



Meet 01: Workshop 01

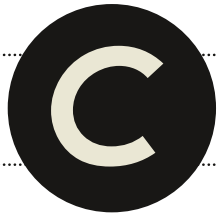
Capstone Project: Practice of Data Analysis for the 2022 Taiwanese Local Election

Instructors: Chung-pei Pien

ZU1920001/ZM1927001



International College of
INNOVATION
National Chengchi University
國立政治大學創新國際學院



CONTENTS

- 1 Introduction
- 2 Workshop
- 3 Data Project Process
- 4 Datasets in Taiwan Elections
- 5 Data Cleaning Practices
- 6 Prepare Your Data Projects

Introduction



For producing high quality data journalisms for Taiwanese local election on November 26th, ICI, CommonWealth, one of leading economic and political magazines in Taiwan, and Election Study Center, open a data journalism course in ICI this Fall.



International College of
INNOVATION
National Chengchi University
國立政治大學創新國際學院



Election Study Center, National Chengchi University



This is a professional training project. We are colleagues. Please call me Torrent or CP during this semester.

01

Introduction

Schedule:

Date	Description
08/26 – 09/11	Online coding skill training
09/15, 17, 18, 22	Workshop
09/26 – 10/13	Produce project proposals (Report proposal on 10/13)
10/14 – 11/17	Carry out projects
11/18 – 11/24	Rehearsal
11/24	Final check
11/26	Election day (4:00pm-11:00pm)
11/26 – 11/28	Produce results

**Every Thursday 9-12 is the project meeting.
Every student should attend to report progresses.**

01

Introduction

Item	Points	%
Online Coding Training	500	18%
Attendance (18 x 50)	900	33%
Weekly Report (7 x 50)	350	13%
Project Proposal	200	7%
Rehearsal	100	4%
Final Check	200	7%
Final Paper	500	18%
Total	2750	100%

Your project may not be published by CommonWealth, but you still can get a good grade.

01

Introduction

Item	Points	%
Project Proposal	200	7%

All teams should represent your proposals on October 13.

	Team Points	PM bonus	Writers bonus
A+	190	6	4
A	184	6	4
A-	180	4	4
B+	170	4	2
B	164	4	2
B-	160	2	2
C+	150	0	0
C	140	0	0

01

Introduction

Item	Points	%
Rehearsal	100	4%

Using CEC's demo data to rehearse the outputs

	Team Points	PM bonus	Writers bonus
A+	95	3	2
A	92	3	2
A-	90	2	2
B+	85	2	1
B	82	2	1
B-	80	1	1
C+	75	0	0
C	70	0	0

01

Introduction

Item	Points	%
Final Check	200	7%

Using CEC's demo data to represent the rehearsal on Nov. 24

	Team Points	PM bonus	Writers bonus
A+	190	6	4
A	184	6	4
A-	180	4	4
B+	170	4	2
B	164	4	2
B-	160	2	2
C+	150	0	0
C	140	0	0

01

Introduction

Item	Points	%
Final Paper	500	18%

Your project's final paper. Commonwealth will set the deadline.

	Team Points	PM bonus	Writers bonus
A+	475	15	10
A	460	15	10
A-	450	10	10
B+	425	10	5
B	410	10	5
B-	400	5	5
C+	375	0	0
C	350	0	0

Workshop

Location: **Room 406**, International Building

9/15: 9:10-12:00

Introduction of Election Data Journalism Reports
and Datasets: Governmental Open Data, Survey
Data, and Text Mining Data

Speaker: Chung-pei Pien

02

Workshop

Location: **Room 304**, International Building

9/17: 9:10-12:00

Taiwan's Democratization and Party System



Speaker: Eric Yu (Research Fellow, Election Study Center, NCCU)

02

Workshop

Location: **Room 304**, International Building

9/17: 13:30-16:30

Topics of Taiwan's Politics: Identity Politics,
Electoral System, and Woman Representation

Speaker: Nathan Batto (Associate Research Fellow,
Institute of Political Science, Academia Sinica)



02

Workshop

Location: **Room 304**, International Building

9/18: 9:10-12:00



Public Outreach for Researchers: Social Media and News Outlet

Speaker: Lev Nachman (Jointly Appointed Assistant Professor of ICI & IDASAT, NCCU)

Lev asks all students to create a Twitter account for his workshop.

Location: **Room 304**, International Building

9/18: 13:30-16:30

Taiwan's Elections and the Visualization Challenges

Speakers:

Silva Shih (史書華, Managing Editor, Commonwealth Magazine)

Sylvia Lee (李郁欣, PM, Commonwealth Magazine)

Steven Yeo (楊時鈞, Graphics Engineer, Commonwealth Magazine)

Ying-Yu Chen (陳盈諭, UI/UX Designer, Commonwealth Magazine)

Kwang-Yin Liu (劉光瑩, Managing Editor, Commonwealth Magazine)

Location: **Room 406**, International Building

9/22: 9:10-12:00

2022 Taiwanese Local Election's Data Cleaning and Management

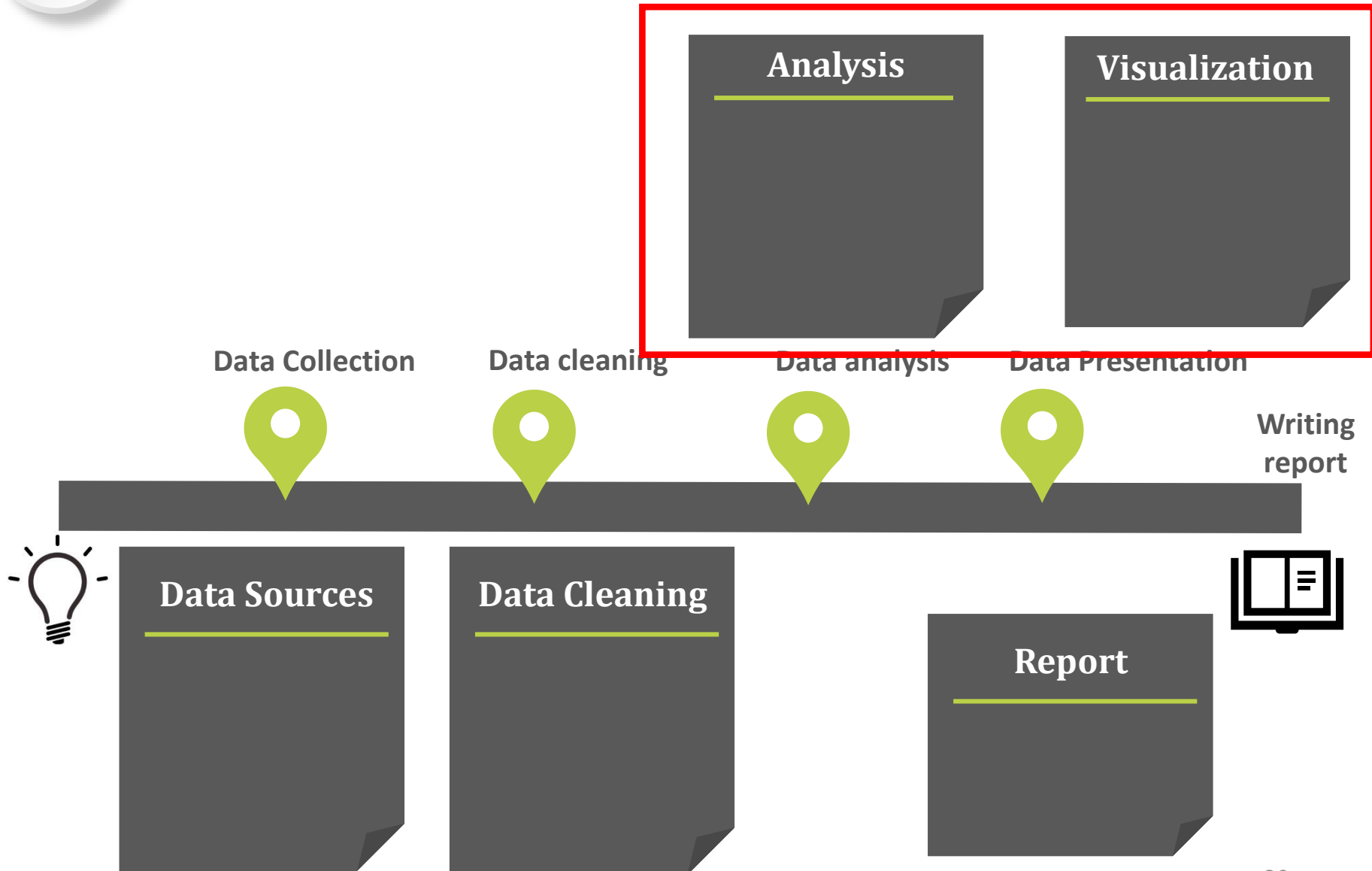
Speaker: Chung-pei Pien

Data Project Process

03

Data Project Process

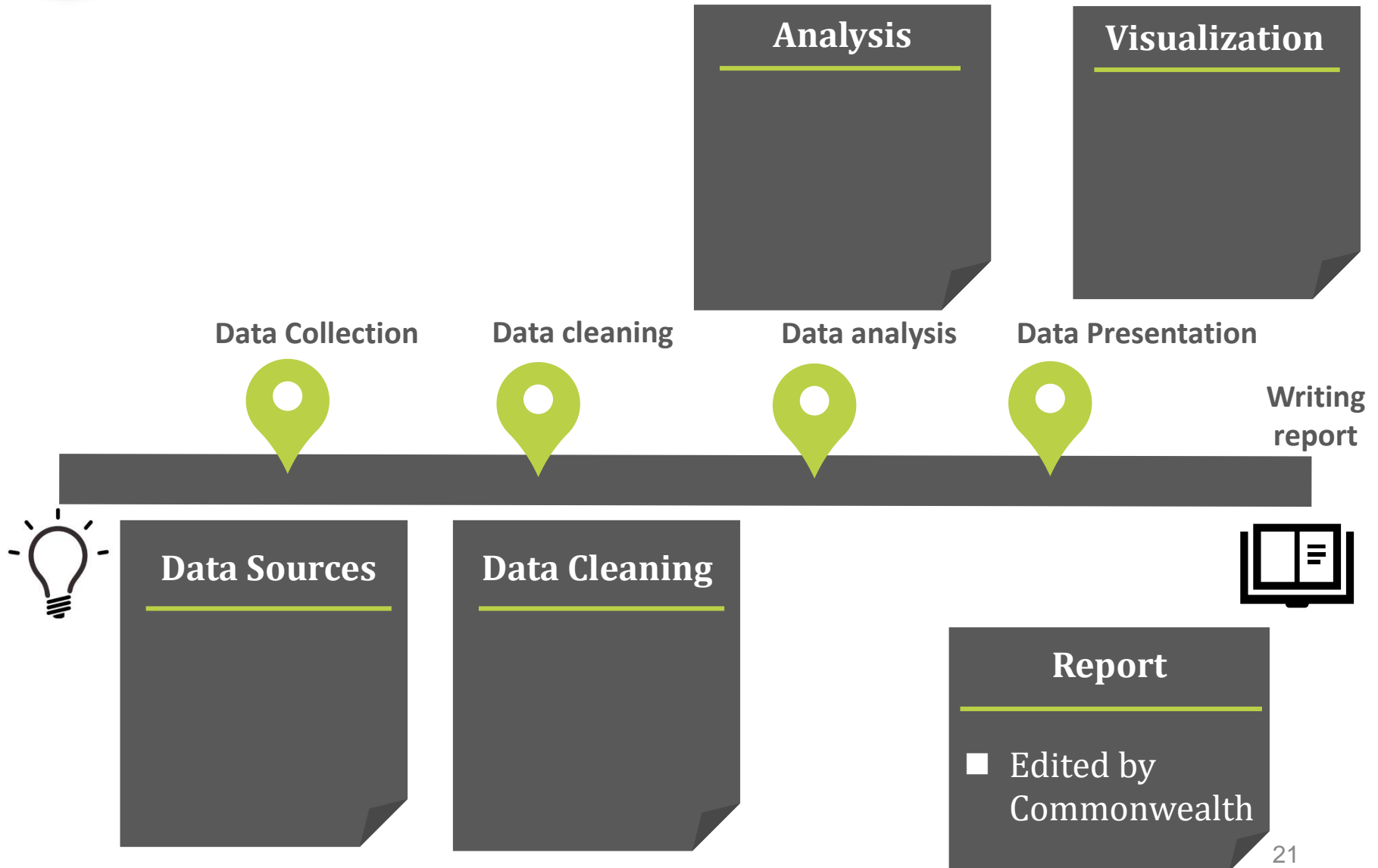
Datacamp's Assignments



03

Data Project Process

Your Projects



Datasets in Taiwan Elections

There are three important data sources:

What are they?

What they can do?

Central Election Commission

Offer all elections' polling stations' results after 1994




<https://data.gov.tw/dataset/13119>

	A	B	C	D	E	F
1	檔案名稱 下載連結網址					
2	選舉資料	http://data.cec.gov.tw/選舉資料庫/votedata.zip				

<http://data.cec.gov.tw/選舉資料庫/votedata.zip>

Central Election Commission

› 下載 › voteData › voteData › 2020總統立委 › 總統

名稱	修改日期
 elbase.csv	2020/1/13 上午 11:52
 elcand.csv	2020/1/13 上午 11:52
 elctks.csv	2020/1/13 上午 11:52
 elpaty.csv	2020/1/13 上午 11:52
 elprof.csv	2020/1/13 上午 11:52

Central Election Commission

	li_id	DPP_2020	KMT_2020	PFP_2020	valid_vote_2020	invalid_vote_2020	turnout_2020	num_voter_2020
16	0902001-007	454	1578	66	1049	5	1054	2787
17	0902001-008	868	2700	112	1840	14	1854	4595
18	0902002-001	766	3238	128	2066	22	2088	5124
19	0902002-002	242	1154	32	714	6	720	1906
20	0902002-003	150	750	22	461	2	463	1119
21	0902002-004	340	1456	52	924	5	929	2226
22	0902002-005	208	1238	54	750	9	759	1971
23	0902002-006	154	882	24	530	5	535	1527
24	0902002-007	208	1316	24	774	13	787	2158
25	0902002-008	328	1438	34	900	10	910	2303
26	0902003-001	762	2876	174	1906	14	1920	4260
27	0902003-002	528	2014	112	1327	18	1345	3193
28	0902003-003	370	1588	62	1010	7	1017	2590
29	0902003-004	320	1224	58	801	9	810	1986
30	0902003-005	410	1964	76	1225	10	1235	3073
31	0902003-006	1748	4662	280	3345	19	3364	7667
32	0902003-007	172	850	28	525	3	528	1465
33	0902003-008	238	1064	40	671	7	678	1623

04

Datasets in Taiwan Elections

Government Open Data

The Socio-Economic Geographic Information System (SEGIS) managed by the Department of Statistics, Ministry of the Interior
<https://segis.moi.gov.tw/>

SEGIS 社會經濟資料服務平台

關於SEGIS 推動成果 資料與服務 統計地圖

現在位置：首頁 > 資料與服務 > 產品與服務查詢

縣市搜尋：請選擇縣市 請選擇類別：請選擇中類 可直接輸入關鍵字查詢 產品名稱 產品單位名稱 查詢

縮小搜尋範圍

資料時間

不限
今年
近三年
~ 年 確定

空間統計單元

不限
縣市別
鄉鎮市區別
村里別
二級發布區
一級發布區

資料搜尋條件：村里別 x

20	第 1 共 144 頁	共 2869 筆
產品與服務名稱	空間範圍	空間統計單元
111年6月行政區人口統計 村里	全國	村里
111年6月行政區人口指標 村里	全國	村里
111年6月行政區五歲年齡組性別人口統計 村里	全國	村里
111年6月行政區十歲年齡組性別人口統計 村里	全國	村里
111年6月行政區二段年齡組性別人口統計 村里	全國	村里
111年6月行政區分齡兒童及少年性別人口統計 村里	全國	村里
111年6月行政區原住民人口指標 村里	全國	村里
111年6月行政區原住民人口統計 村里	全國	村里

SEGIS provides city, district and li levels' population, education, and income data. You can merge them into election data.

2020 年總統選舉提名初選民調問卷格式

二〇一九年三月六日第十八屆第二十五次中常會通過

您好！這裡是○○○○所進行的一項電話訪問。

我們想請教您一些有關總統選舉的看法。

（問題不多，只要幾分鐘就好）

（您的電話是電腦隨機抽出來的，您的意見是很有代表性的）

1. 請問這裡是住家電話，還是公司電話？

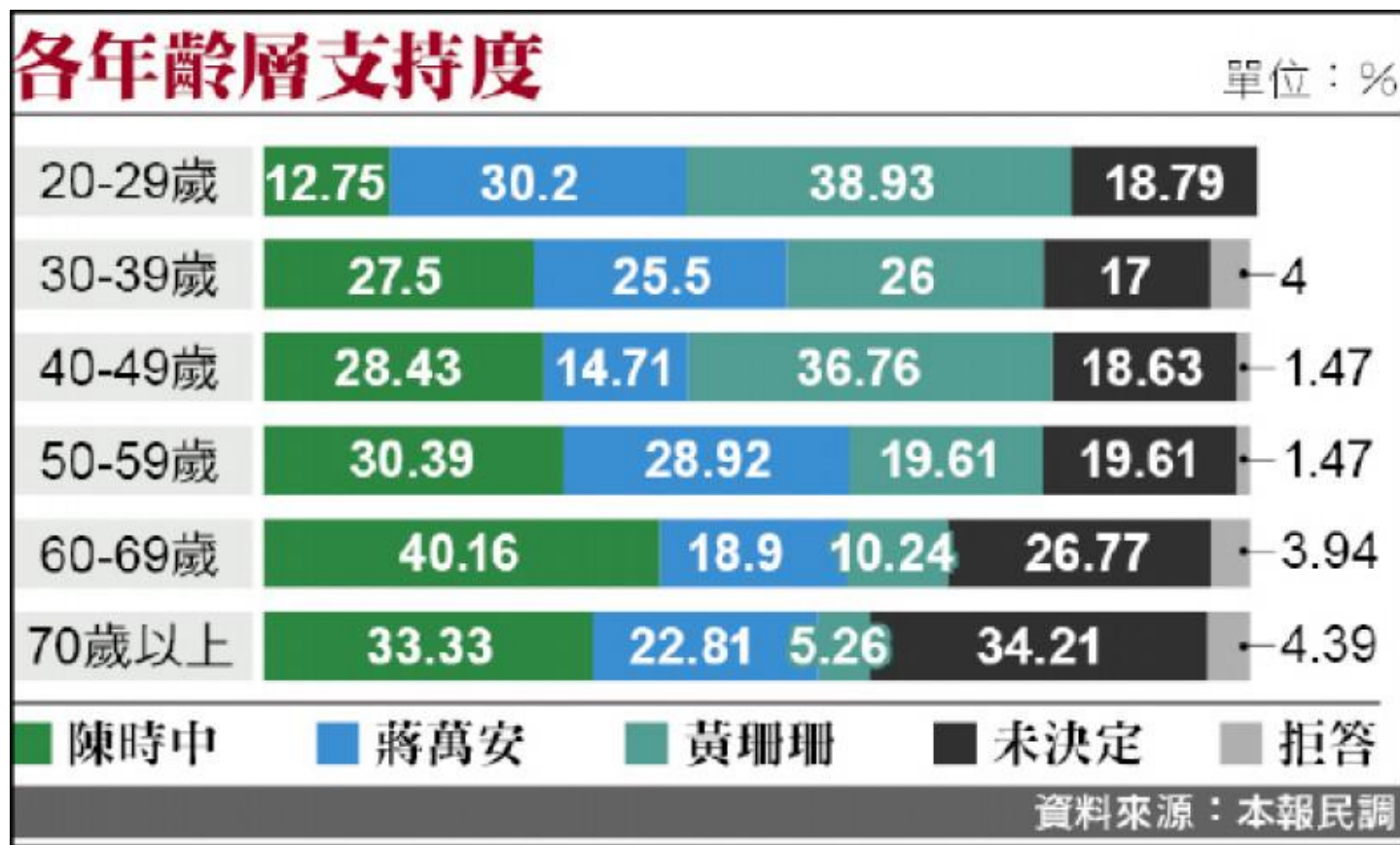
(01) 住家電話

(02) 住商合一

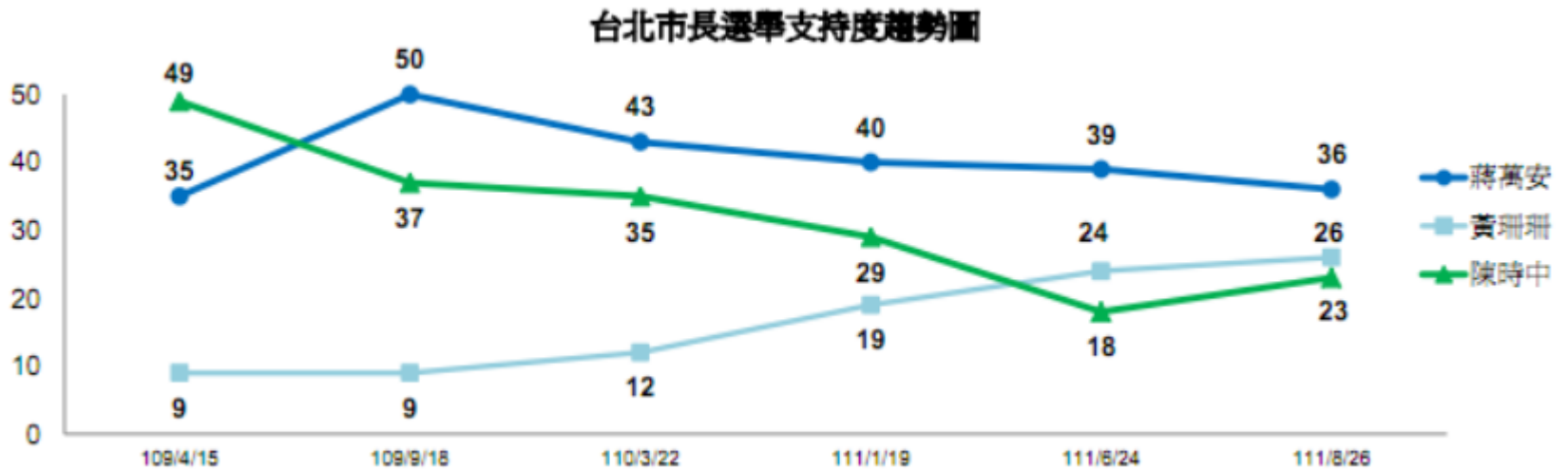
2. 請問您住在這裡嗎？

請問您是否滿 20 歲有投票權了？

3. 請問您目前的戶籍在哪一個縣市？



Survey Data



Survey is the only way to gather individual attitudes and summarize different groups' attitudes.

Suppose Village01 has the following characteristics:

KMT	60%
DPP	40%

Male	60%
Female	40%

Can we say V01's male voters are more likely to vote KMT?

We can't. If we do, it's called **ecological fallacy**.

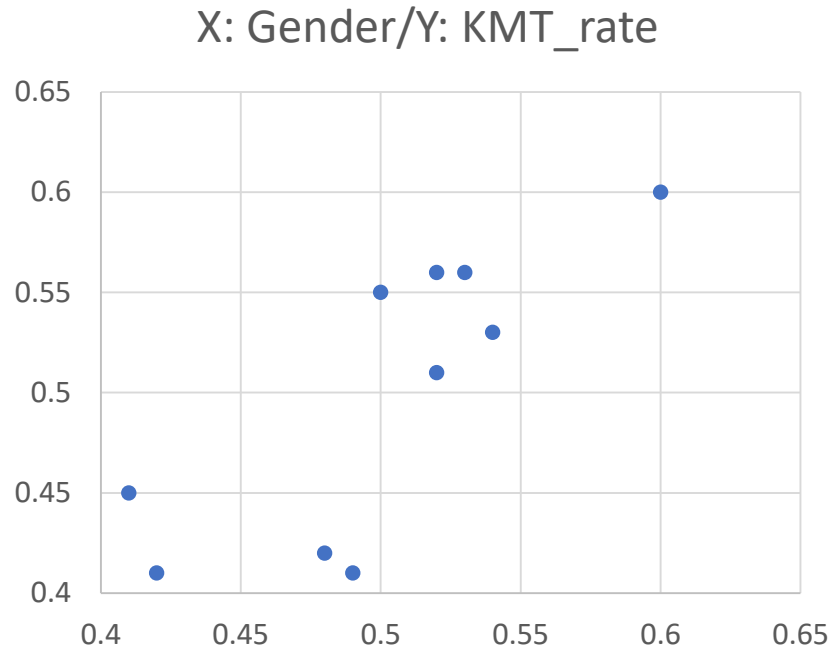
Maybe all female vote KMT and 20% male vote KMT.

04

Datasets in Taiwan Elections

Survey Data

Village	Male	KMT_rate
V01	0.60	0.60
V02	0.53	0.56
V03	0.54	0.53
V04	0.49	0.41
V05	0.41	0.45
V06	0.52	0.51
V07	0.48	0.42
V08	0.52	0.56
V09	0.42	0.41
V10	0.50	0.55



We only can say: **a village** with high ratio of male is more likely to vote KMT.

Individual	Gender	Party
001	Male	KMT
002	Female	KMT
003	Male	DDP
004	Male	KMT
005	Male	KMT
006	Female	DDP
007	Male	KMT
008	Female	KMT
009	Male	DDP
010	Female	DDP

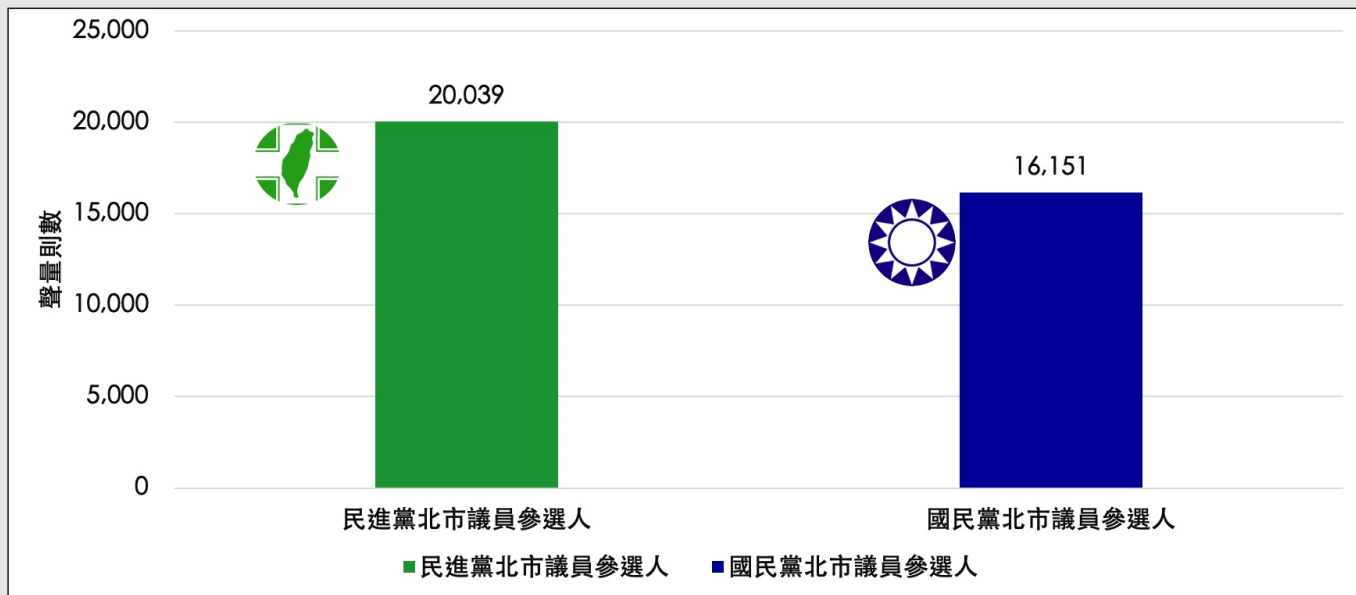
	KMT	DDP
Male	4	2
Female	2	2

Collect data from social media: Facebook, Decard, ptt, or Twitter

Using text mining skills to show hot words and sentiments of specific issues.

DDP and KMT candidates' interactions when they registered.

民進黨、國民黨北市議員參選人：登記參選期間網路聲量



數據區間：2022/8/26-2022/9/1

Data Cleaning Practices

05






Data Cleaning Practices

When you choose one of three data sources, you need to know one important thing:

There is a huge gap between raw data and the table files you received in my Big Data and Social Analysis.

> 下載 > voteData > voteData > 2020總統立委 > 總統

名稱

 elbase.csv
 elcand.csv
 elctks.csv
 elpaty.csv
 elprof.csv

修改日期

2020/1/13 上午 11:52
 2020/1/13 上午 11:52
 2020/1/13 上午 11:52
 2020/1/13 上午 11:52
 2020/1/13 上午 11:52

IL_id	DPP_2020	KMT_2020	PFP_2020	valid_vote_2020	invalid_vote_2020	turnout_2020	num_voter_2020
16	0902001-007	454	1578	66	1049	5	1054
17	0902001-008	868	2700	112	1840	14	1854
18	0902002-001	766	3238	128	2066	22	2088
19	0902002-002	242	1154	32	714	6	720
20	0902002-003	150	750	22	461	2	463
21	0902002-004	340	1456	52	924	5	929
22	0902002-005	208	1238	54	750	9	759
23	0902002-006	154	882	24	530	5	535
24	0902002-007	208	1316	24	774	13	787
25	0902002-008	328	1438	34	900	10	910
26	0902003-001	762	2876	174	1906	14	1920
27	0902003-002	528	2014	112	1327	18	1345
28	0902003-003	370	1588	62	1010	7	1017
29	0902003-004	320	1224	58	801	9	810
30	0902003-005	410	1964	76	1225	10	1235
31	0902003-006	1748	4662	280	3345	19	3364
32	0902003-007	172	850	28	525	3	528
33	0902003-008	238	1064	40	671	7	678

05

Data Cleaning Practices

A	B	C	D	E	F
Open data		Survey		Text Mining	
Data Cleaning		Data Cleaning		Data Cleaning	
Project I	Project II	Project III	Project IV	Project V	Project VI

I suggest teams with same data sources can clean data together.






05

Data Cleaning Practices

In this capstone course, you have to fill out the gap:
To clean the data based on your project's demand.

Let's try to clean 2020 Presidential Election results.

› 下載 › voteData › voteData › 2020總統立委 › 總統

名稱	修改日期
 elbase.csv	2020/1/13 上午 11:52
 elcand.csv	2020/1/13 上午 11:52
 elctks.csv	2020/1/13 上午 11:52
 elpaty.csv	2020/1/13 上午 11:52
 elprof.csv	2020/1/13 上午 11:52

選舉資料庫系統管理操作手冊

轉檔檔案

檔案名稱	檔案說明	備註
elbase.csv	行政區基本資料	
elcand.csv	候選人基本資料	
elpaty.csv	政黨基本資料	
elprof.csv	選舉概況檔	
elctks.csv	候選人得票檔	
elrepm.csv	不分區政黨代表人基本檔	只有不分區及僑居國外國民立委選舉才需要
elretks.csv	不分區政黨得票檔	只有不分區及僑居國外國民立委選舉才需要

備註：

- 檔案內容編碼須為UTF-8編碼
- 欄位間以逗號隔開
- 檔案內容請勿使用逗點符號

Prepare Your Data Projects

When you start your professional data projects:

1. What does Commonwealth want?
2. Produce precise results
3. You should use your adept analysis and collect an analysis that your audiences can understand

This capstone doesn't encourage you to learn fancy models or data visualization.

Simple is better!

To prepare the codes for this election's results

1. Merge this election's results with previous ones and other data (Check your codes during the rehearsal)
2. Scenario prediction

To prepare diverse scenarios:

KMT win

DDP win

What is win?

06

Prepare Your Data Projects

Next Thursday, all teams should report:

The 2018 election's code.

Your initial ideal of the project.