B277873

B277873

2024-12-09

Contents

Introduction

Audience

• The audience of the report includes policymakers in Portugal, Ministers of government departments, and secretaries of state (Machado et al. 2024).

Data Story

The public appears to be preferring medical consultation through the internet as `Teleconsultations` usage in Portugal has dramatically risen after the COVID-19 pandemic ("OECD Health Data: Health Care Utilisation" n.d.). Analysis of Organisation for Economic Co-operation and Development (OECD) healthcare data from 2017 to 2022 reveals that `Teleconsultations` per person increased sharply from less than 1 in 2019 to over 2 by 2021 in a rapidly escalating fashion ("OECD Health Data: Health Care Utilisation" n.d.). This upward trend highlights how teleconsultations is becoming a mandatory component of healthcare delivery, and it appears to be staying with us for some time which requires policy change to meet the requirements to both mitigate some of the side effects of such shift in the mode of healthcare delivery and to maximise benefits (OECD 2023). In order to inform policymakers to adopt better policies that are costumed to the needs of the healthcare sector in Portugal, analysis of consultations data and internet access in Portugal is required.

Context

The rapid adoption of teleconsultation during the pandemic was driven by the need for accessible and safe healthcare solutions (Marshall et al. 2022). Policymakers need to investigate and update policies and regulations to accommodate the technical requirements that are needed to fully adopt the move towards telemedicine practices as a separate mode of healthcare delivery through teleconsultation (Machado et al. 2024). Since access to the internet is mandatory for a satisfactory result, an analysis of the percentage of households that have internet access in Portugal and its correlation to the rise in teleconsultation and to the incidence of COVID-19 is required (Marshall et al. 2022).

Data Sources and Data Wrangling

Data Sources

The data for this report is derived from the Organisation for Economic Co-operation and Development (OECD) Healthcare Utilisation dataset ("OECD Health Data: Health Care Utilisation" n.d.) and OECD ICT Access and Usage by Households dataset("ICT Access and Usage by Households and Individuals" n.d.).

Data Gaps and Missing Information

Missing Data:

- Some years or consultation types has missing observations ("OECD Health Data: Health Care Utilisation" n.d.).
- Efficacy and health outcomes in relation to `Teleconsultations` data is missing, limiting the ability to perform detailed trend analysis.

Ambiguities:

- Certain variables are not clearly defined in the dataset ("OECD Health Data: Health Care Utilisation" n.d.).
- There are not enough detailed data to show the correlation between different demographics and age groups' reliance on `Teleconsultations`("OECD Health Data: Health Care Utilisation" n.d.).

Contextual Data:

Patient demographics or healthcare outcomes associated with the consultations were not included, which could provide additional insights ("OECD Health Data: Health Care Utilisation" n.d.).

Data Wrangling

In order to highlight the need for data analysis of both teleconsultations and in person consultations, we need to highlight the trend for both variables first and how they compare to each other across time. Conventional `In-person` consultation is the standardised mode of healthcare delivery, thus, comparing its levels to `Teleconsultations` per person highlights the significance of the latter. Since data visualisation is aimed to highlight data for policymakers in portugal, we are only interested in consultations data from Portugal. A line graph is perfect to highlight the differences between two numerical data across time. COVID-19 is arguably the main reason for the spike in teleconsultation after 2020 so it might be associated for the change in its levels. Timeline chosen must therefore include data from before and after the global pandemic time period. This is also to communicate the compound effect of both the increase of household internet access and COVID-19 in increasing the public's appetite for `Teleconsultations`.

Filtering data for the first plot

Data from the OECD dataset was derived to create the raw data frame that is used for initial data filtering. Initial viewing of the data is conducted to ensure observations data types. In order not to mix results, only `In-person` and `Teleconsultations` data were filtered to isolate it from the `Total` consultations per person data. observations with null values were filtered out of the data frame then to facilitate plotting in the next step. Although it is not required for plotting, the `Reference area` was also selected to preserve coherence of the newly filtered data frame. The period between 2017 and 2022 was selected to illustrate the before and after effect of COVID-19 on `Teleconsultations`(Marshall et al. 2022).

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Load data
raw_data <- read_csv("OECD.ELS.HD,DSD_HEALTH_PROC@DF_KEY_INDIC,+all.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 531602 Columns: 54
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (44): STRUCTURE, STRUCTURE_ID, STRUCTURE_NAME, ACTION, REF_AREA, Referen...
## dbl  (2): TIME_PERIOD, OBS_VALUE
## lgl  (8): Time period, Observation value, OBS_STATUS3, Observation status 3,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(raw_data)
```

```
## Rows: 531,602
## Columns: 54
## $ STRUCTURE              <chr> "DATAFLOW", "DATAFLOW", "DATAFLOW", "DATAFLOW",~
## $ STRUCTURE_ID           <chr> "OECD.ELS.HD:DSD_HEALTH_PROC@DF_KEY_INDIC(1.1)"~
## $ STRUCTURE_NAME         <chr> "Healthcare utilisation", "Healthcare utilisati~
## $ ACTION                 <chr> "I", "I", "I", "I", "I", "I", "I", "I", "I", "I~
## $ REF_AREA               <chr> "AUS", "AUS", "AUS", "AUS", "AUS", "AUS", "AUS"~
## $ `Reference area`       <chr> "Australia", "Australia", "Australia", "Austral~
## $ MEASURE                <chr> "CONSULT", "CONSULT", "CONSULT", "CONSULT", "CO~
## $ Measure                <chr> "Consultations", "Consultations", "Consultation~
## $ UNIT_MEASURE           <chr> "CN_PS", "CN_PS", "CN_PS", "CN_PS", "CN_PS", "C~
## $ `Unit of measure`      <chr> "Consultations per person", "Consultations per ~
## $ MEDICAL_PROCEDURE      <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Medical procedure`    <chr> "Not applicable", "Not applicable", "Not applic~
## $ OCCUPATION             <chr> "OC221", "OC221", "OC221", "OC221", "OC221", "O~
## $ Occupation             <chr> "Medical doctors", "Medical doctors", "Medical ~
## $ DIAGNOSTIC_TYPE        <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Diagnostic category`  <chr> "Not applicable", "Not applicable", "Not applic~
## $ PROVIDER               <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ Provider               <chr> "Not applicable", "Not applicable", "Not applic~
## $ FUNCTION               <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Healthcare function`  <chr> "Not applicable", "Not applicable", "Not applic~
## $ MODE_PROVISION         <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Mode of provision`    <chr> "Not applicable", "Not applicable", "Not applic~
## $ CARE_TYPE              <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Care type`            <chr> "Not applicable", "Not applicable", "Not applic~
## $ HEALTH_FACILITY        <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Medical technology`   <chr> "Not applicable", "Not applicable", "Not applic~
## $ AGE                    <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ Age                    <chr> "Not applicable", "Not applicable", "Not applic~
## $ SEX                    <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ Sex                    <chr> "Not applicable", "Not applicable", "Not applic~
## $ DATA_SRCE              <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Data source`          <chr> "Not applicable", "Not applicable", "Not applic~
## $ WAITING_TIME           <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Waiting time`         <chr> "Not applicable", "Not applicable", "Not applic~
## $ DISEASE                <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ Disease                <chr> "Not applicable", "Not applicable", "Not applic~
## $ CANCER_SITE            <chr> "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z", "_Z",~
## $ `Cancer site`          <chr> "Not applicable", "Not applicable", "Not applic~
## $ CONSULTATION_TYPE      <chr> "_T", "_T", "_T", "_T", "_T", "_T", "_T", "_T",~
## $ `Consultation type`    <chr> "Total", "Total", "Total", "Total", "Total", "T~
## $ TIME_PERIOD            <dbl> 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022,~
## $ `Time period`          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ OBS_VALUE              <dbl> 6.8, 7.0, 7.1, 7.3, 7.3, 7.4, 7.8, 8.5, 7.4, 0.~
## $ `Observation value`    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ OBS_STATUS             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ `Observation status`   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ OBS_STATUS2            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ `Observation status 2` <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ OBS_STATUS3            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
## $ ‘Observation status 3‘ <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ UNIT_MULT              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ ‘Unit multiplier‘      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ DECIMALS               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Decimals               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```r
# Filter data for Portugal and relevant years
portugal_data <- raw_data %>%
  filter(
    `Reference area` == "Portugal",
    `Unit of measure` == "Consultations per person",
    `Consultation type` %in% c("Teleconsultations", "In-person"),
    TIME_PERIOD >= 2017 & TIME_PERIOD <= 2022,
    Occupation == "Medical doctors",
    !is.na(OBS_VALUE)
  ) %>%
  select(`Reference area`,
         OBS_VALUE,
         TIME_PERIOD,
         `Consultation type`)
```

Visualization: Line Graph

A point of reference was created first to higlight the point of the spike in `Teleconsultations` value. Plotting is straight forward here, since necessary filtering was completed in the previous step. color coding is used to highlight the different consultation type levels . There is no need to alter the scale of the Y or X axis as both lines represent data of the same type and are approxiamte in value. However, a clear title is needed to guide the reader through the context of the plot.

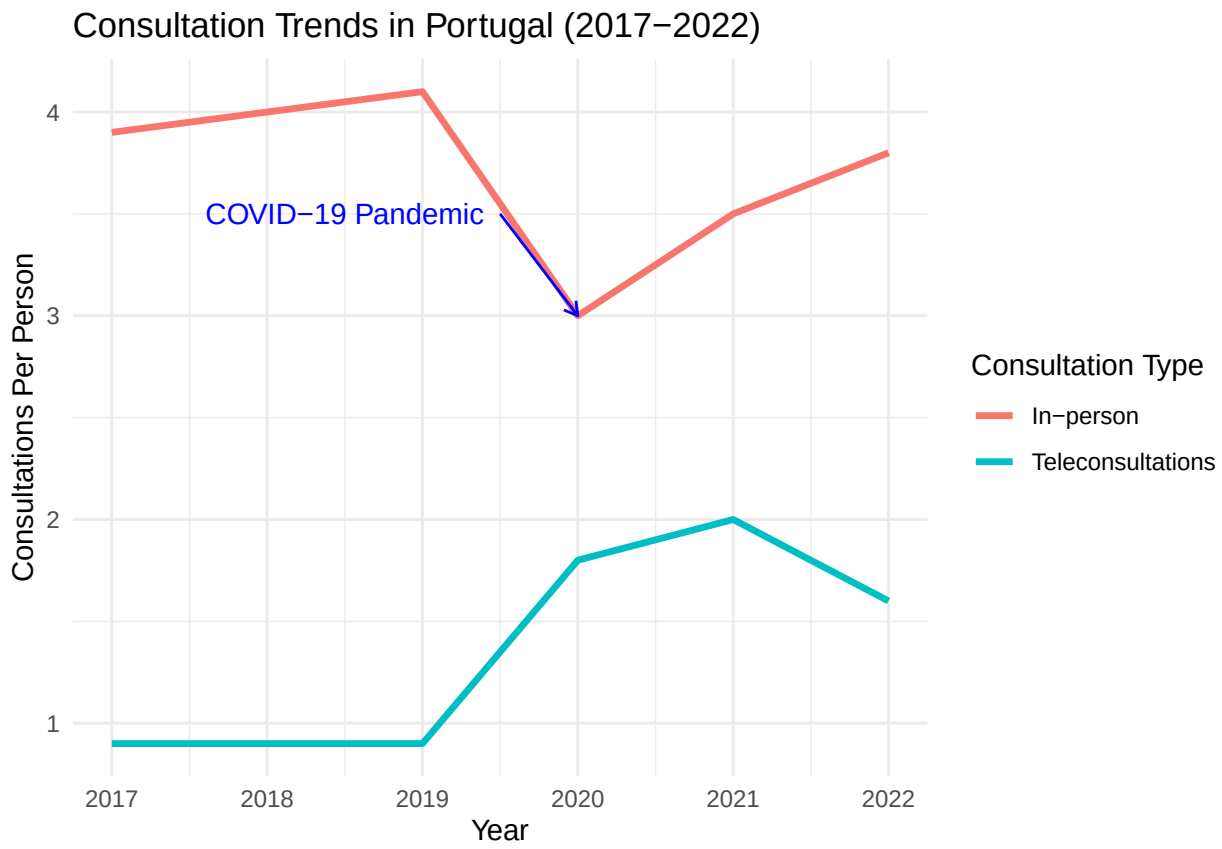
```r
library(ggplot2)

max_year <- 2020
  max_value <- max(portugal_data $OBS_VALUE [portugal_data $TIME_PERIOD == max_year], na.rm = TRUE)

# Line Graph
portugal_data %>% ggplot(aes(x = TIME_PERIOD, y = abs(OBS_VALUE), color = `Consultation type`)) +
  geom_line(size = 1.2) +
  annotate("text", x = max_year - 1.5, y = max_value + 0.5,
             label = "COVID-19 Pandemic", size = 4, color = "blue") +
    # Add arrow pointing to the max value
  annotate("segment", x = max_year - 0.5, xend = max_year,
             y = max_value + 0.5, yend = max_value,
             arrow = arrow(length = unit(0.2, "cm")), color = "blue") +
  labs(
    title = "Consultation Trends in Portugal (2017-2022)",
    x = "Year",
    y = "Consultations Per Person",
    color = "Consultation Type"
  ) +
  theme_minimal()
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
```

```
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



Filtering internet access data for second plot

Data is derived from the OECD ICT Access and Usage by Households dataset to create the raw data frame that is filtered to focus of data on `Households with Internet access at home` between 2017-2022 in Portugal. To keep the data tidy, data is filtered by users `Income.group` and gegraphical area to include the total users of the country. Both internet access and consultation per person per year data were joined to prepare for plotting.

```
raw_access <- read.csv("raw_intaccess.csv")

portugal_access_data <- raw_access %>%
  filter(
    `Reference.area` == "Portugal",
    `Measure` == "Households with Internet access at home",
    `Income.group` == "Total",
    TIME_PERIOD >= 2017 & TIME_PERIOD <= 2022,
    `Geographical.area` == "Total",
    !is.na(OBS_VALUE)
  )
portugal_access_data <- portugal_access_data %>%
```

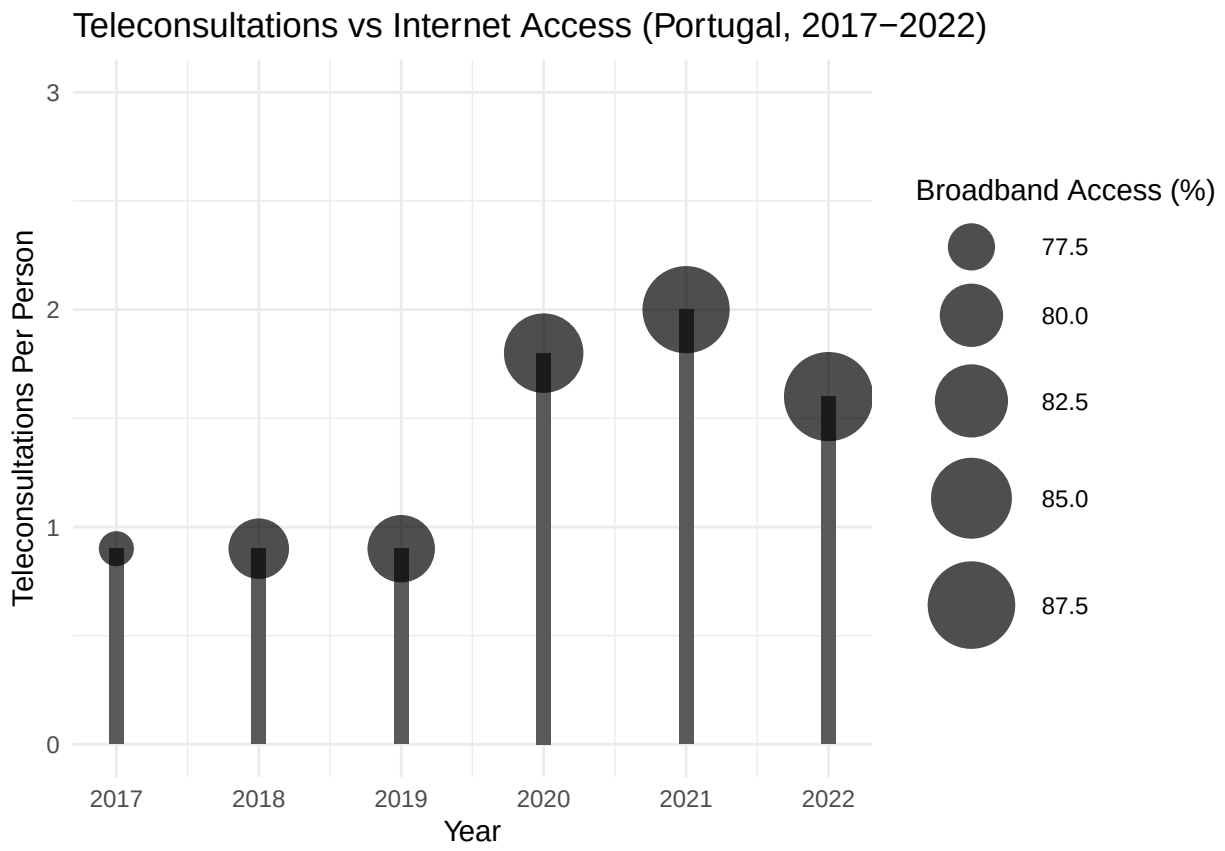
```
  rename(PERCENTAGE = OBS_VALUE)

joined_data <- portugal_data %>%
  filter(`Consultation type` == "Teleconsultations") %>%
  left_join(portugal_access_data,
            by = "TIME_PERIOD")
```

Plotting internet access data

To better visualise the association of internet access percentage and `Teleconsultations` rate, a combination of bar plot and point plot is used. The x-axis represents the time period, y-axis indicates the `Teleconsultations` per person per year, while each point's size corresponds to the percentage of households with internet access across Portugal, providing a visual representation of internet accessibility in correspondence to `Teleconsultation` rates.

```
joined_data %>% ggplot(aes(x = TIME_PERIOD, y = OBS_VALUE)) +
  geom_bar(stat = "identity",
           width = 0.1,
           position = "identity",
           alpha = 2) +
  geom_point(aes(size = PERCENTAGE), alpha = 0.7) +
  scale_size_continuous(name = "Broadband Access (%)",
                        range = c(5.5,15)) +
  scale_y_continuous(
    name = "Teleconsultations Per Person",
    limits = c(0, 3)
  ) +
  labs(
    title = "Teleconsultations vs Internet Access (Portugal, 2017-2022)",
    x = "Year",
    y = "Teleconsultations Per Person"
  ) +
  theme_minimal() +
  theme(
    legend.position = "right")
```



Recommendations and Discussion

Recommendations

1. **Enhance telecommunications infrastructure**:
   - The data calls for needed investments in telecommunication infrastructure for healthcare delivery purposes (Séror 2002).
   - As teleconsultation becomes more of a mainstream method for healthcare delivery, more attention to issues related to privacy and cybersecurity is needed and continuous evaluation programs for the pivacy of internet connections may be required.
2. **Hybrid Consultation** :
   - A combination of teleconsultation and in-person services protocols developments to improve patient access and optimize resource allocation may be needed(OECD 2023).
   - Push for healthcare workforce training programs to adopt the new trend (Modi et al. 2018).
3. **Monitor Effectiveness**:
   - Implement audits and regular assessments of teleconsultation's impact on patient outcomes and healthcare efficiency (Miranda et al. 2023).
   - continuous reserch is to be conducted to investigate patients opinions and practices to improve user experiences (Carrillo De Albornoz, Sia, and Harris 2022).

Discussion

Data analysis shows a clear escalating trend in `Teleconsultations` rate per person in Portugal, especially during crises. This upward trend can be associated with the crises itself or the increasing availability of internet service which points out the importance of such services for the healthcare sector. Policymakers should pay attention for the phenomenon of teleconsultation and adopt policies to improve communications infrastructure across the country in rural areas where internet might not be easily accessed to maximise benefits of `Teleconsultations` (Machado et al. 2024). Telecommunication firms should work hand in hand with policy makers to facilitate the adoption of such policies (Machado et al. 2024). Although the data analysis highlights the overall big picture, further analysis is required to compare areas with households internet access percentages that are lower than other areas to highlight commonalities and mitigate common issues. Analysis of the type of internet connection technology can further guide policy-making to target the right sector for future investments in communication infrastructure (Marshall et al. 2022).

Reflection

Description

The project requires a new set of skills and a new approach towards problem solving than what I am used to. Although I have attended some of CS50 classes online to be more familiar to how programmers think prior to this course, R seems to have a more steep learning curve. The program's curriculum that I need to get my programming skills from is unorganised as it mixes general data science material with programming training in the same week, which does not fit my learning style of focusing on one subject at a time.

Feelings

I felt lost in the complexity of the program's material, the data set and how to data wrangle it. I did not know where to start the project and I got more frustrated the more time passed by without me achieving any progress. Even though I finished most of the program's material, I did not feel confident with my data wrangling skills.

Evaluation

since the material of the program is not well organised. I decided to re-organise it after week 8. I started going through the theoretical material of data science without going through any coding material. Although this did not really improve my programming skills, it cleared my mind from the ambiguity towards the program's material to focus only on programming then. When I started the coding material, a combination of using ChatGPT with the Feynman technique appears to be giving much better results than relying on tutorial explanation videos (Ambion et al. 2020). Using ChatGPT to write code from a text prompt was not efficient as it starts to hallucinate easily and goes off topic. However, learning to code everything by hand after following tutorials was time consuming and very un-organised, as the tutorials go through multiple concepts rapidly. Instead of following the tutorials explanation, I copy and paste code chunks and ask ChatGPT to explain it in bullet points and simple language and then try to teach the chatbot what I have learned using my own words and see if it approves it or not. Using this technique made it much easier to create mind maps to facilitate navigating through skills and syntax that I need to finish the project.

Analysis

Using ChatGPT as a teacher and a student at the same time appears to be an effective method of applying reflective active learning practice along the way. By simplifying complex concepts and abstracting them

using the Feynman technique (Ambion et al. 2020), making mind-maps and learning syntax in a hierarchical structure becomes easier. Since ChatGPT creates comparatively clean code, learning this way makes it much easier to learn good coding practices passively without getting used to bad ones.

Conclusion

The project experience made me invent new styles of learning to best suit my thinking process when learning a new skill. After honing my skills in data wrangling, getting to the execution phase of the project becomes easier, especially with a dictionary provided to explain the data set I am using. ChatGPT proved to be a better tool of learning coding than writing the whole code from ground zero.

Action Plan

Future projects will focus on:

- adopting new methods of active learning techniques to develop new skills.
- Expanding skills in advanced visualisation techniques, including new plot styles and interactive ones.
- focusing on data visualisation and custom data wrangling according to data visualisation needs.

References

Ambion, Ronnel Ian A, Rainier Santi C De Leon, Alfonso Pio Angelo R Mendoza, and Reinier M Navarro. 2020. "The Utilization of the Feynman Technique in Paired Team Teaching Towards Enhancing Grade 10 ANHS Students' Academic Achievement in Science." In *2020 IEEE Integrated STEM Education Conference (ISEC)*, 1–3. IEEE.

Carrillo De Albornoz, Sara, Kah-Ling Sia, and Anthony Harris. 2022. "The Effectiveness of Teleconsultations in Primary Care: Systematic Review." *Family Practice* 39 (1): 168–82. https://doi.org/10.1093/fampra/cmab077.

"ICT Access and Usage by Households and Individuals." n.d. OECD. Accessed December 9, 2024. https://doi.org/10.1787/b9823565-en.

Machado, Sara, Ana Moura, Francisca Vargas Lopes, Diogo Marques, and Luis Sa. 2024. "A Health System Assessment Approach to Analysis of Political Parties' Health Proposals, Portugal." *Bulletin of the World Health Organization* 102 (11): 820–27. https://doi.org/10.2471/BLT.24.291831.

Marshall, Victoria K., Melody N. Chavez, Tina M. Mason, and Dinorah Martinez Tyson. 2022. "The Use of Telehealth to Provide Continuity of Cancer Care During the COVID-19 Pandemic: Advantages, Disparities, and Implications to Promote Health Equity." *Hospital Topics* 100 (4): 166–76. https://doi.org/10.1080/00185868.2021.1942360.

Miranda, Rafael, Mónica Duarte Oliveira, Filipa Matos Baptista, and Isabel Albuquerque. 2023. "Telemonitoring in Portugal: Where Do We Stand and Which Way Forward?" *Health Policy* 131 (May): 104761. https://doi.org/10.1016/j.healthpol.2023.104761.

Modi, Parth K., David Portney, Brent K. Hollenbeck, and Chad Ellimoottil. 2018. "Engaging Telehealth to Drive Value-Based Urology." *Current Opinion in Urology* 28 (4): 342–47. https://doi.org/10.1097/MOU.0000000000000508.

OECD. 2023. *The COVID-19 Pandemic and the Future of Telemedicine*. OECD Health Policy Studies. OECD. https://doi.org/10.1787/ac8b0a27-en.

"OECD Health Data: Health Care Utilisation." n.d. OECD. Accessed December 9, 2024. https://doi.org/10.1787/data-00542-en.

Séror, Ann C. 2002. "Internet Infrastructures and Health Care Systems: A Qualitative Comparative Analysis on Networks and Markets in the British National Health Service and Kaiser Permanente." *Journal of Medical Internet Research* 4 (3): e21. https://doi.org/10.2196/jmir.4.3.e21.