

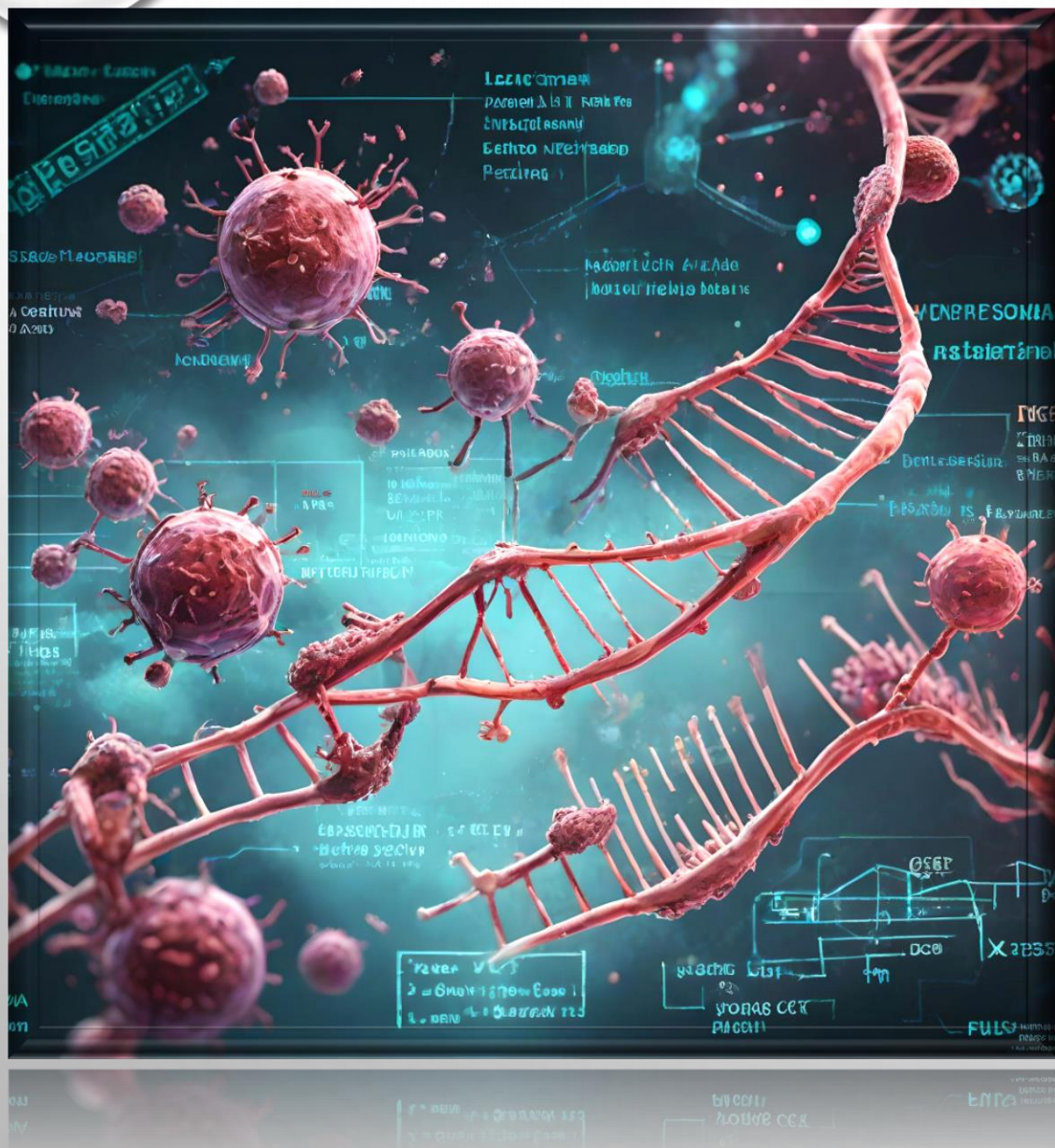
癌症發生統計分析

民國68-110年縣市別性別癌症別發生率資料

指導老師：簡志峰 老師

報告人員：王國昌 Brad

報告日期：2024/07/28



目

錄



01. 資料選取動機



02. 圖表說明



03. 問題與解決



04. 總結與感想



01. 資料選取動機



01. 資料選取動機

1. 選擇癌症統計分析的原因？

專業背景主要是生物科技，主要為接觸基礎分子生物實驗為主，本身對於癌症醫藥方面的部分也有興趣，因此想多了解近年來癌症發生與分佈的情形。

2. 資料來源與檔案內容

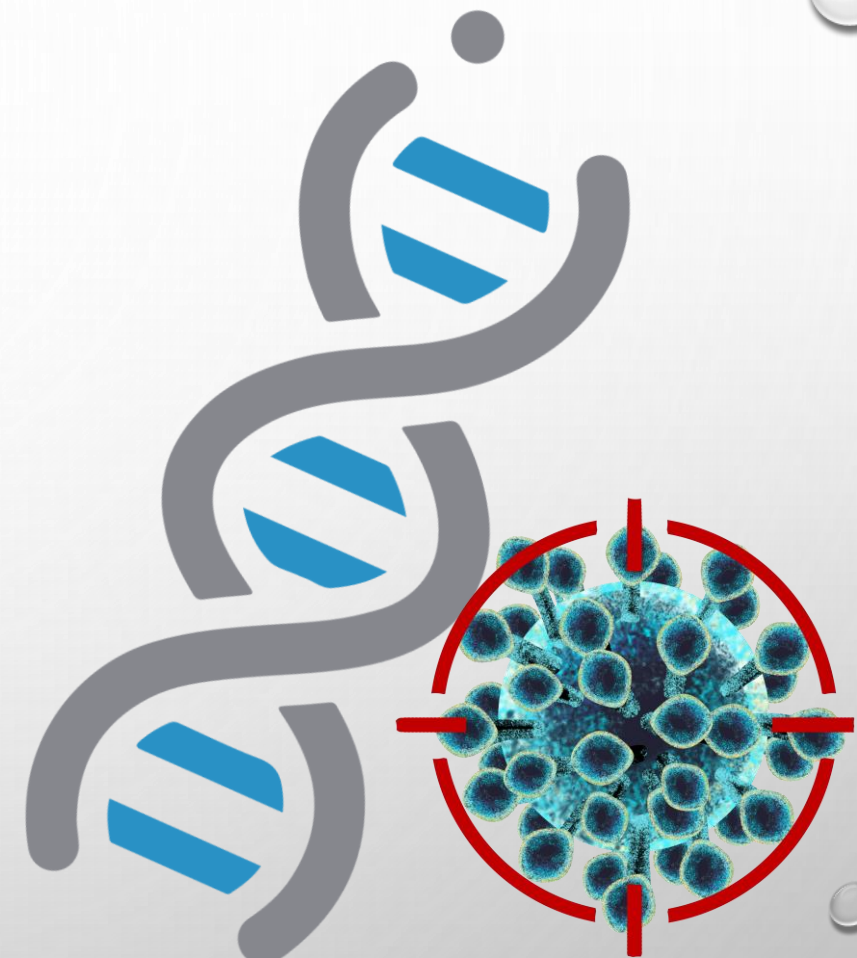
資料來源：

- 政府資料開放平臺



檔案內容：

- CSV檔案：癌症發生統計，提供我國癌症發生統計資料





01. 資料選取動機- 搜尋與觀察資料來源 CSV檔

政府資料開放平台
網站導覽 X f Language 小幫手 線上客服 會員登入

資料集 高應用價值主題專區 資料故事館 互動專區 消息專區 諮詢小組 授權條款 關於平台

癌症發生 搜尋癌症發生
熱門關鍵字：即時交通、空氣品質、歷史交通

資料集服務分類

1 癌症發生統計
提供我國癌症發生統計資料，供各界使用
主要欄位說明：癌症診斷年、性別、縣市別、癌症別、年齡標準化發生率 WHO 2000世界標準人口 (每10萬人口)、癌症發生數、平均年齡、年齡中位數、粗率 (每10萬人口)
資料提供屬性：檔案資料
衛生福利部國民健康署 © 2024-07-15 16:09:26 更新 © 50132 17292 22

2 高雄市癌症標準化發生率
癌症標準化發生率
主要欄位說明：年別、女性癌症標準化發生率、男性癌症標準化發生率
資料提供屬性：檔案資料
高雄市政府主計處 © 2024-07-17 17:00:08 更新 © 1081 107 0

3 十大癌症發生率
當年度每10萬人中發生癌症的比率
主要欄位說明：414 十大癌症發生率(86-109).pdf
資料提供屬性：檔案資料
行政院 © 2024-01-10 14:15:21 更新 © 1885 623 0

4 歷年惡性腫瘤發生率(攝護腺癌、肺癌、大腸直腸癌、女性乳癌、子宮頸癌)
資料來源：本署癌症登記資料庫。備註：標準化率係以2000年W.H.O.世界人口年齡結構為基準。
Source：Cancer Registry Databas(...詳內)
主要欄位說明：依年度別分by year、攝護腺癌Prostate Cancer(以2000年世界標準人口年齡結構計算之年齡標準化率，每十萬人)、肺癌Lung Cancer(以(...詳內)
資料提供屬性：檔案資料
衛生福利部國民健康署 © 2024-07-15 16:14:13 更新 © 14891 5540 2

癌症發生統計

提供我國癌症發生統計資料，供各界使用

主要欄位說明：癌症診斷年、性別、縣市別、癌症別、年齡標準化發生率 WHO 2000世界標準人口 (每10萬人口)、癌症發生數、平均年齡、年齡中位數、粗率 (每10萬人口)

資料提供屬性：檔案資料

衛生福利部國民健康署 © 2024-07-15 16:09:26 更新 © 50132 17292 22

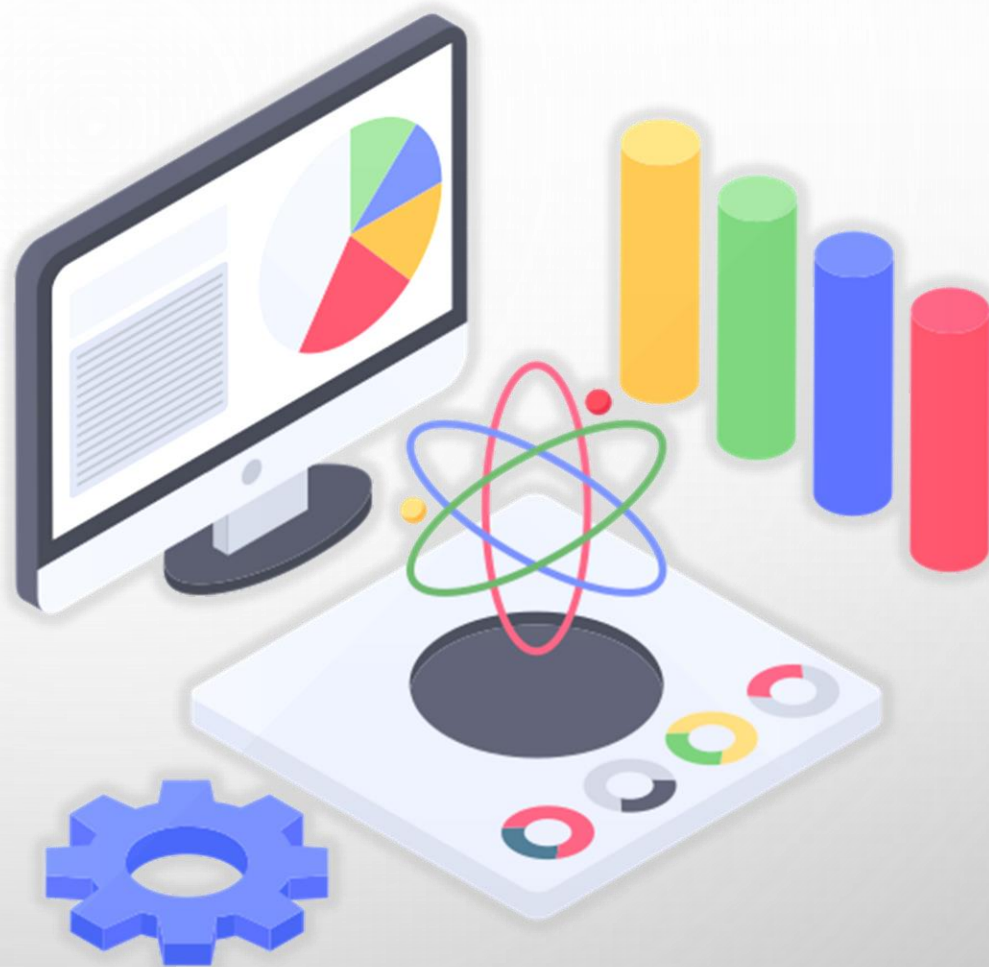
癌症診斷年	性別	縣市別	癌症別	年齡標準化發生率	癌症發生數	平均年齡	年齡中位數	粗率 (每10萬人口)
1974	全	全國	口腔、口腔	3.4	456	51.12	55	3.5
1975	全	全國	口腔、口腔	3.36	45	46.62	47	3.28
1976	全	全國	口腔、口腔	3.71	1,247	46.82	47	3.97
1977	全	全國	口腔、口腔	3.16	375	56.13	59	2.15
1978	全	全國	口腔、口腔	12.02	1,471	51.18	52	8.39
1979	全	全國	口腔、口腔	9.32	37	51.41	59	9.21
1980	全	全國	口腔、口腔	10.08	1,254	54.21	57	7.15
1981	全	全國	口腔、口腔	8.76	1,138	51.8	53	6.69
1982	全	全國	口腔、口腔	8.51	64	56.55	57.5	9.36
1983	全	全國	口腔、口腔	1.11	138	56.64	58	0.79
1984	全	全國	口腔、口腔	0.13	20	75.3	32	0.11
1985	全	全國	口腔、口腔	0.13	14	59.27	81	0.08
1986	全	全國	口腔、口腔	0.61	78	52.04	55	0.44
2021	男	台東縣	子宮頸	11.61	15	51.8	48	14.31
2021	女	台東縣	子宮頸	18.94	28	54.31	55	23.81
2021	男	台東縣	前列腺、輸精	3.72	3	48.35	47	2.95
2021	女	台東縣	睪丸	1.87	4	50.75	49	3.85
2021	男	台東縣	睪丸	0.66	1	54	54	0.96
2021	女	台東縣	睪丸	1.72	4	81	80	3.82
2021	男	台東縣	睪丸及睪丸	0.11	1	81	80	0.96
2021	女	台東縣	睪丸	4.43	5	54.6	64	4.41
2021	男	台東縣	睪丸及睪丸	24.21	33	59.5	51	31.29
2021	女	台東縣	睪丸及睪丸	0.85	1	58	58	0.96
2021	男	台東縣	睪丸及睪丸	0.34	2	74.5	74.5	1.41
2021	女	台東縣	睪丸及睪丸	3.37	8	61.87	65	3.71
2021	男	台東縣	睪丸及睪丸	2.1	12	58.25	64	11.52
2021	女	台東縣	睪丸及睪丸	1.65	4	75	73	3.85
2021	男	台東縣	睪丸及睪丸	279.35	511	62.66	63	416.71
2021	女	台東縣	睪丸及睪丸	0.62	1	84	84	1.41
2021	男	台東縣	睪丸及睪丸	5.53	7	60.57	64	8.96
2021	女	台東縣	睪丸及睪丸	13.82	19	60.67	63	18.76
2021	男	台東縣	睪丸及睪丸	2.72	4	79.25	78.5	3.92
2021	女	台東縣	睪丸及睪丸	1.62	2	86	86	3.83
2021	男	台東縣	睪丸及睪丸	13.33	18	66.83	65.5	15.25
2021	女	台東縣	睪丸及睪丸	3.7	8	80	80	1.41
2021	男	台東縣	睪丸及睪丸	28.51	31	54.29	52	43.38
2021	女	台東縣	睪丸及睪丸	1.9	2	59.5	58.5	3.82
2021	男	台東縣	睪丸及睪丸	4.32	5	54.4	52	3.84
2021	女	台東縣	睪丸及睪丸	2.6	8	52.5	58	11.21
2021	男	台東縣	睪丸及睪丸	4.61	5	53.4	55	7.34
2021	女	台東縣	睪丸及睪丸	3.52	3	64	63	4.25
2021	男	台東縣	睪丸及睪丸	7.72	8	51	59	11.21
2021	女	台東縣	睪丸及睪丸	0.38	1	84	84	1.41
2021	男	台東縣	睪丸及睪丸	1.86	1	47	47	1.41
2021	女	台東縣	睪丸及睪丸	3.85	5	71	79	7.34
2021	男	台東縣	睪丸及睪丸	100.18	121	62.36	58	178.36
2021	女	台東縣	睪丸及睪丸	11.49	1	81	80	17.46
2021	男	台東縣	睪丸及睪丸	16.21	1	35	35	17.44
2021	女	台東縣	睪丸及睪丸	26.78	2	45	45	24.87
2021	男	台東縣	睪丸及睪丸	10.99	1	71	71	17.46

癌症診斷年數據來源：
1974年到2021年

性別資料分為：
全體、男性、女性

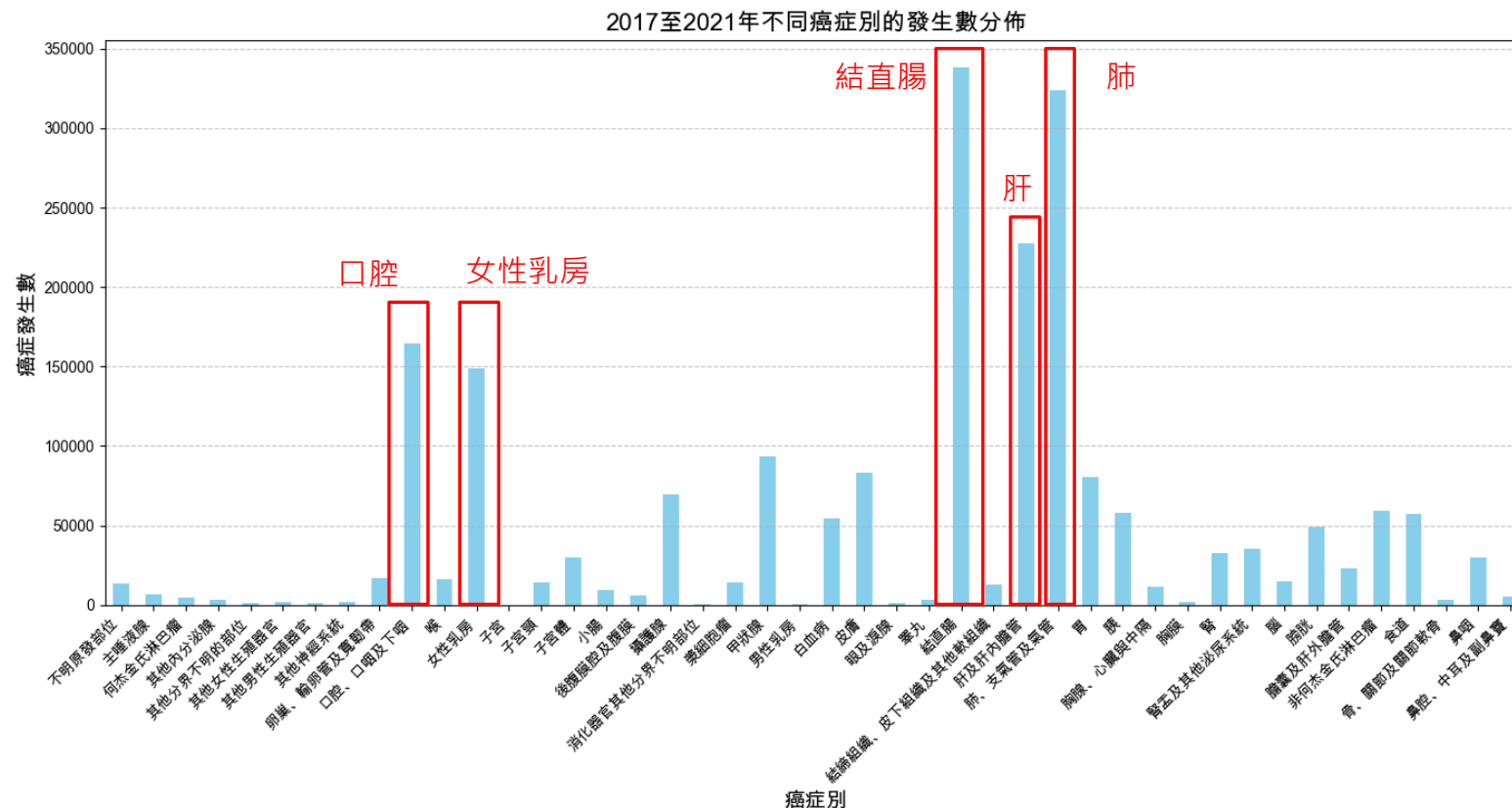
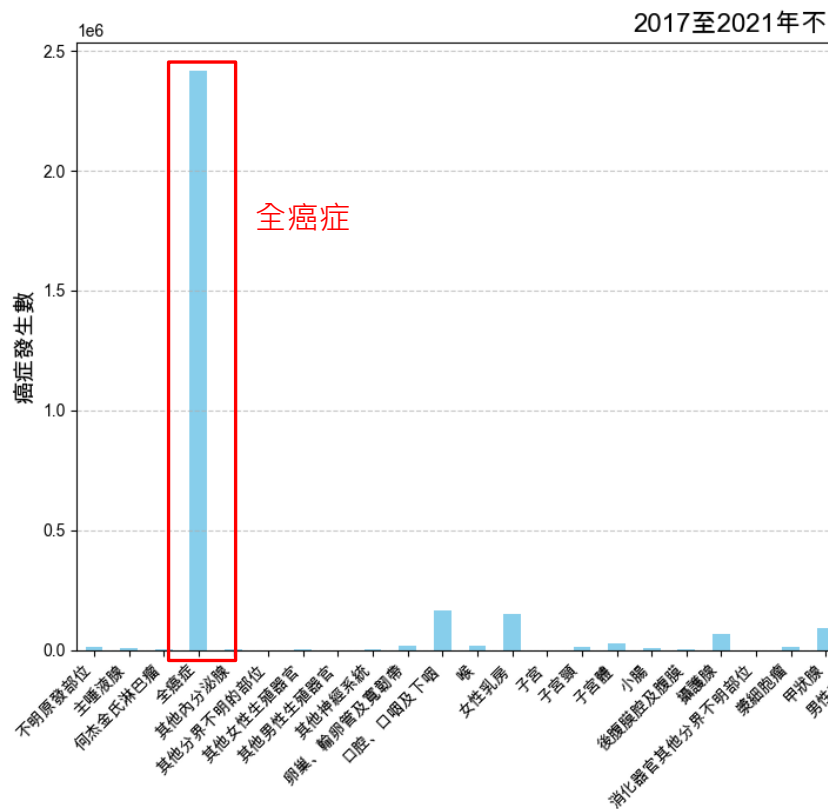
縣市資料為：
區域性與全國

02. 圖表說明





02. 圖表說明-資料分析(2017-2021年癌症分佈情形)



分析癌症發生統計結果：
考量結果數據的差異性，本次分析以2017至2021年的結果進行比較，統計結果中，男性與女性共有44種不同類別的癌症發生，其中以全癌症的記錄最多。

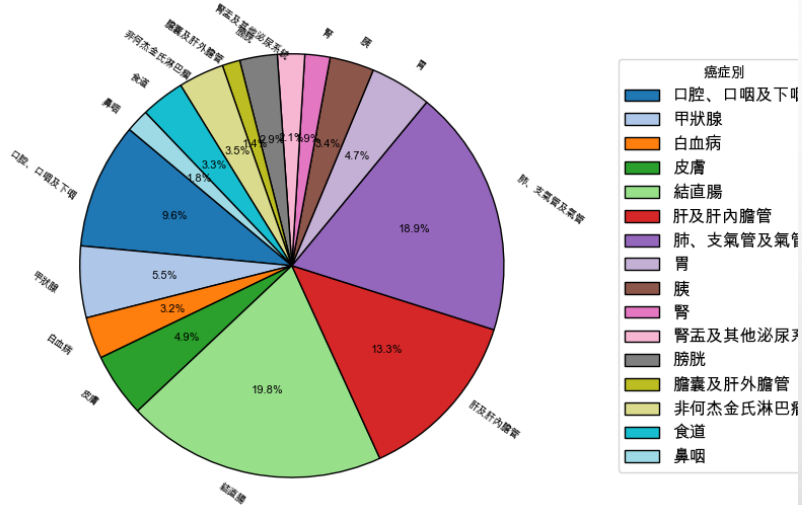
本次分析統計分析結果因全癌症較難明確定義，因此將全癌症結果剔除，不列入本次專案分析內。
其中以口腔、女性乳房、結直腸、肝及肺發生率最多



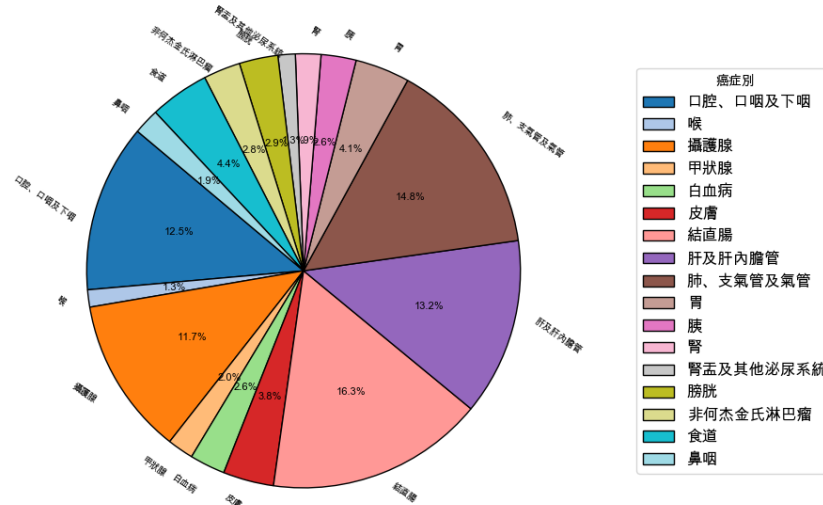
02. 圖表說明-資料分析(2017-2021年癌症分佈情形)

製圖說明：製作不同性別在2017年到2021年癌症發生的情形，觀察各性別癌症發生的類別有哪些。

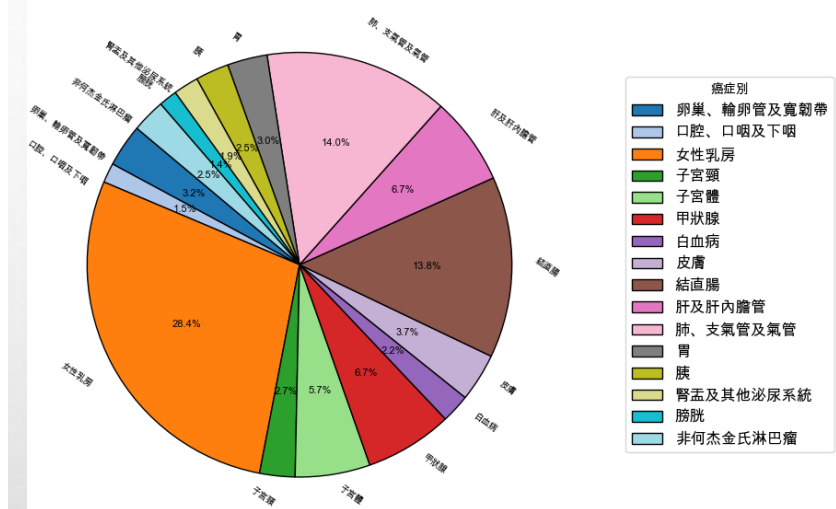
全 的癌症別發生數分佈 (2017至2021年)



男 的癌症別發生數分佈 (2017至2021年)



女 的癌症別發生數分佈 (2017至2021年)



在此次結果中，剔除癌症發生率少於1%的結果。

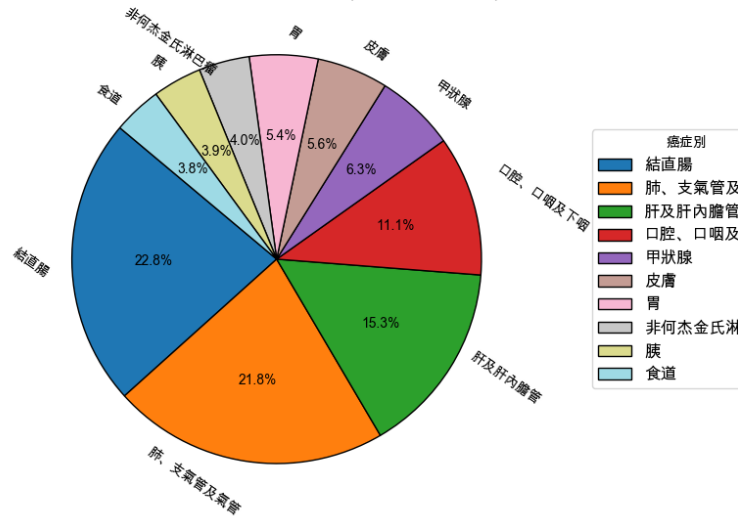
篩選後的結果，可以發現在各性別中，癌症發生的情形主要以肺、肝、口腔、結直腸、攝護腺、女性乳房居多



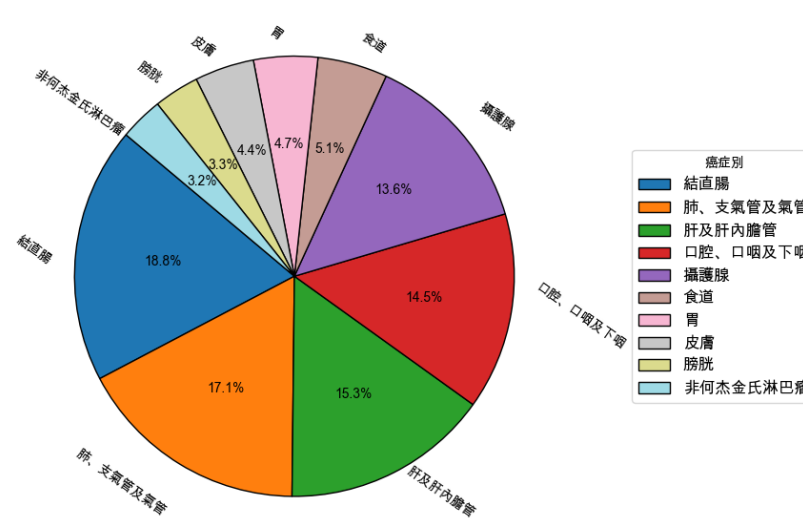
02. 圖表說明-資料分析(2017-2021年癌症分佈情形)

製圖說明：分析在2017年至2021年中，各性別中發生的前十大癌症類別有哪些。

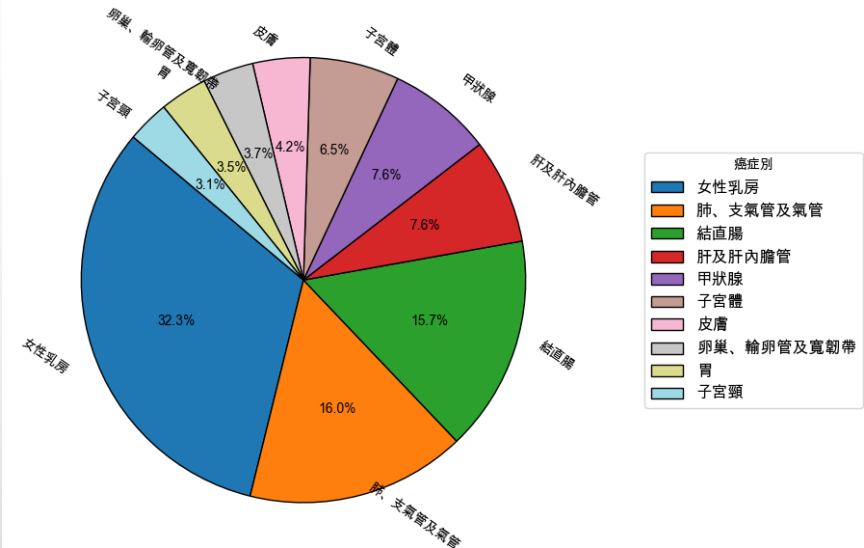
全的前十大癌症別發生數分佈 (2017至2021年)



男的前十大癌症別發生數分佈 (2017至2021年)



女的前十大癌症別發生數分佈 (2017至2021年)



在此次結果中，篩選出各性別的所有癌症中，比例佔最多數的結果。

全性別：結直腸(22.8%)、肺(21.8%)、肝(15.3%)、口腔(11.1%)、甲狀腺(6.3)

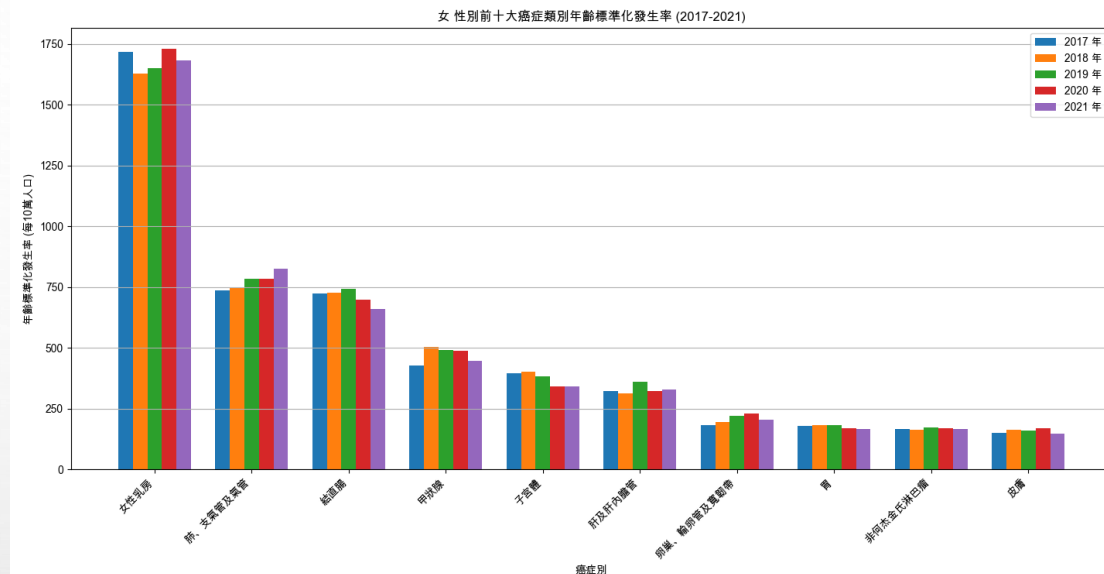
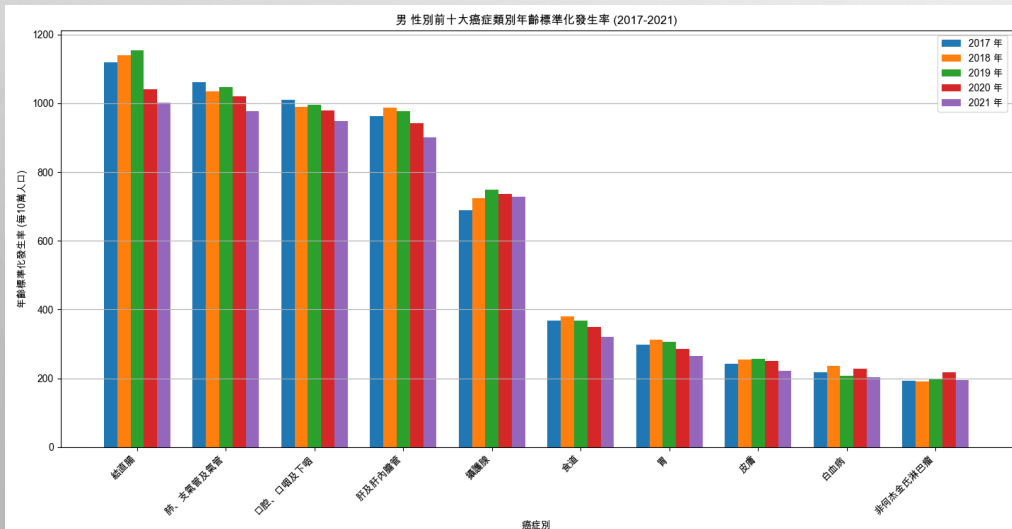
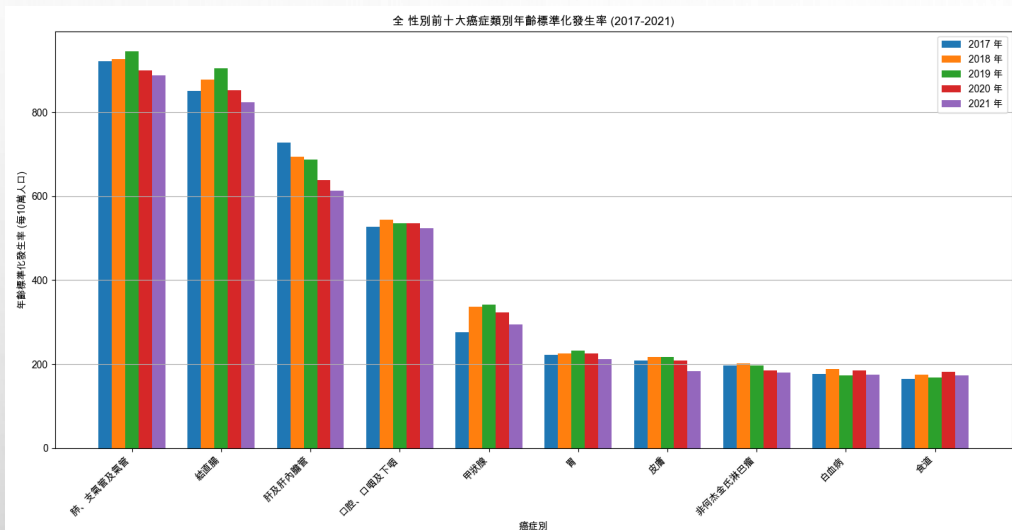
男性：結直腸(18.8%)、肺(17.1%)、肝(15.3%)、口腔(14.5%)、攝護腺(13.6%)

女性：女性乳房(32.3%)、肺(16.0%)、結直腸(15.7%)、肝與甲狀腺(7.6%)、子宮(6.5%)



02. 圖表說明-資料分析(2017-2021年癌症分佈情形)

製圖說明：觀察在2017年至2021年中，各性別中發生的前十大癌症的變化與每10萬人發生機率。



最後，2017年到2021年間，觀察在不同性別中，每10萬人口發生的比例

全性別：每10萬人就會有肺癌、結直腸癌、肝癌會超過600人左右發生

男性：每10萬人就會有肺癌、結直腸癌、口腔、肝癌會超過900人左右發生

女性：每10萬人就會有女性乳房、肺癌、結直腸、甲狀腺會超過500人左右發生，其中女性乳房最為明顯約1600人左右



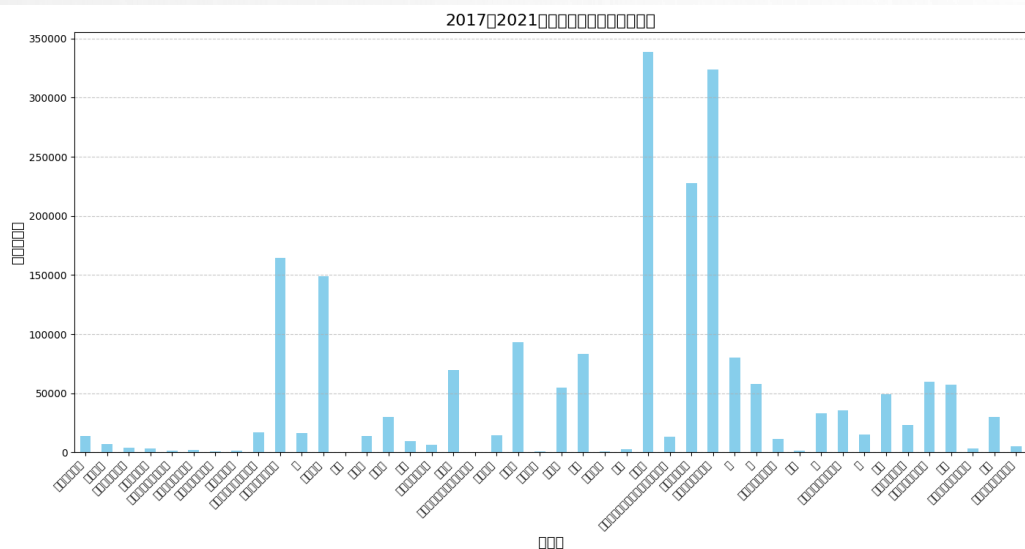
03. 問題與解決



03. 問題與解決

問題一、plt圖表呈現時，無法轉換並輸出為中文字型

網路上搜尋相關解決辦法



```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import matplotlib.font_manager as fm
4 import numpy as np
5
6
7 # 設定中文字型
8 def plt_chinese():
9     # 使用系統中可用的中文字型
10    available_fonts = fm.findSystemFonts(fontpaths=None, fonttext='ttf')
11    for font in available_fonts:
12        if 'SimHei' in font or 'Microsoft JhengHei' in font:
13            plt.rcParams['font.sans-serif'] = [font.split('/')[1].replace('.ttf', '')] # 提取字型名稱
14            break
15    else:
16        plt.rcParams['font.sans-serif'] = ['Arial Unicode MS'] # 避免找不到字型時使用的備用字型
17    plt.rcParams['axes.unicode_minus'] = False # 顯示負號
18
19 # 設定中文字型
20 plt_chinese()
21
```

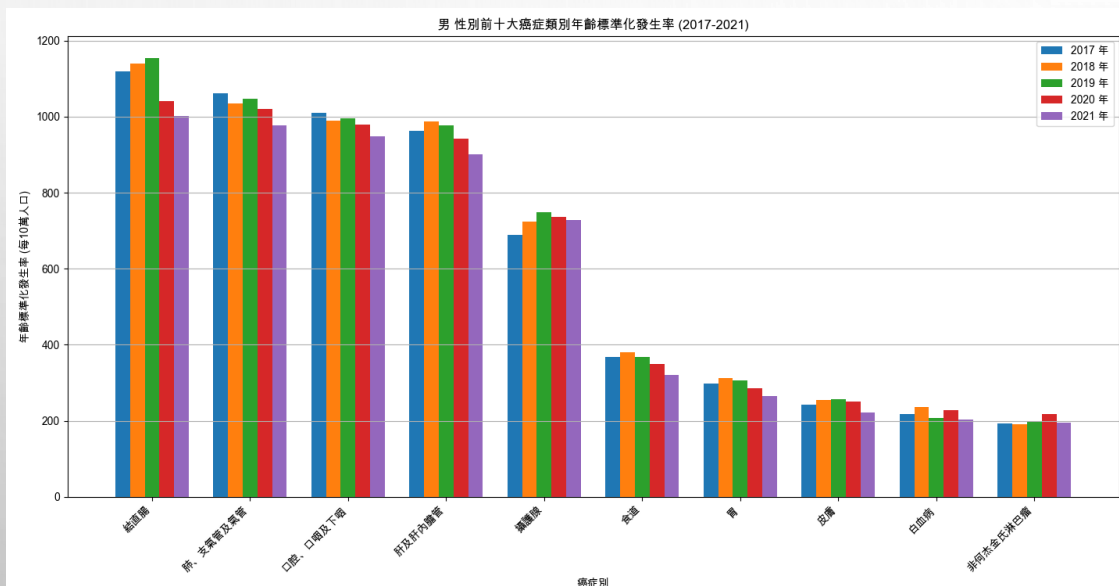
1. 使用matplotlib.font_manager 模組進行中文字型設定
2. 確認系統內中文字型格式與名稱，



03. 問題與解決

問題二、多筆數據結果顯示不一致，且不明白如何並排呈現結果

網路上搜尋相關解決辦法



```
# 確保每個性別和年份的數據長度一致
def ensure_length_consistency(df, top_n=10):
    cancer_types = df['癌症別'].unique()
    if len(cancer_types) < top_n:
        missing_count = top_n - len(cancer_types)
        missing_rows = pd.DataFrame({
            '癌症別': [''] * missing_count,
            '年齡標準化發生率 WHO 2000世界標準人口 (每10萬人口)': [0] * missing_count,
            '癌症診斷年': [df['癌症診斷年'].iloc[0]] * missing_count,
            '性別': [df['性別'].iloc[0]] * missing_count
        })
    df = pd.concat([df, missing_rows], ignore_index=True)
    return df
```

1. 使用consistency 定義數據，增加數據的一致性，確保癌症別數量維持為10類
2. 確認是否在癌症別中，年齡發生的結果是否有空值，確保數據列一致
3. 以pd.concat將2017至2021年數據合併

04. 總結與感想





04. 總結與感想

- 在癌症統計分析結果中，由於資料數據龐大，故只選擇了近5年的結果進行分析，透過分析也可以瞭解到男性與女性的癌症發生率與分佈結果可能會有些許差異，但整體來看主要還是以肺、肝、結直腸癌為主。
- 未來可以透過不同的分析與比較使醫療與基礎研究人員可以進一步瞭解癌症發生的原因及如何早期預防的情形。
- 最後，在程式碼執行的過程中，有些不清楚如何執行的部分，依舊是透過網路資料搜尋讓自己可以完成本次的專案作品，未來也會持續加強程式語法的基礎觀念與知識，讓自己可以更得心應手。



報告結束

感謝各位聆聽！

指導老師：簡志峰 老師

報告人員：王國昌 Brad

報告日期：2024/07/28

