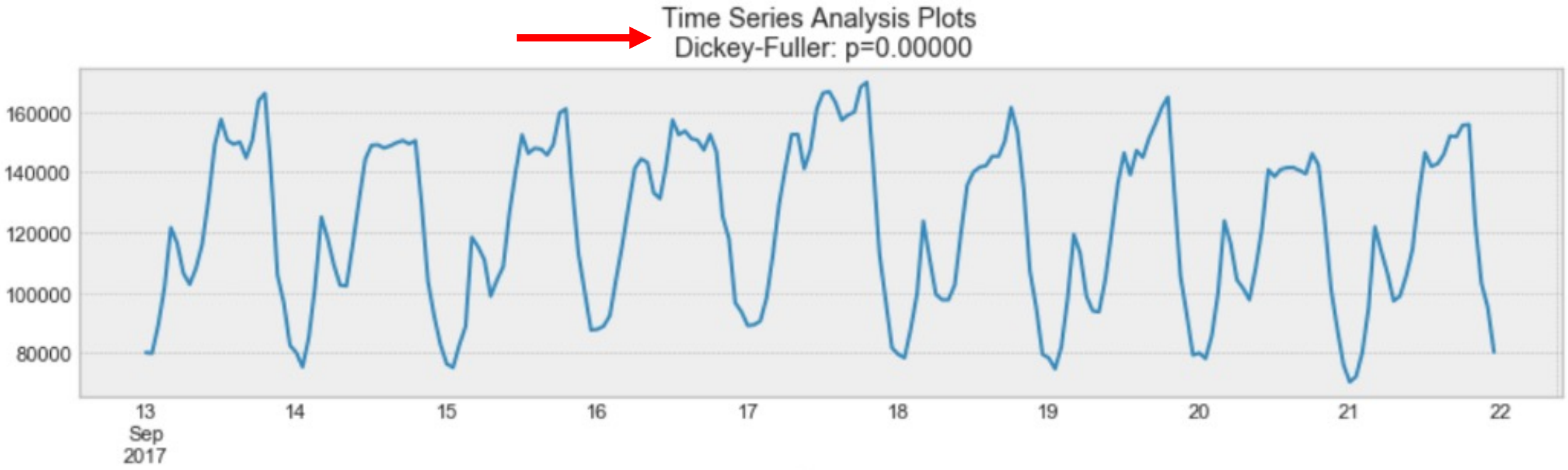
- a medium-term variation caused by circumstances that repeat in irregular intervals.
- E.G.

  A recession in the global economy like -

  1. The panic of 1785
  2. The 1815–1821 depression
  3. The great recession from 2007–2008

Stationarity

- The data has **constant mean and variance**

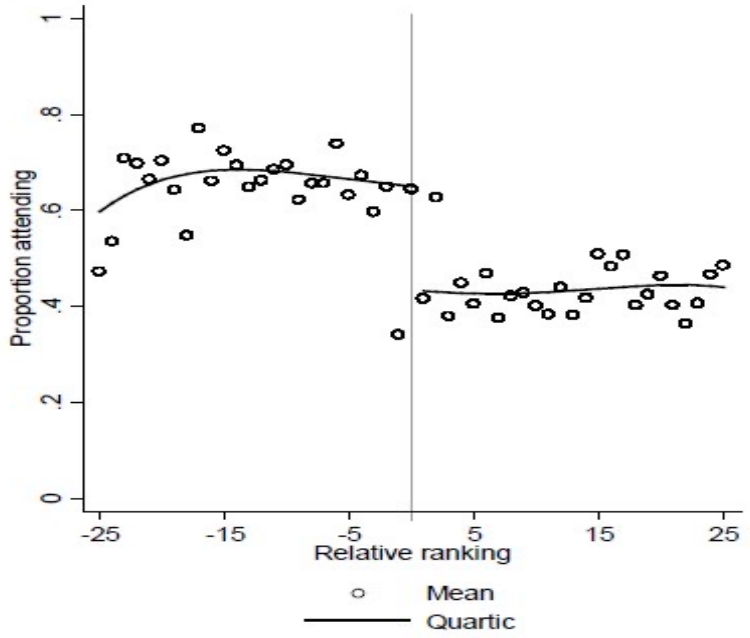


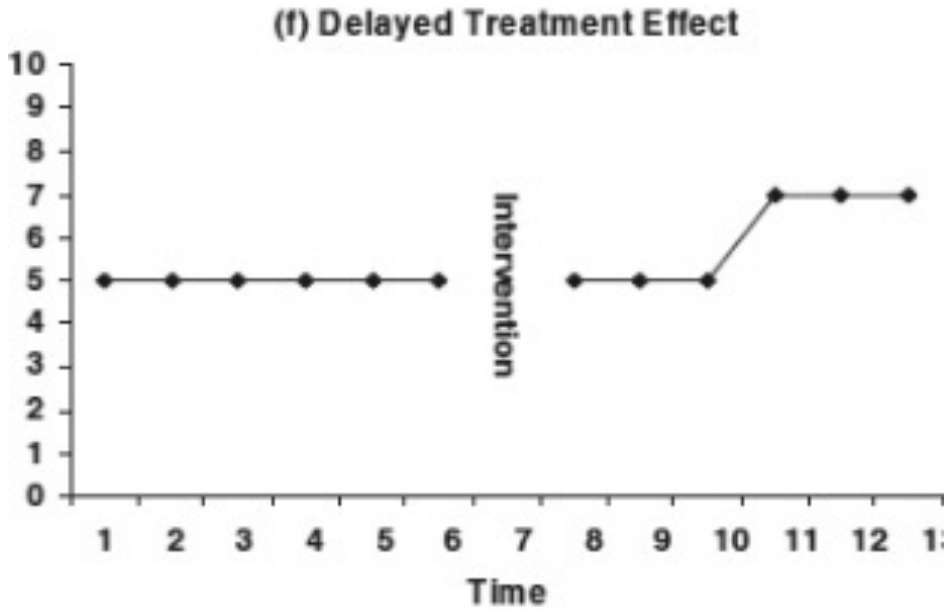
Non- stationary

- Image source — https://chartink.com/

Types of Effects – other dimensions

- Continuous & Discontinuous
- Immediate & delayed

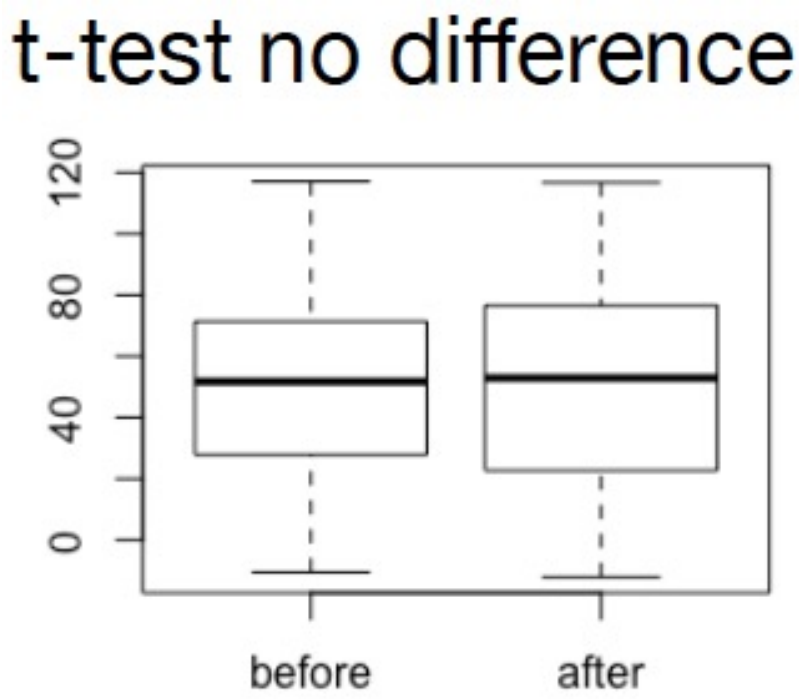


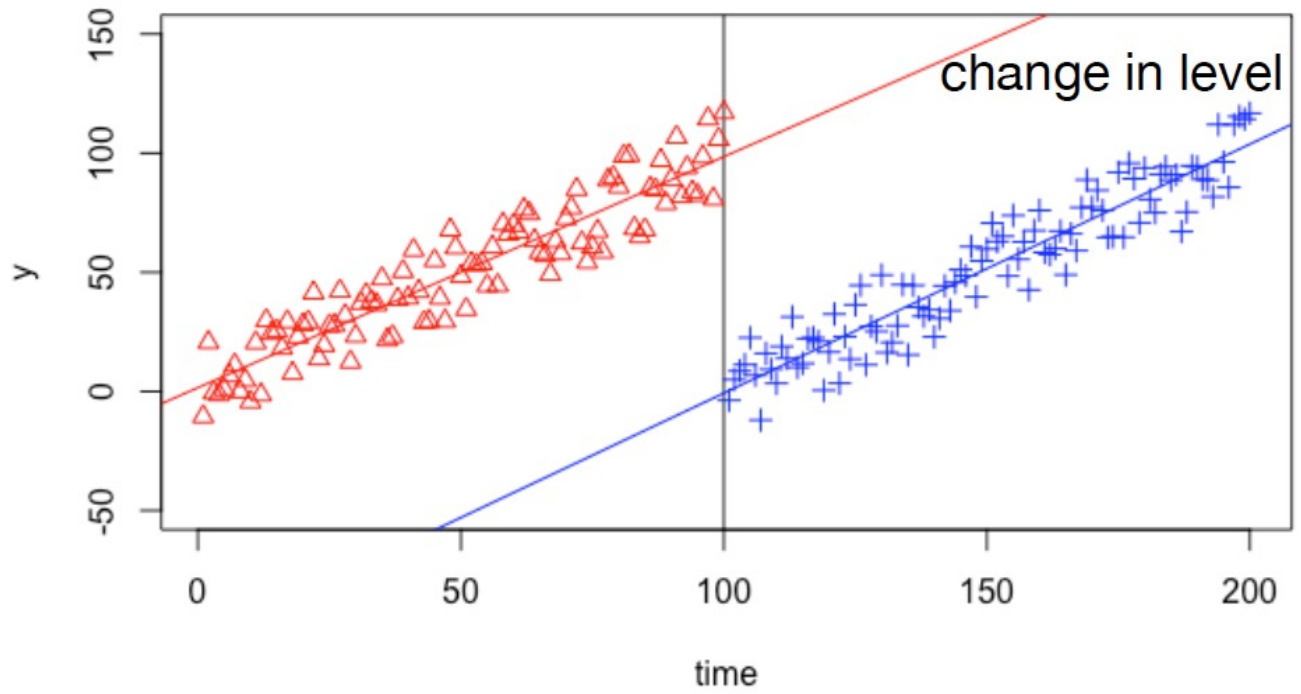


Types of Effects

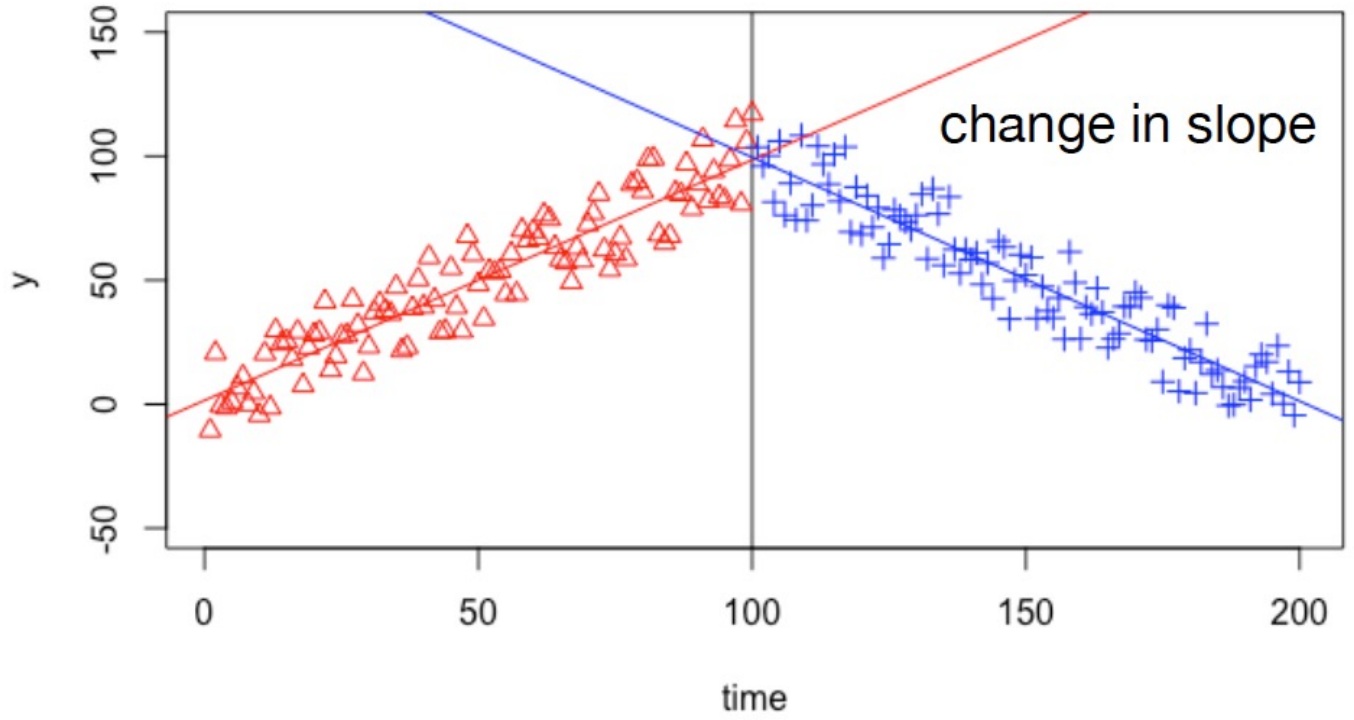
- Three dimensions
  - form (level, slope, variance, cyclicity)
  - permanence (continuous or discontinuous)
  - immediacy (immediate or delayed)

Ordinary statistics cannot be used





Ordinary statistics cannot be used



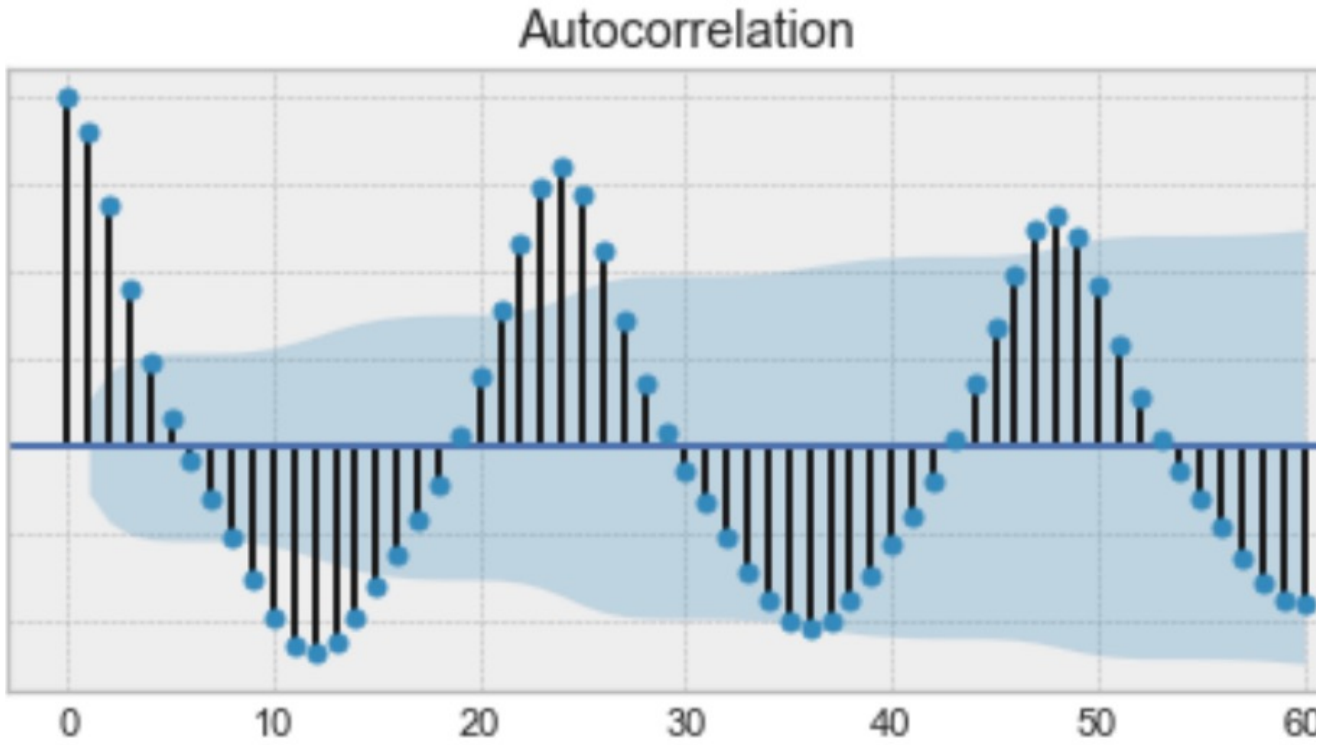
Autocorrelation

- serial correlation/serial dependence

X	Y	Y[-1]
1	0.397	0.157
2	0.157	-0.083
3	-0.083	-0.243
4	-0.243	-0.323
5	-0.323	-0.243
6	-0.243	-0.083
7	-0.083	0.077
8	0.077	0.347
9	0.347	

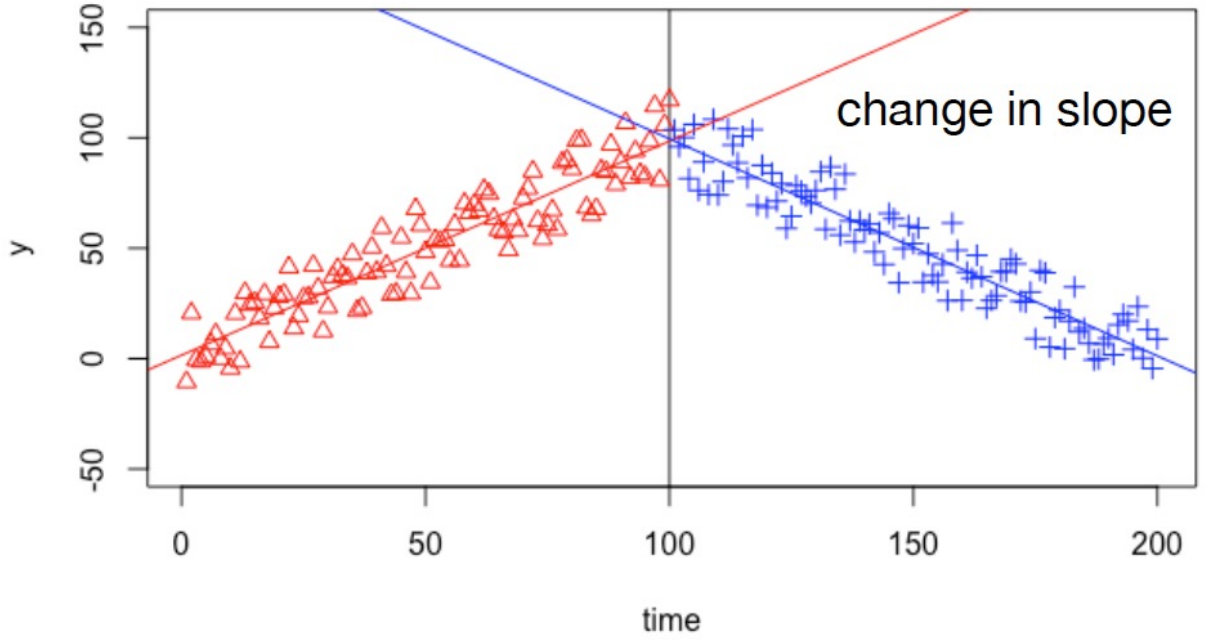
correlation = 0.64

Autocorrelation



- Autocorrelation is diagnosed using a *correlogram* (*ACF plot*) and can be tested using the *Durbin-Watson test*.
- Implication:
  - model is *misspecified*
  - *standard errors, p-values are misleading*

Takeaway: Ordinary statistics cannot be used for time-series analysis



EXAMPLE

Simple Interrupted Time Series

In March 1974, Cincinnati Bell began charging 20 cents per call to local directory assistance

Simple Interrupted Time Series

$O_1 \rightarrow O_2 \rightarrow O_3 \rightarrow X \rightarrow O_4 \rightarrow O_5 \rightarrow O_6$

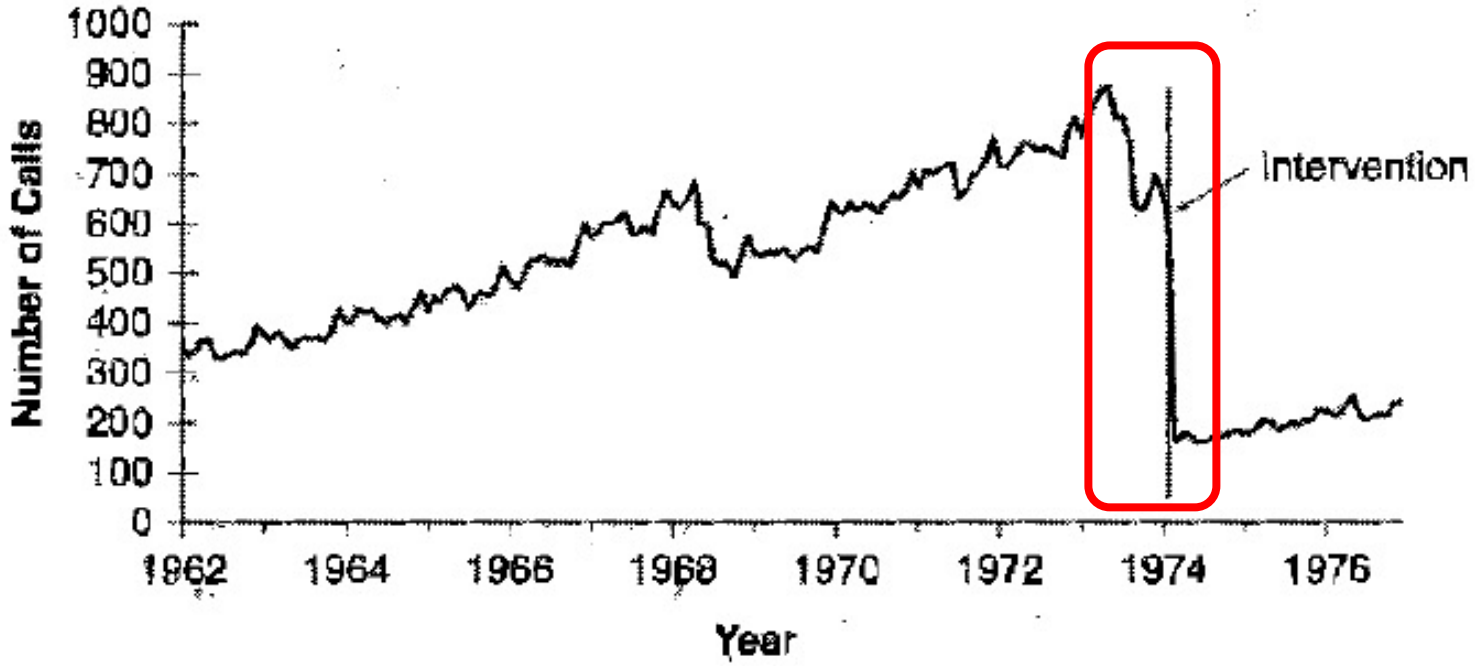


FIGURE 6.1 The effects of charging for directory assistance in Cincinnatii

In March 1974, Cincinnati Bell began charging 20 cents per call to local directory assistance

Rival Hypotheses?

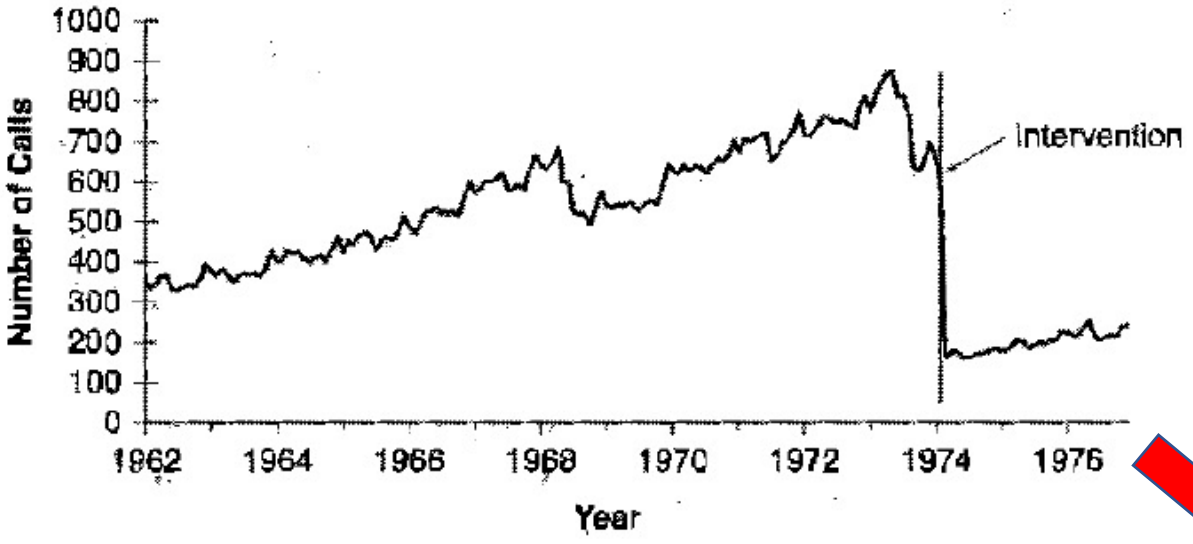
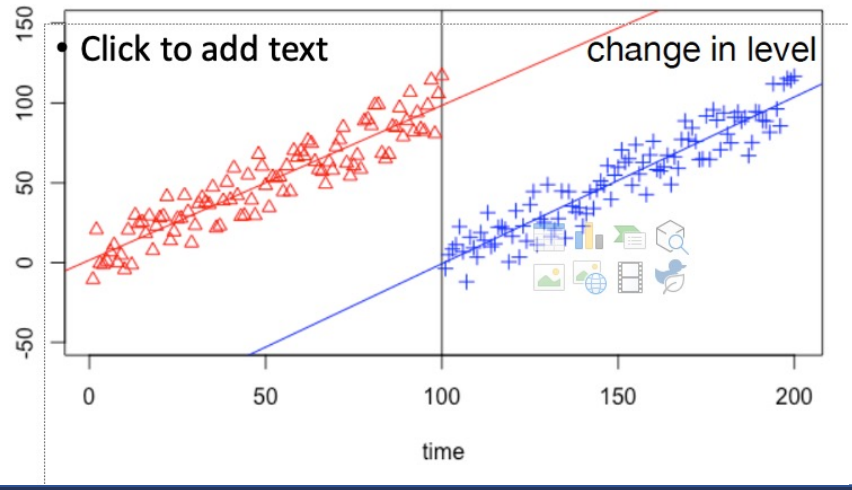


FIGURE 6.1 The effects of charging for directory assistance in Cincinnatii

Internal Validity?

- Regression?
- Selection?
- Attrition?

a change in level/intercept

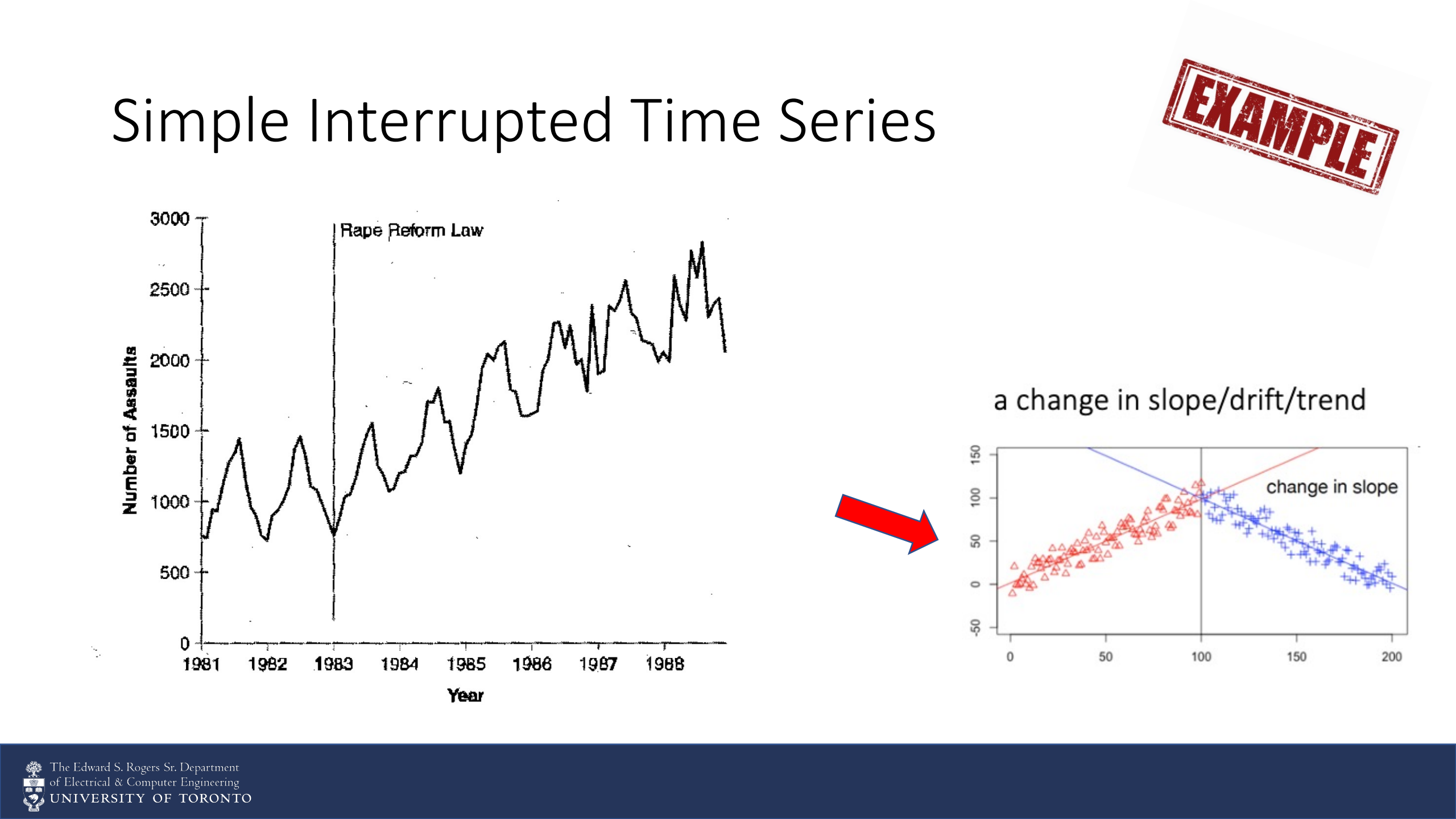


Simple Interrupted Time Series

- In 1983, Canada reformed its Criminal Code pertaining to sexual assault.

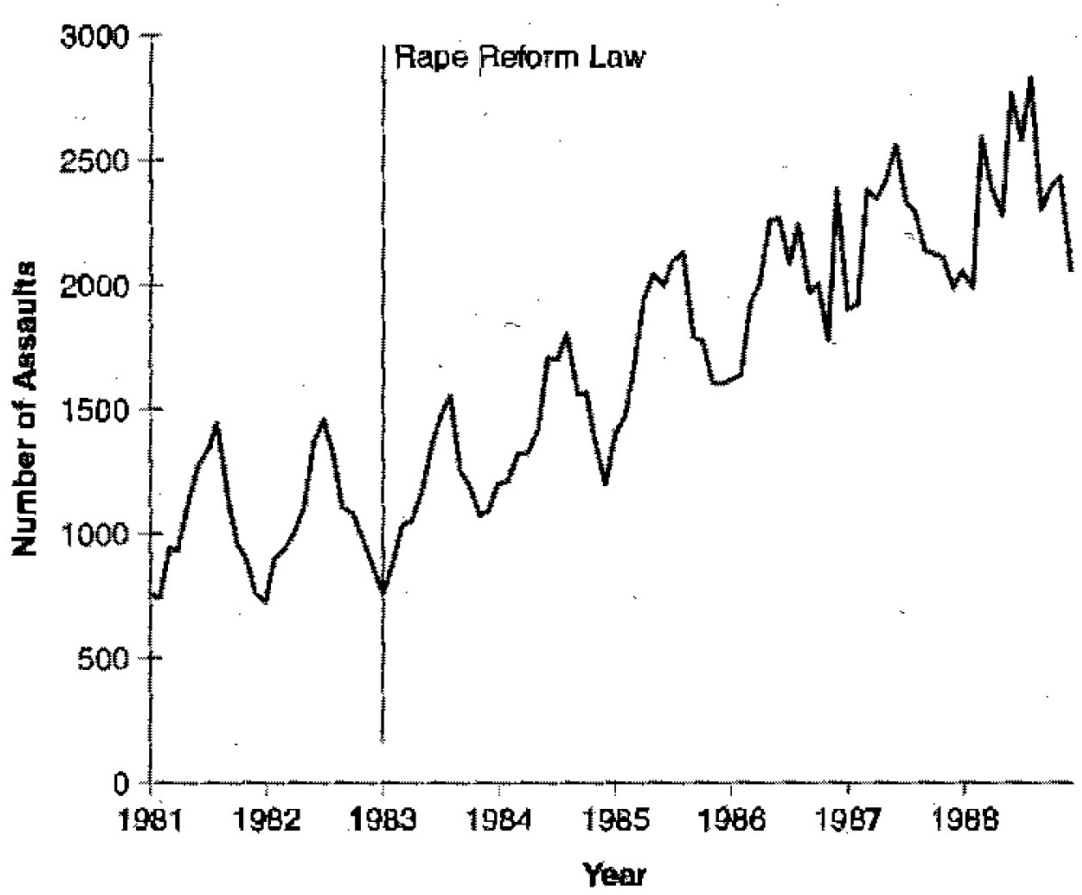
EXAMPLE

Simple Interrupted Time Series



a change in slope/drift/trend

Rival Hypotheses?



Construct validity?

reform legislation

or

reform legislation w/ great publicity

Simple Interrupted Time Series

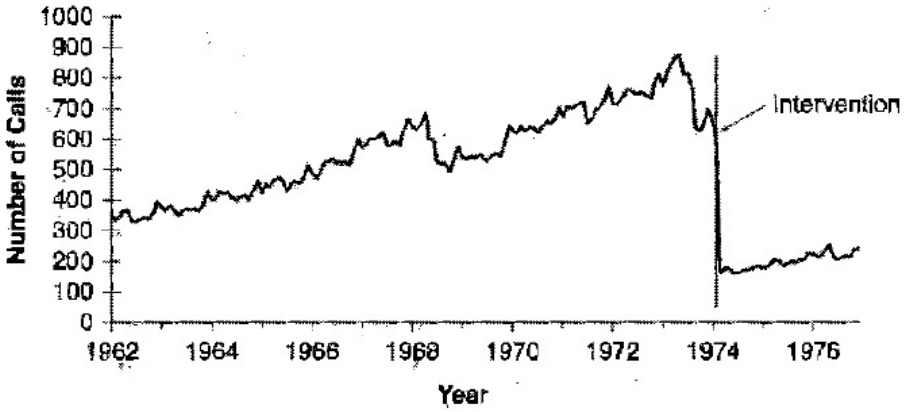
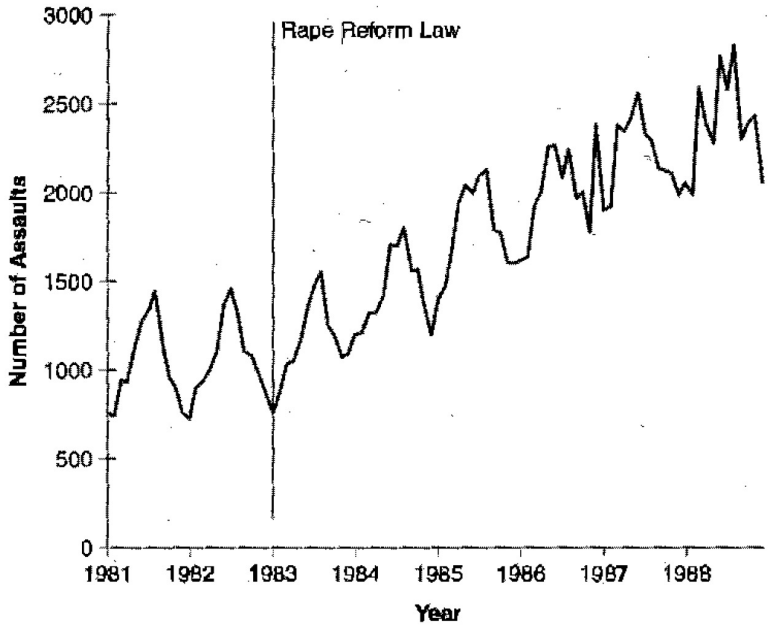


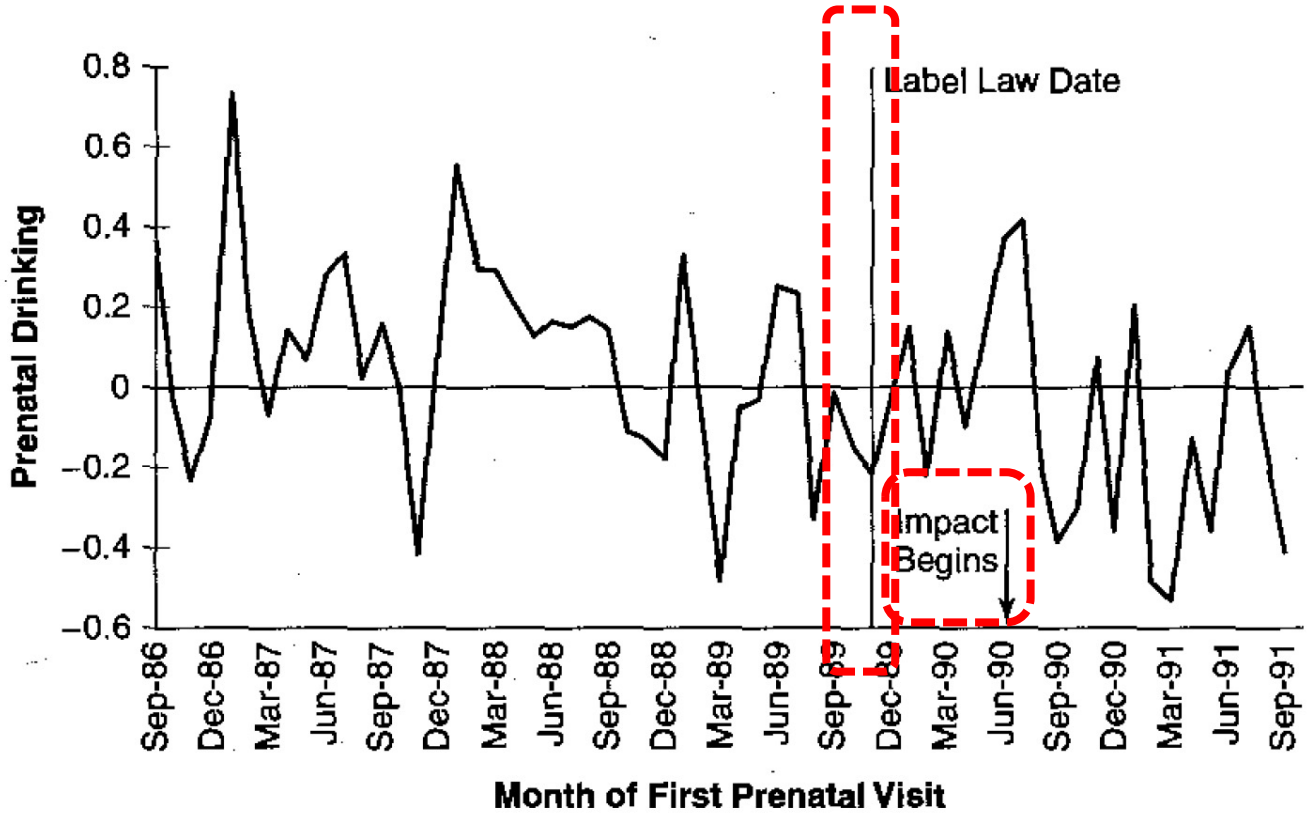
FIGURE 6.1 The effects of charging for directory assistance in Cincinnati



Simple Interrupted Time Series

Starting November 18,1989, a federal law required a warning label on all alcohol containers

Weak and Delayed Effects



Adding other design features to the basic interrupted time series

Adding a Nonequivalent No-Treatment Control Group Time Series

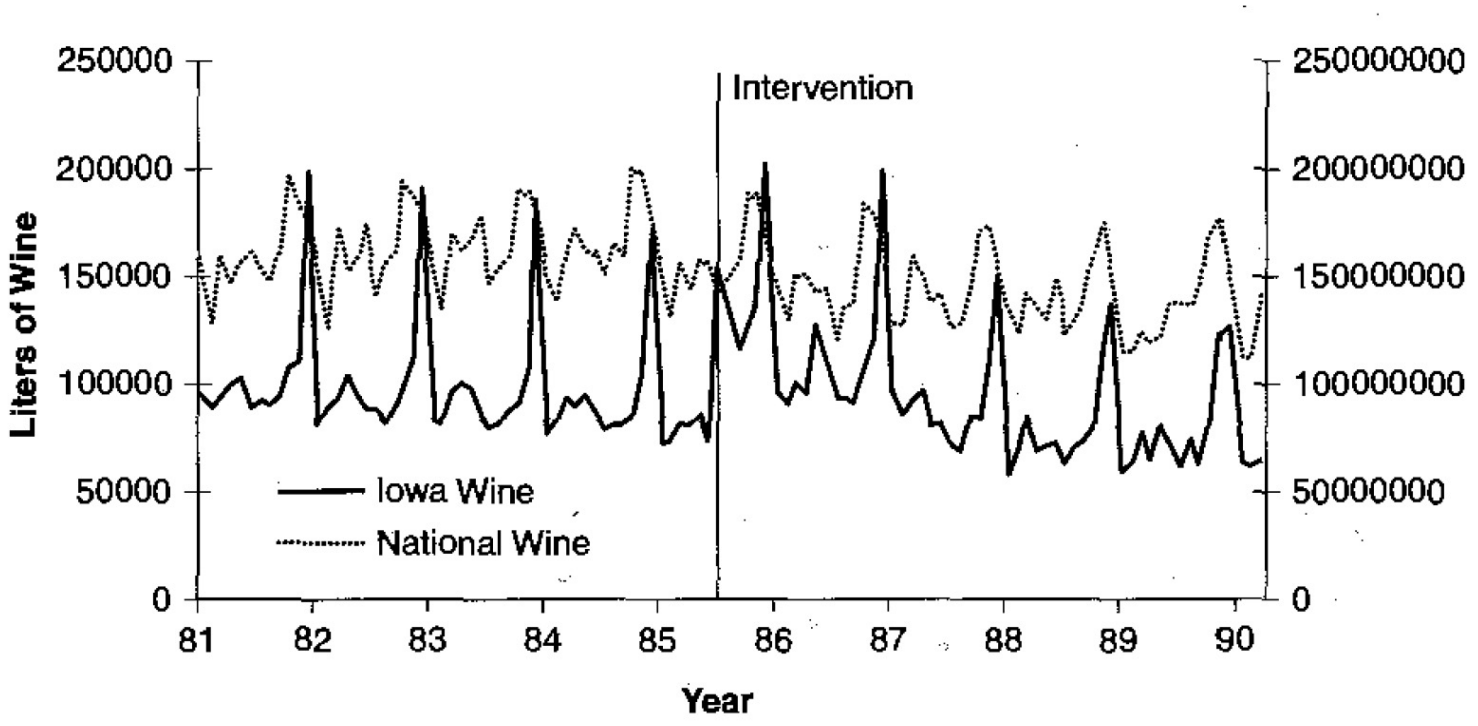
O_1	O_2	O_3	O_4	O_5	X	O_6	O_7	O_8	O_9	O_{10}
O_1	O_2	O_3	O_4	O_5		O_6	O_7	O_8	O_9	O_{10}

July 1, 1985, Iowa ended its public monopoly on liquor sales. 200 state-owned stores → 1200 private stores

An early time-series analysis examined alcohol sales for 2.5 years after the law took effect (until December 1987) and found that wine consumption increased by 93%

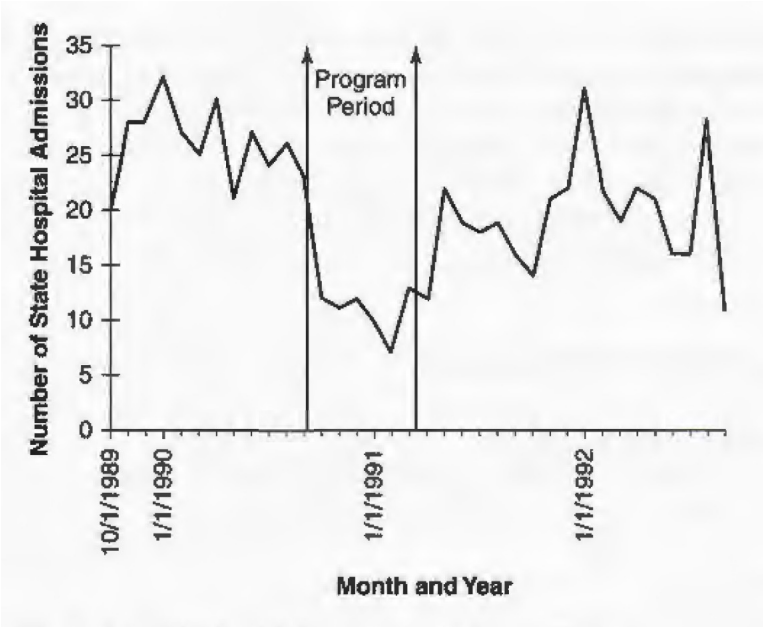
Mulford et al. (1992) investigated this matter further by adding a control series and extending the data in 1990.

Adding a Nonequivalent No-Treatment Control Group Time Series



Removing the Treatment at a Known Time

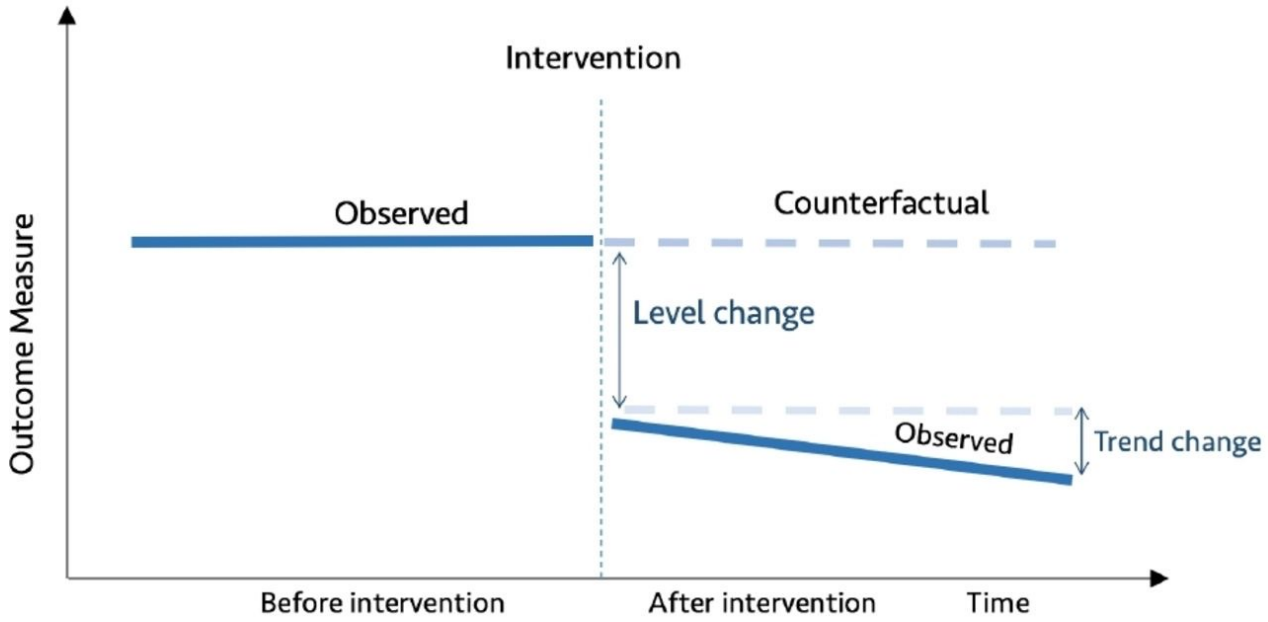
O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8 O_9 ~~X~~ O_{10} O_{11} O_{12} O_{13}

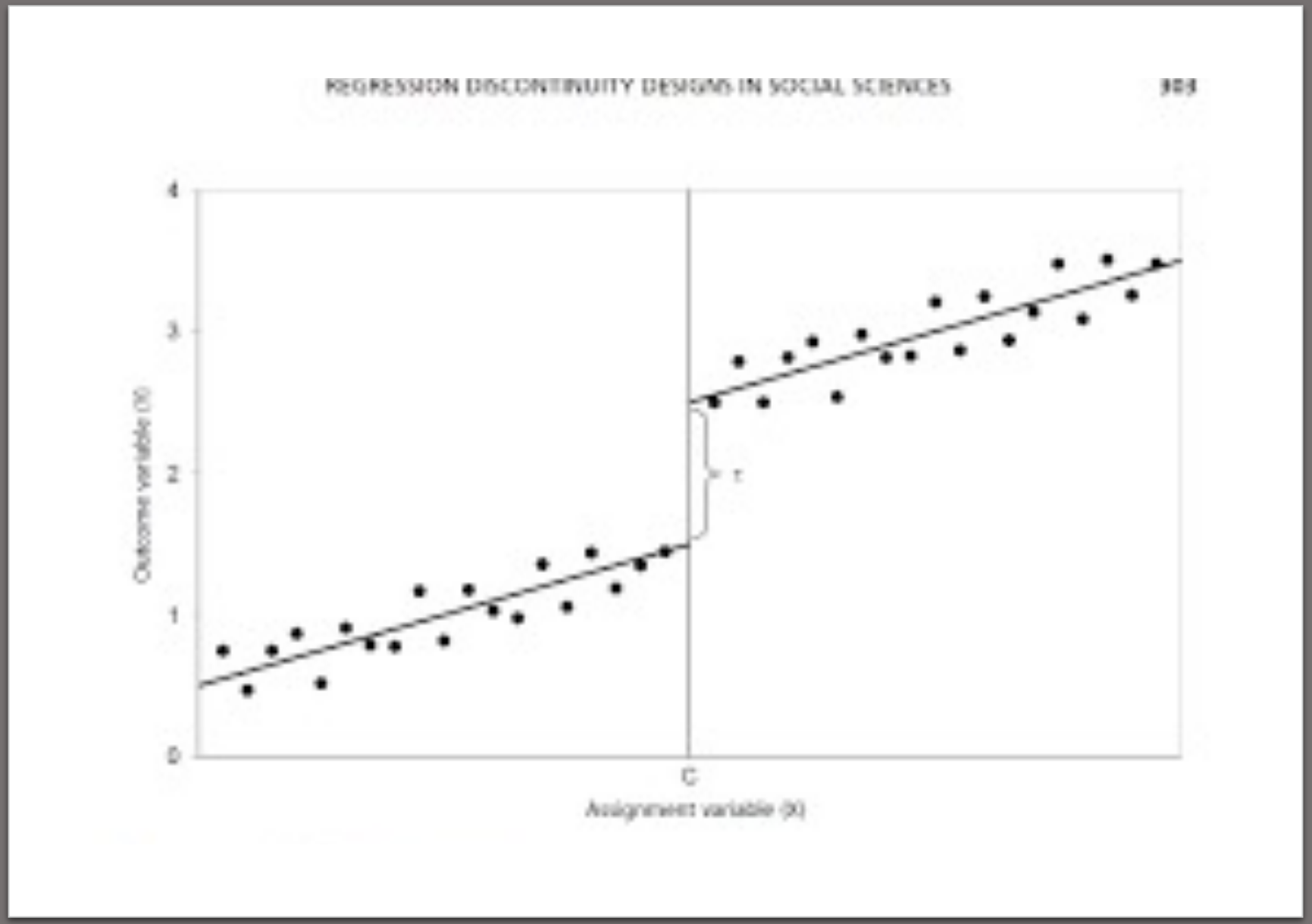


Interrupted Time Series

Causal hypothesis –

the observations after treatment will have a different slope or level from those before treatment.





RDD -- Regression Discontinuity Design

RD (Regression Discontinuity) Design

When prison inmates are first release the oven like a job and other finance financial resources to help them become productive members of society.

- Do some of them return to crime after leaving prison in order to get those resources?

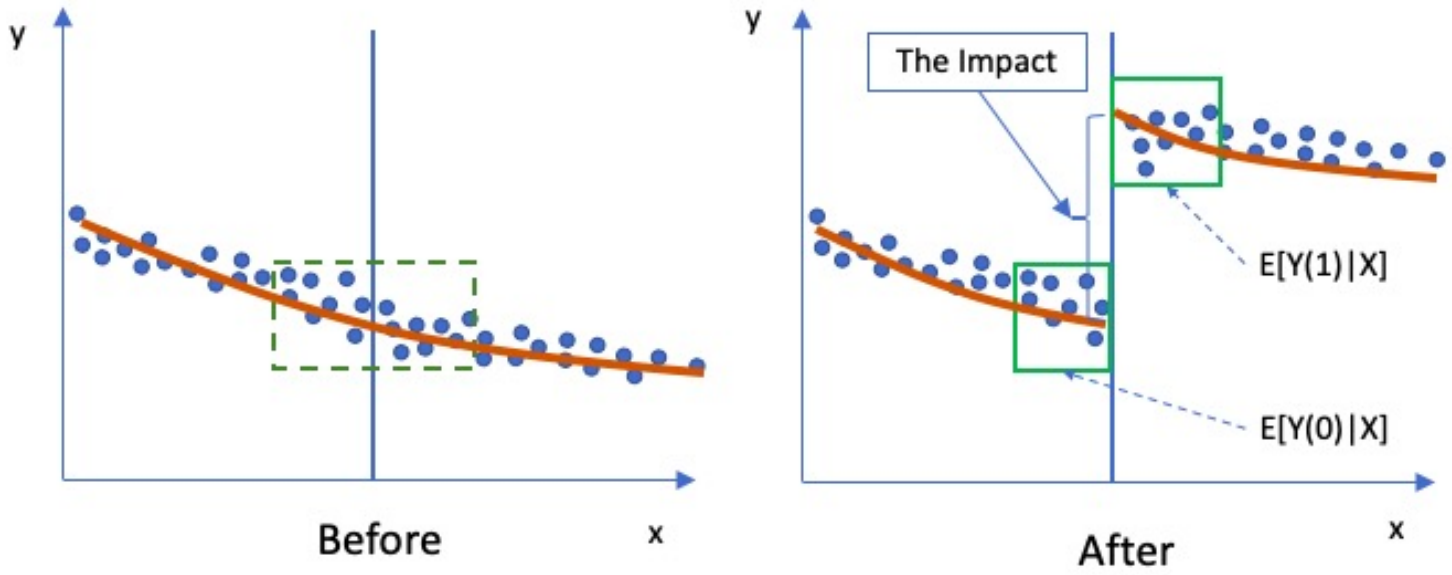
- Will providing them with funds and other release reduce further offending?

RDD – Basic structure

Assigns units to conditions **on the basis of a cutoff score on an assignment variable**

O_A	C	X	O_2
O_A	C		O_2

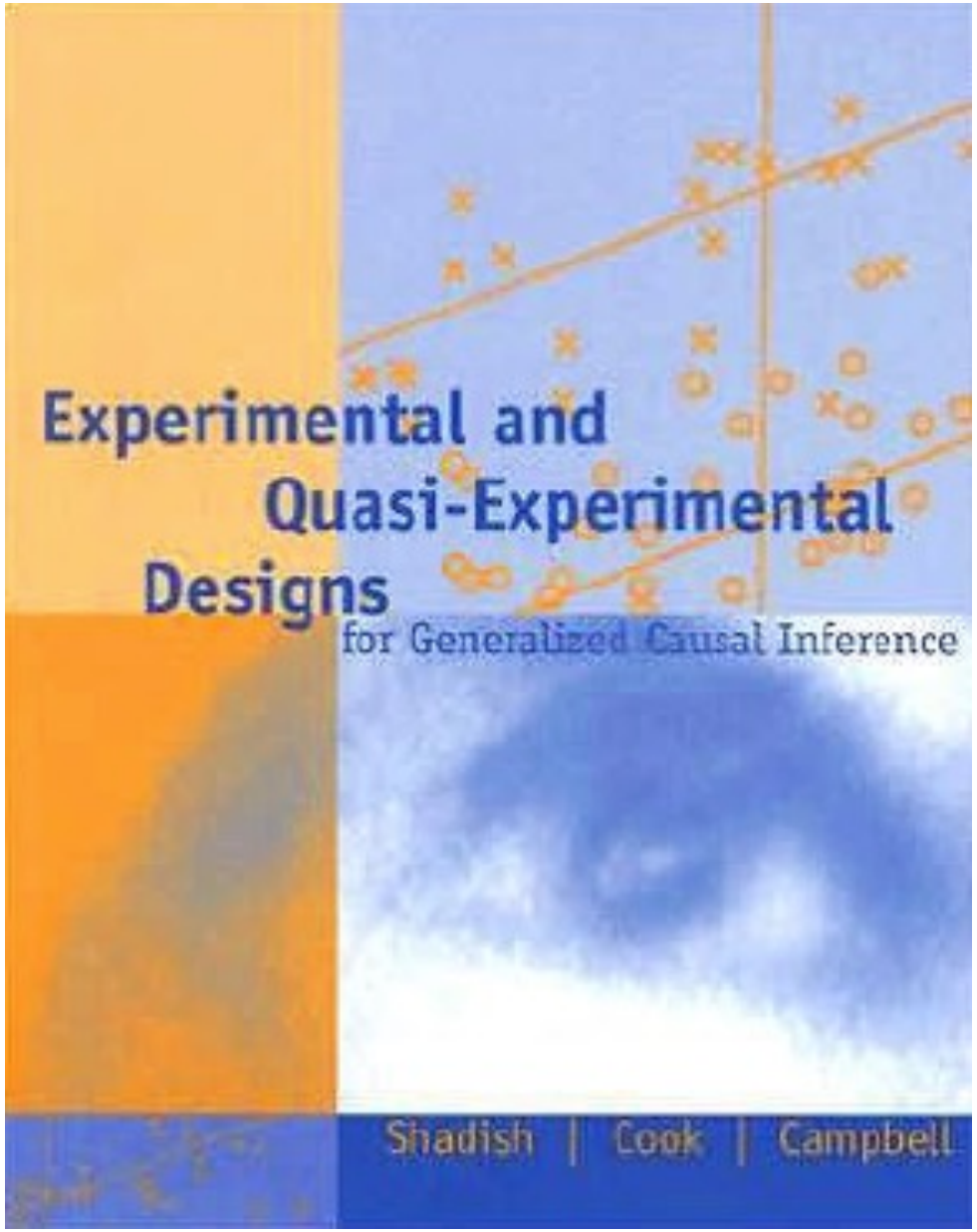
- O_A -- pre-assignment measure of the assignment variable
- C – cutoff score



The Assignment Variable and the Cutoff

- The assignment variable often assesses merit or need.
- The assignment variable cannot be caused by treatment
- The best assignment variable is a continuous variable
- The assignment variable need not be a pretest.
- Choice of cutoff score depends on many considerations.
- Assignment to treatment must be based only on the cutoff score

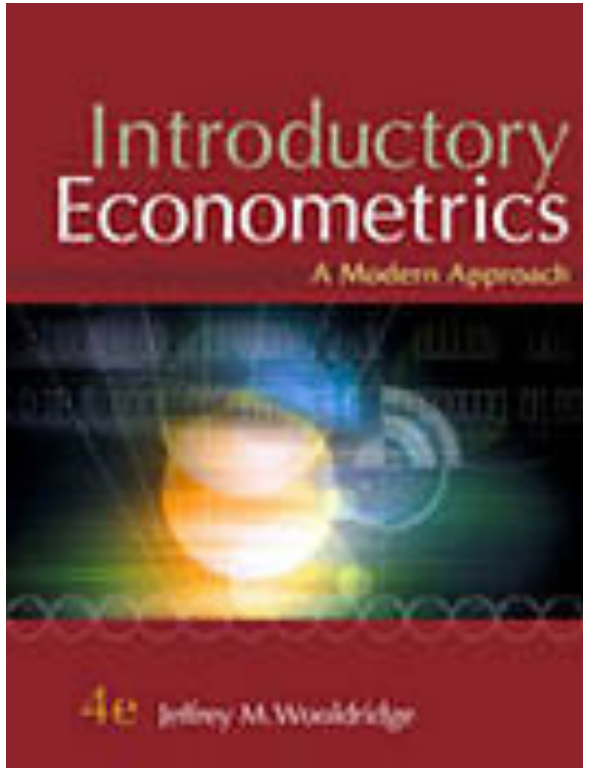
References

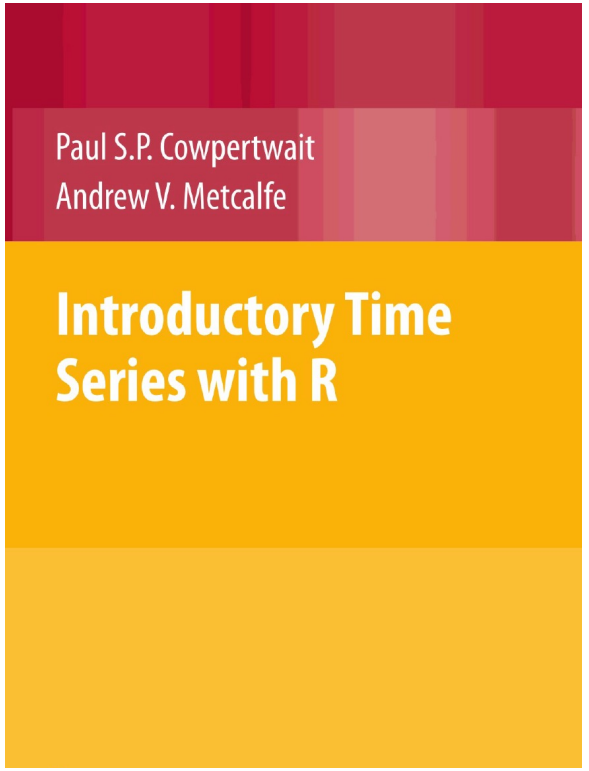

Experimental and
Quasi-Experimental
Designs
for Generalized Causal Inference
Shadish | Cook | Campbell

Other Readings

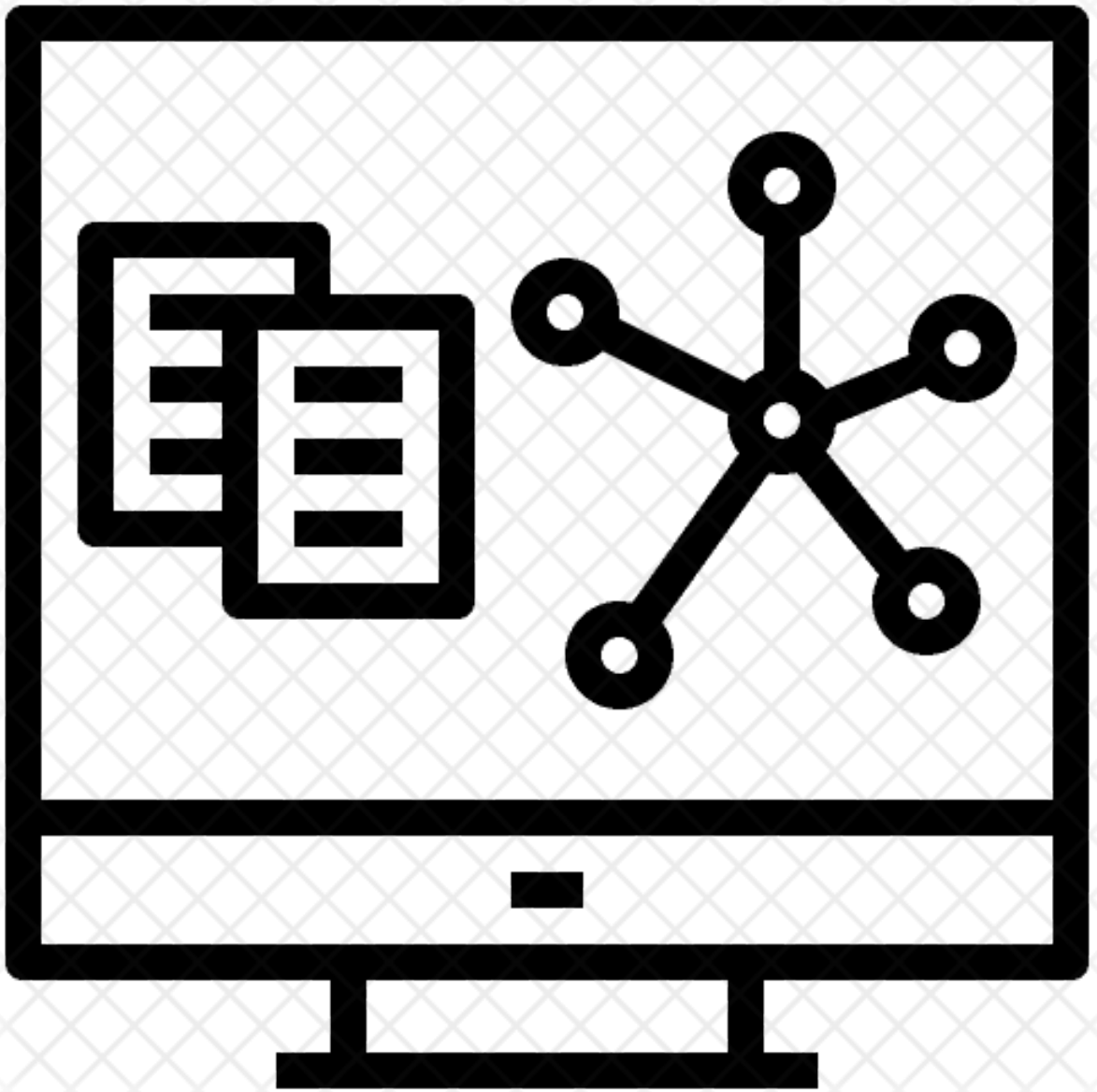
- Cowpertwait, P. S., & Metcalfe, A. V. (2009). Introductory time series with R. Springer Science & Business Media.

- Woolridge, J. M. (2003). Introductory econometrics: A modern approach. Thomson, Mason. Chapter 10 - Time series



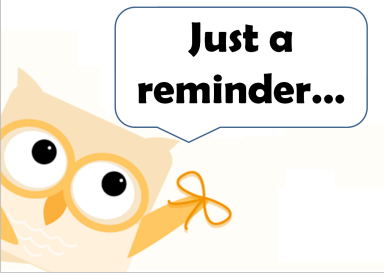


Modelling time series



Stationarity

- The data has **constant mean and variance**



Autoregression (AR)

- A statistical model is said to be autoregressive if it predicts future values based on previous values.
- ***AR(parameter)*** -- parameter is the number of independent variables or the count of past values considered for forecasting.

  E.G., AR (n), n = [0,1,2,...]

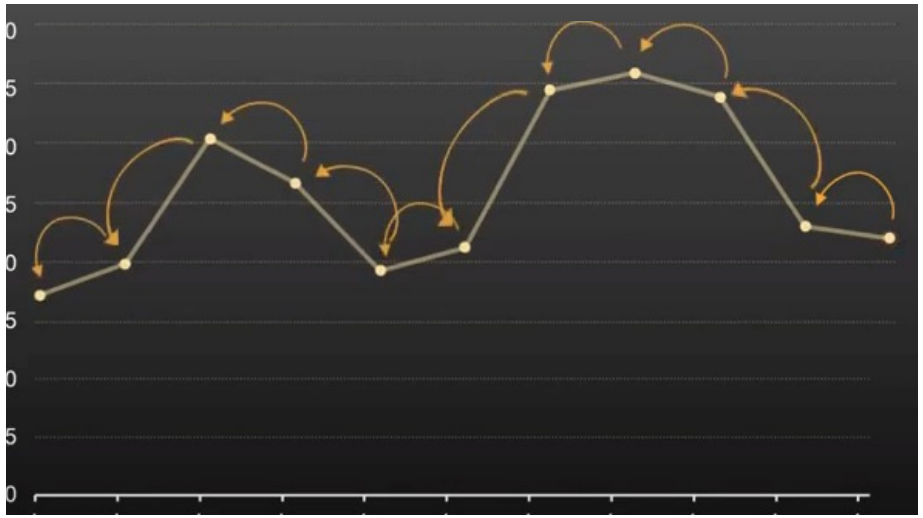
target

lagged target

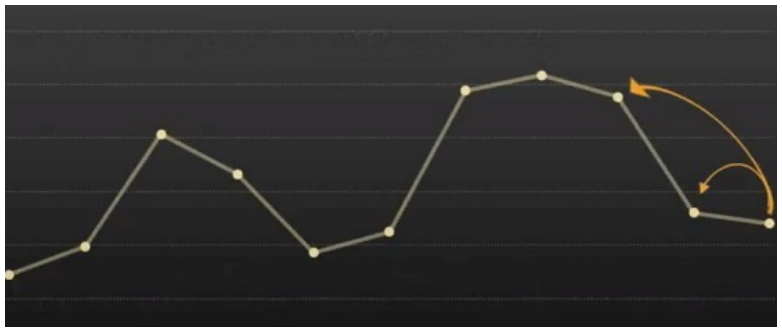
AR (4): $Yt = B1\,(Yt\text{-}1) + B2\,(Yt\text{-}2) + B3\,(Yt\text{-}3) + B4\,(Yt\text{-}4)$ + Errors

Date	Yt	Yt-1	Yt-2	Yt-3	Yt-4
01-01-2019	169				
01-02-2019	175	169			
01-03-2019	168	175	169		
01-04-2019	180	168	175	169	
01-05-2019	140	180	168	175	169
01-06-2019	120	140	180	168	175
01-07-2019	108	120	140	180	168
01-08-2019	100	108	120	140	180
01-09-2019	110	100	108	120	140
01-10-2019	121	110	100	108	120
01-11-2019	120	121	110	100	108
01-12-2019	105	120	121	110	100
01-01-2020	To be forecasted				
02-01-2020					

1 lag



2 lag



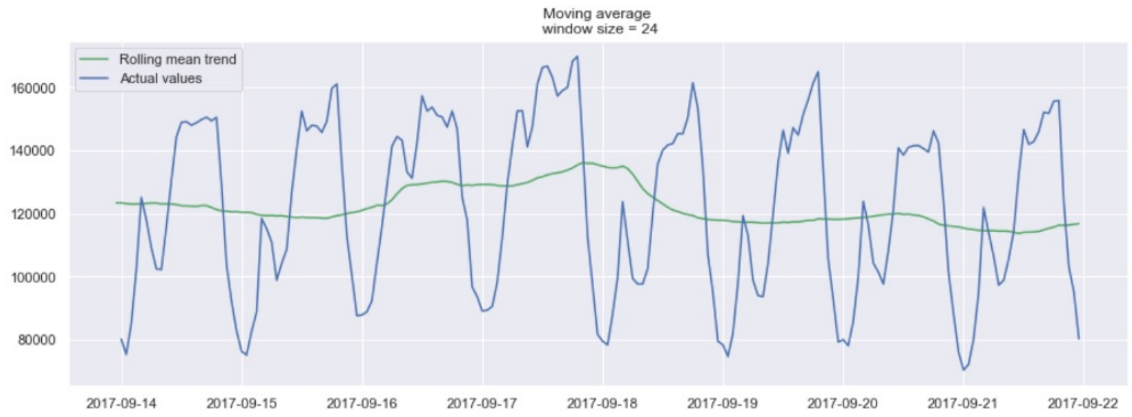
Autoregression (AR)

- A statistical model is said to be autoregressive if it predicts future values based on previous values.
- ***AR(parameter)*** -- parameter is the number of independent variables or the count of past values considered for forecasting.

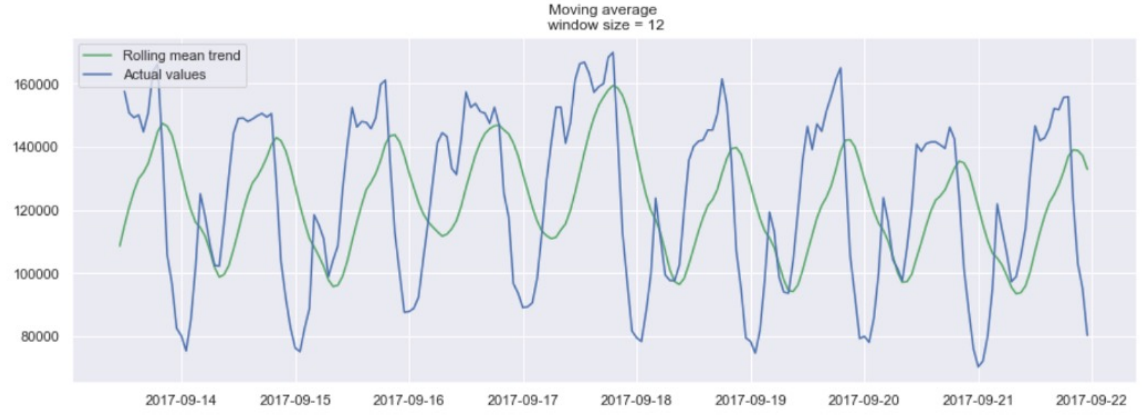
  E.G., AR (n), n = [0,1,2,...]
- forecast a series based solely on the past values in the series -- lags

Moving Average (MA)

- Statement: the next observation is the mean of all past observations.



24h window



12h window

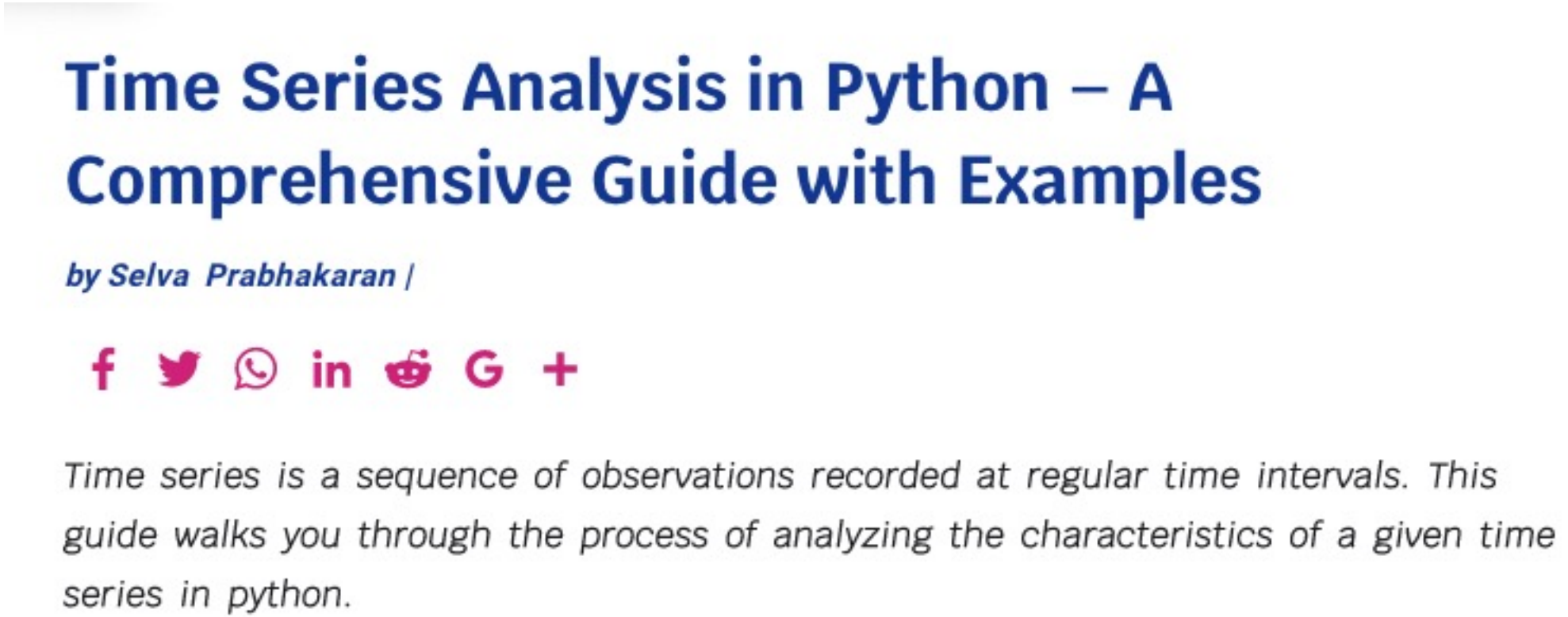
How to Use a Moving Average to Buy Stocks

By CORY MITCHELL | Updated Jan 7, 2021

https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp

ARMA (Auto-regressive moving average)

ARIMA (Auto-regressive integrated moving average)



Time Series Analysis in Python – A Comprehensive Guide with Examples

by Selva Prabhakaran |

Time series is a sequence of observations recorded at regular time intervals. This guide walks you through the process of analyzing the characteristics of a given time series in python.

https://www.machinelearningplus.com/time-series/time-series-analysis-python/

statsmodels is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. An extensive list of result statistics are available for each estimator. The results are tested against existing statistical packages to ensure that they are correct. The package is released under the open source Modified BSD (3-clause) license. The online documentation is hosted at statsmodels.org.

https://www.statsmodels.org/stable/tsa.html