CERVICAL CANCER SCREENING AND MORTALITY RATE ANALYSIS IN OECD COUNTRIES

Exam Number: B244416

2023-12-11

Table of Contents

knitr::opts\_chunk$set(include = TRUE)

# Introduction

Worldwide, Cervical Cancer is the fourth most common cancer among women and the second leading cause of death in women aged 15–44 years, with an estimated 527,624 new cases and 265,653 deaths in 2012 (Bruni et al., March 24, 2015). Cervical cancer is among the most preventable cancers: besides being prevented through HPV vaccination, pre-cancerous lesions can easily be detected through screening before they become cancer (WHO, 2023). When screening detects pre-cancerous lesions, these can be treated, and cancer avoided. While all OECD countries actively engage in screening for cervical cancer, screening rates within these countries vary disproportionately (Weisband, Y. L., 2021). Despite a decline in mortality rates due to cervical cancer over time in some countries, disparities persist within the OECD countries due to various factors such as campaigns, guidelines, political will, and policy shifts regarding cervical cancer that affect screening rates (Devaux, M. 2014).

# Context

Cervical cancer screening rates and mortality rates due to cervical cancer have been a subject of concern and interest including OECD countries. Disparities in screening rates and mortality rates continue to grow, and understanding the data trends and patterns is crucial for informed decision-making and policy formulation. In this data storytelling exercise, we will utilize cervical cancer screening rate and cervical cancer mortality rate datasets to explore, and analyse relationships between the two variables from 2010 to 2022 for selected OECD countries.

# Audience

This report is prepared for the World Health Organization (WHO) senior management team who are responsible for making decisions, and policies that impact public health specifically preventive health. To do this, these stakeholders must be adequately informed with clear insights that will guide their decisions, and consequently lead to appropriate actions to address the underlying causes. The change in audience did not necessarily result in selecting a different topic of the story. However, visuals have somewhat been changed to enrich understanding of the data story.

# Dataset and limitations.

The datasets include information on cervical cancer screening rates and mortality rates across OECD countries from 2010 to 2022. The data dictionary below provides a summary of the variables, variable types, as well as descriptions.

# KEY VARIABLES

The following are the key variables used on the dataset.

## Cervical cancer screening rates dataset.

Year (Integer/date) : Represents the period in years in which the screening rates were recorded

Country (Character): Represents the country names where cervical cancer screening rates are recorded

Screening rates (Double): Represents the percentage or percentage of cervical cancer screening in the selected OECD country, and year.

## Cervical cancer mortality rate dataset

Year (Integer/date): Represents the time in years when the mortality rates were recorded.

Country (Character): Represents the country names where cervical cancer mortality rates were recorded.

Mortality rates (Double): Shows the percentage or rates of cervical cancer as a cause of mortality in selected OECD countries.

# Limitations of the dataset

The data story has a handful of limitations. Variability in data quality and reporting standards among different countries has been noted during data wrangling, affecting the comparability of results. Adding to this, some countries or specific years have missing values or incomplete data, which could impact the accuracy of analyses. Furthermore, differences in healthcare systems and cervical cancer screening programs among OECD countries may supposedly introduce complexities in interpreting the data. For instance, some countries use organized screening programs, whereas other countries use opportunistic screening.

While high cervical cancer screening rates, especially those close to or above 80% are good indicators of progress, without other compounding variables the screening rates will give us a limited understanding. One of the hypotheses in this data story is that the high screening rates result in cervical cancer mortality reduction overtime. Therefore, I have extracted data from the OECD website on cervical cancer mortality rate to form part of analysis. Additionally, I would want to appreciate the impact of cervical mortality with respect to age groups, however this could not be possible due to the structure of the data (The data structure only includes one age band from 20-69 years).

# Objectives of the data story

* To examine which countries are doing well in terms of screening and which may need improvements.
* To determine the relationship between cervical cancer screening and mortality rate.

# Missing elements from the data story

The story did not consider the HPV vaccination program that may also have profound impact on cervical cancer mortality rate trend. HPV vaccine rates which are considered a primary preventive measure before screening would affect the mortality as well as the cervical cancer screening rate trend as confirmed by the existing literature. A study done in in the US showed decreased cervical cancer incidence and mortality among women and girls aged 15 to 24 years after HPV vaccine introduction (Tabibi, Tara et al, 2021). Quick to note is also the fact that the data is not disaggregated by age categories, this aspect makes it impossible to appreciate the nuanced impact screening rate has on age, and to categorize which age groups have the highest mortality rates to inform custom made interventions. Finally, to improve the data story I would use the survey data since this type of data considers the socio-economic factors that influence individual choices to participate in cervical cancer screening.

# Data loading, wrangling and coding

Data Wrangling is the process of getting data into a useful form for visualisation, summary, and modelling (Heer, J, 2018). . Data wrangling is important in data science particularly in the health space because data seldomly comes in a manner and form that fits the defined means of analysis. To this end, cervical cancer screening and mortality rates datasets were also be subjected to wrangling before further analysis

cervical\_cancer\_screening\_data<-read\_csv("ccs\_screening.csv")#Data source: https://stats.oecd.org/index.aspx?queryid=30159  
cervical\_cancer\_mortality\_data<-read\_csv("causes\_mortality\_cc.csv")# Data source: https://stats.oecd.org/index.aspx?queryid=30121

#Cleaning the cervical cancer screening dataset to turn variable names to lower cases and place underscores in spaces and dashes for ease of analysis. The data was then pivoted wider to enhance analysis  
ccs\_cleaned<-cervical\_cancer\_screening\_data %>%   
 clean\_names() %>%   
 pivot\_wider(names\_from = c(variable,   
 measure),  
 values\_from = value,  
 id\_cols = c(year,   
 country)) %>%   
 clean\_names()

#The cleaned data was filtered to remove other cancer types and remain with cervical cancer, year, and country, only,   
ccs\_filtered<-ccs\_cleaned %>%   
 select(country,year,cervical\_cancer\_screening\_programme\_data\_percent\_of\_females\_aged\_20\_69\_screened) %>%   
 filter(complete.cases(.))

# The filtered data, was then cleaned again, anf further filtered for complete cases to remove all missiong values  
ccs\_filtered <- ccs\_cleaned %>%   
 select(country,  
 year,  
 cervical\_cancer\_screening\_programme\_data\_percent\_of\_females\_aged\_20\_69\_screened) %>%   
 rename(screening\_rate=cervical\_cancer\_screening\_programme\_data\_percent\_of\_females\_aged\_20\_69\_screened) %>%   
 filter(complete.cases(.))

#Cleaning was also done for cervical cancer mortality rate dataset  
ccm\_cleaned<-cervical\_cancer\_mortality\_data %>%   
 clean\_names() %>%   
 pivot\_wider(names\_from = c(variable,   
 measure),  
 values\_from = value,  
 id\_cols = c(year,   
 country)) %>%   
 clean\_names()

#The cervical cancer dataset was also filtered  
ccm\_filtered<-ccm\_cleaned%>%   
 select(country,  
 year,   
 malignant\_neoplasms\_of\_cervix\_uteri\_deaths\_per\_100\_000\_females\_standardised\_rates) %>%   
 rename(Deaths\_per\_100000\_females=malignant\_neoplasms\_of\_cervix\_uteri\_deaths\_per\_100\_000\_females\_standardised\_rates) %>%   
 filter(complete.cases(.))

#After cleaning and filtering both datasets, the two datasets were then joined using the left join command.  
cc\_joined<-ccs\_filtered %>%   
 left\_join(ccm\_filtered, by = c("country"  
 ,"year")) %>%   
 filter(complete.cases(.)) #%>%   
 #view()

# Data Visualisation

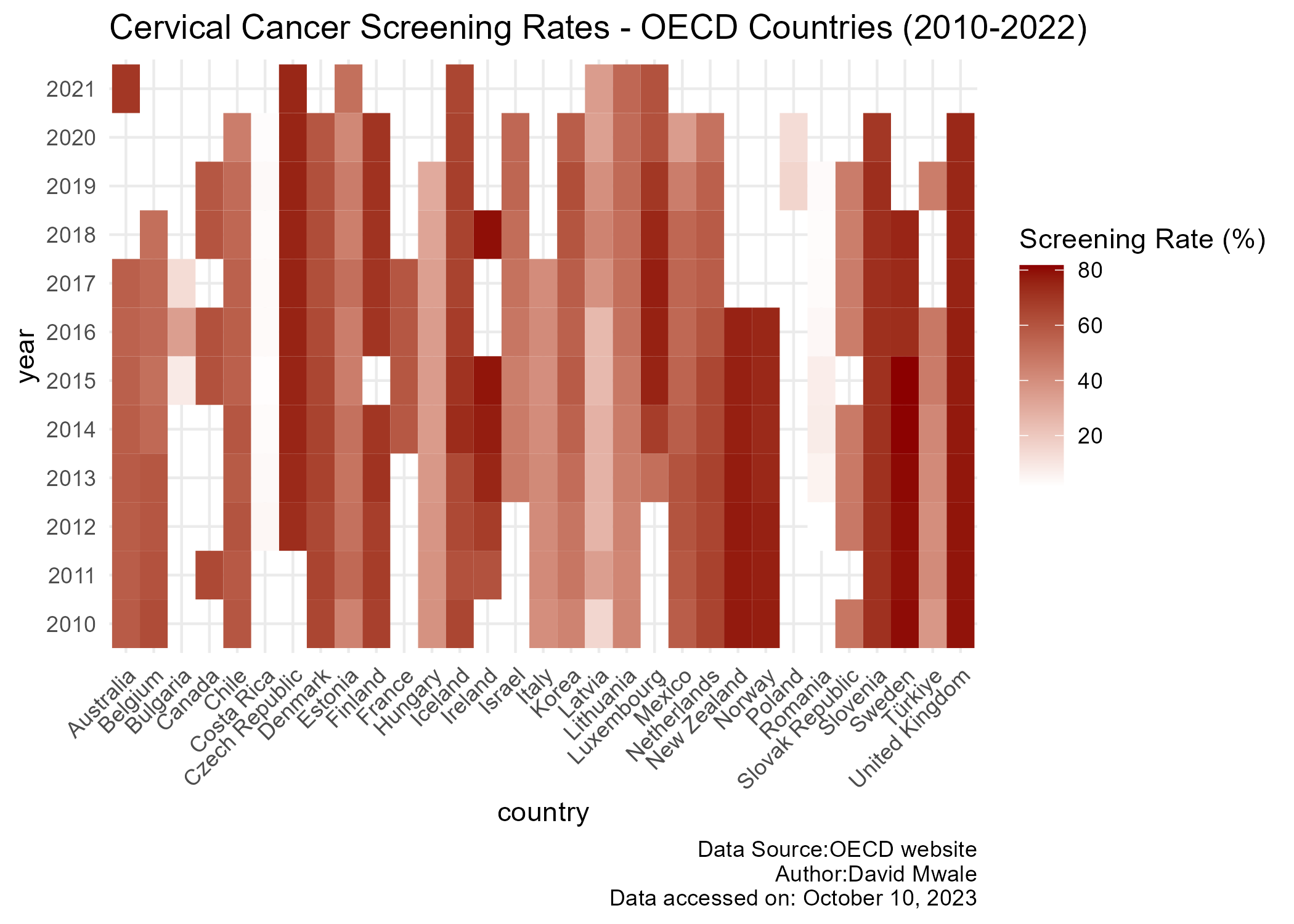
In this data story firstly, I will use the geomtile chart, selecting this was an easy option as it explicitly shows concentration, and readers can effortlessly identify areas of interest such as countries with the most screening rates over time. The second choice of visuals used in this data story is geompoint (scatter plot) to illustrate the distribution of mortality and cervical cancer screening rates.

## Discussion and Analysis

#The combined datasets saved as CC\_joined is then used to make an analysis  
p6<-cc\_joined %>%   
 filter(complete.cases(.)) %>%   
ggplot(aes(x = factor(country),   
 y = factor(year),   
 fill = screening\_rate)) +  
 geom\_tile() +  
 scale\_fill\_gradient(low = "white",   
 high = "darkred") +  
 labs(title = "Cervical Cancer Screening Rates - OECD Countries (2010-2022)",  
 x = "country",   
 y = "year",  
 fill = "Screening Rate (%)",   
 caption = "Data Source:OECD website\nAuthor:David Mwale\nData accessed on: October 10, 2023") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45,   
 hjust = 1))  
  
p6

ggsave(p6, file = "Plot6.png",   
 width = 7,   
 height = 5)

knitr::include\_graphics("C:/Users/DavidMwale/Videos/SCHOOL\_UNIVERSITY OF EDINBURGH/R\_practice/Plot6.png")



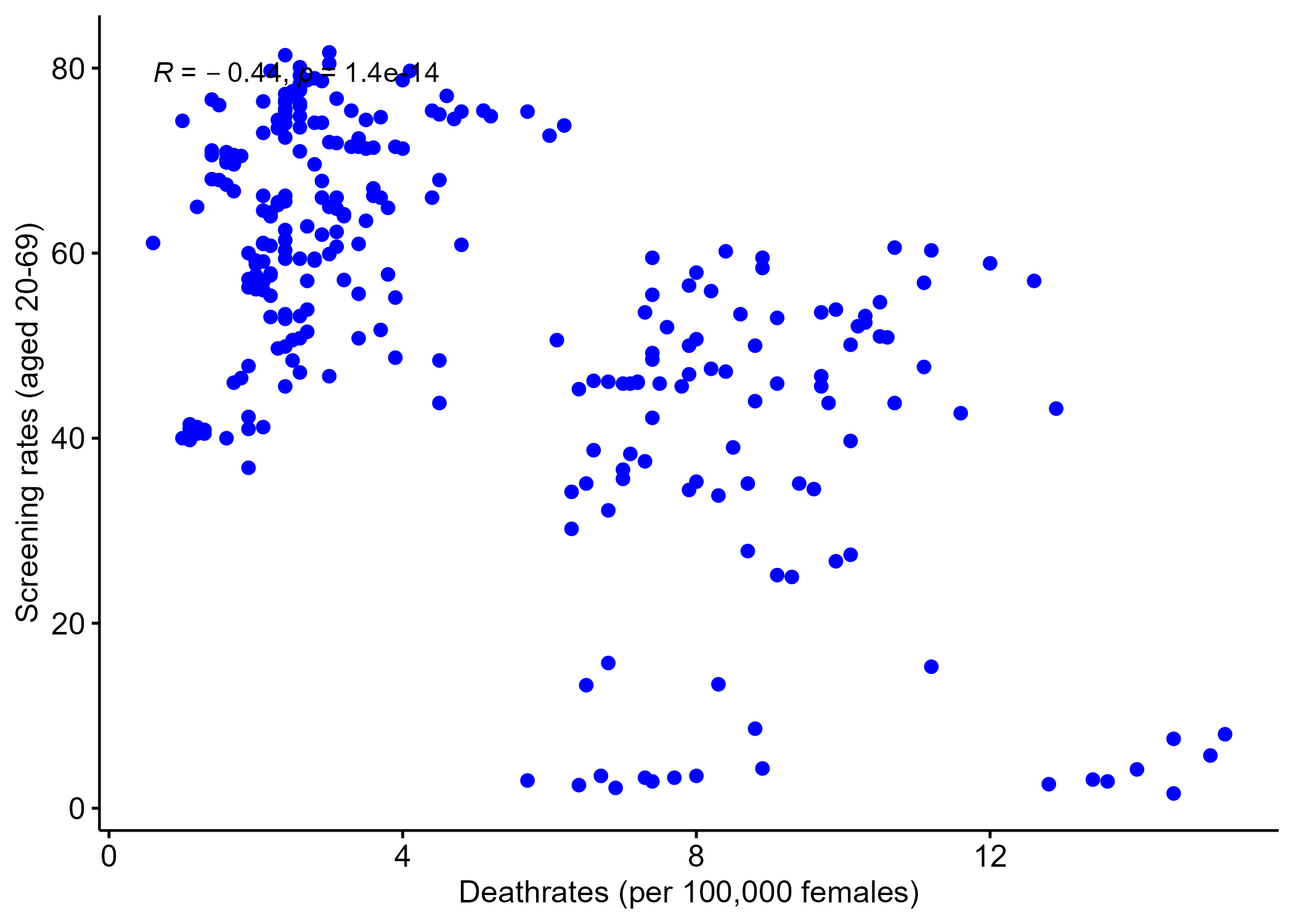
From the chart above, it can be clearly noted that the screening rates are varied, and disproportionate among different countries. Countries like Australia, Belgium, Denmark, Iceland, New Zealand, Finland, Czech Republic, Norway, Slovenia, Sweden, and United Kingdom have higher screening rates of approximately 80%. On the other hand, countries Latvia, Romania, Poland, Hungary, and Bulgaria have lower screening rates. The uneven screening rates are related to inequalities to access to screening services, social and economic determinants such as sex, gender biases, and poverty (WHO, 2023). The chart only shows disparities among oecd countries and shows evidence of the existence of inequalities in cervical cancer screening rates because of different underlying factors. However, if we marry this data to the mortality rate because of cervical cancer, additional insights will unfold. The scatter plot below shows the relationship between cervical cancer screening and mortality rate.

p9<-cc\_joined %>%   
 filter(.,   
 country %in% c("Australia",   
 "Latvia",   
 "Italy",   
 "Denmark",  
 "Belgium",  
 "United Kingdom",  
 "Finland",  
 "Czech Republic",  
 "Costa Rica",  
 "Romania","Iceland","Bulgaria","Canada","Chile","Estonia","France","Hungary","Ireland","Mexico","Poland","Slovenia","Sweden")) %>%  
 filter(complete.cases(.)) %>%  
 group\_by(country) %>%  
 summarize(Deaths\_per\_100000\_females = mean(Deaths\_per\_100000\_females),  
 screening\_rate = mean(screening\_rate)) %>%  
 ggplot(aes(x = Deaths\_per\_100000\_females,   
 y = screening\_rate)) +  
 geom\_point(aes(color = country), size = 2) +  
 geom\_text(aes(label = country, hjust = -0.1, vjust = 1.5)) +  
 geom\_smooth(method = "loess", se = FALSE, color = "cadetblue", size = 0.6) +  
 labs(x = "Deaths per 100,000 females (standardised rates)", y = "Screening rates (aged 20-69)",  
 title = "Cervical Cancer Mortality Rates vs. Screening Rates",  
 caption = "Data source: oecd website\n Author: David Mwale") +  
 expand\_limits(x = c(0, 20), y = c(-1, 80)) +  
 coord\_cartesian(expand = FALSE) +   
 theme(legend.position = "none")  
  
p12<-ggscatter(cc\_joined, x = "Deaths\_per\_100000\_females", y = "screening\_rate",  
 color = "blue", cor.coef = TRUE,   
 cor.method = "spearman",  
 xlab = "Deathrates (per 100,000 females)", ylab = "Screening rates (aged 20-69)")  
  
p12

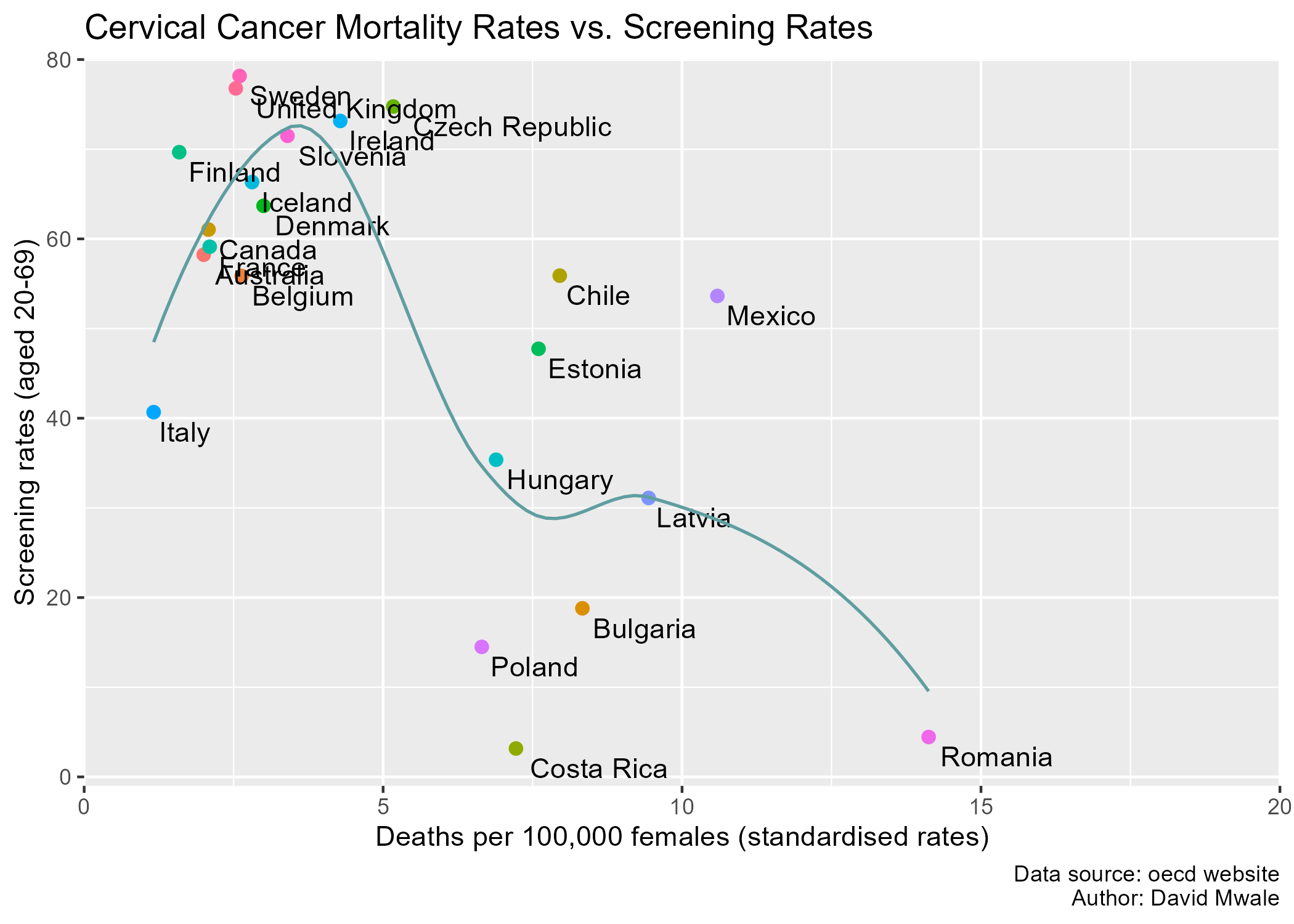
ggsave(p12, file = "plot12.png", width = 7, height = 5)  
p9

ggsave(p9, file = "Plot9.png", width = 7, height = 5)

par(mfrow=c(1, 2))  
knitr::include\_graphics("C:/Users/DavidMwale/Videos/SCHOOL\_UNIVERSITY OF EDINBURGH/R\_practice/Plot12.png")



knitr::include\_graphics("C:/Users/DavidMwale/Videos/SCHOOL\_UNIVERSITY OF EDINBURGH/R\_practice/Plot9.png")



par(mfrow=c(1, 1))

There is an inverse relationship between cervical cancer screening rates, and mortality rates. Countries with low screening rates have higher mortality rates, whereas countries with higher screening rates have relatively lower mortality rates. This relationship means that increased cervical cancer screening rates supposedly result in reduced cervical cancer mortality rates. Countries such as Sweden, The United Kingdom, Finland have evidently high screening rates, and simultaneously low screening rates. On the other hand, the chart clearly shows that low screening rates are persistent in countries like Romania, Costa Rica, Latvia, Poland, and Bulgaria. Even thou Italy has relatively lower cervical cancer screening rates, it seems to have the lowest deaths per 100,000 females, suggesting that there are other underlying factors such as vaccination programs against HPV that contribute a further reduction in mortality rate that is not accounted in the data. In the higher echelon of the cervical cancer screening continuum are countries that have almost one thing in common; the organized screening strategy for cervical cancer, while in the lower epoch of the screening continuum are countries that have lower screening rates as well as higher mortality rates. These latter countries also do not follow the organized screening program.

Although the data has been used to identify the existing relationship between cervical cancer, and cervical cancer mortality rate, it can also be used to answer other questions some of which have been stated below.

\* What optimal point does cervical cancer screening no longer results in reduced mortality rates?  
  
\* What age category is the most affected by cervical cancer mortality?

# **Conclusion and Recommendations**

The aim of the analysis was to examine countries that have lower screening rates, as well as determine the relationship between cervical cancer screening, and mortality rate. Uncovered evidence shows that they are countries such as Latvia, Poland, Costa Rica have extremely lower screening rates, whereas countries like Finland, Sweden, the United Kingdom have relatively higher screening rates. In addition, countries with higher screening rates have lower mortality rates, and countries with lower screening rates have higher mortality rates. There is need to be cautious about these findings in making generalized statements for the entire population as the analysis used program data. In terms of recommendations, it is important that

\* They urge countries with lower screening rates and higher mortality rates should adopt and implement organized screening in order to increase screening rates, and consequently lower cervical cancer mortality rate .  
  
  
\* They encourage countries with higher screening rates, and lower mortality rates to strengthen their invitations system.   
  
  
\* All countries should invest in educational awareness campaigns on the importance of cervical cancer screening.

# **Reflective practice**

My journey throughout the introductory course in Data Science for health and social has been a lot of things. To say the least. It has been exciting, it has been hard. It has been eventful. Overall, it was full of lessons.

At the onset of the program, 2 weeks after the orientation, we started attending classes online. I should admit. At this point in time, I felt collected, excited, and confident about the program. As we descended further into the program content, I began to lose it, and felt like I did not hold it together. Part of the feeling emanated from the fact that I have limited coding background and learning that I would be using R programming language throughout made me more anxious, as R was foundational to the success of the datathon project. I started to participate in the online discussions and realized that the feeling I had was collective as fellow students admitted openly to having zero background in R programming language. In addition, the breakaway sessions during class made me feel uncomfortable, precisely because I did not know how to lead, and still panicked even when I was supporting the one leading.

Even though this part of the experience was bad, it also taught me to seek initiative, reach out to staff, as well as fellow students and learn from them, and I consider this as something good. During the breakaway sessions, the right codes I suggested made me feel not only positive but also confident, and on the flip side the errors that we got because of the wrong codes, gave me negative energy.

Among the many reasons that things went well despite the challenges are the insights expressed in discussion boards, the teaching staff that helped me especially during the break away sessions. In one of the breakaway sessions, I remember Dr Holly Tibble saying “come on you can do this, I believe in you”. Further to this, an article by Bledsoe. T.S., & Baskin, J.J. (2014) on recognizing student fear was particularly useful that I would encourage other students to read.

Apart from learning that fear should never supersede my academic ambitions, I also realized that seeking for assistance in times of confusion is so fundamental. In the future, I plan to still use the strategies outlined by Bledsoe. T.S., & Baskin, J.J. (2014) for helping students to overcome fear.

# References

Landy, R., Pesola, F., Castañón, A., & Sasieni, P. (2016). Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case-control study. British journal of cancer, 115(9), 1140–1146. <https://doi.org/10.1038/bjc.2016.290>

World Health Organization (2020) A cervical cancer-free future: First-ever global commitment to eliminate cancer. Available at: <https://www.nhs.uk/conditions/coronavirus-covid-19/> (Accessed: 30 November 2023).

Alam, Z., Cairns, J. M., Scott, M., Dean, J. A., & Janda, M. (2023). Interventions to increase cervical screening uptake among immigrant women: A systematic review and meta-analysis. PloS one, 18(6), e0281976. <https://doi.org/10.1371/journal.pone.0281976>

Svihrova, V., Kocsis, L., Svihra, J., & Szaboova, V. (2022). Barriers to the Cervical Cancer Screening by CPC-28 Questionnaire: A Pilot Study. Archives of clinical and biomedical research, 6(5), 764–770. <https://doi.org/10.26502/acbr.50170289>

World Health Organization (2023) Cervical cancer. Available at: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer> (Accessed: 3 December 2023).

World Health Organization (2023) Supporting the prevention, detection and treatment of cervical cancer. Available at: <https://www.who.int/europe/activities/supporting-the-prevention-detection-and-treatment-of-cervical-cancer> (Accessed: 3 December 2023).

Tabibi T, Barnes JM, Shah A, Osazuwa-Peters N, Johnson KJ, Brown DS (2022). Human Papillomavirus Vaccination and Trends in Cervical Cancer Incidence and Mortality in the US. JAMA Pediatr. 2022;176(3):313-316. <doi:10.1001/jamapediatrics.2021.4807> T. Scott Bledsoe [tsbledsoe@apu.edu](mailto:tsbledsoe@apu.edu) & Janice J. Baskin (2014) Recognizing Student Fear: The Elephant in the Classroom, College Teaching, 62:1, 32-41, DOI: 10.1080/87567555.2013.831022

Blankinship B(PhD). (2023) “Wrangling”. HEIN110372023-4SV1FLEX: Introduction to data science in health and social care. University of Edinburgh. Unpublished.

Bruni L, Serrano B, Roura E, Alemany L, Cowan M, Herrero R, Poljak M, Murillo R, Broutet N, Riley M.L, Sanjose S(2022). Cervical cancer screening programmes and age-specific coverage estimates for 202 countries and territories worldwide: a review and synthetic analysis. The Lancet Global HealthVol. 10No. 8e1115–e1127. <https://doi.org/10.1016/S2214-109X(22)00241-8>

Devaux M. (2015). Income-related inequalities and inequities in health care services utilisation in 18 selected OECD countries. The European journal of health economics : HEPAC : health economics in prevention and care, 16(1), 21–33. <https://doi.org/10.1007/s10198-013-0546-4>

Weisband, Y. L., Torres, L., Paltiel, O., Sagy, Y. W., Calderon-Margalit, R., & Manor, O. (2021). Socioeconomic Disparity Trends in Cancer Screening Among Women After Introduction of National Quality Indicators. Annals of family medicine, 19(5), 396–404. <https://doi.org/10.1370/afm.2715>

Heer, J., Hellerstein, J.M., Kandel, S. (2018). Data Wrangling. In: Sakr, S., Zomaya, A. (eds) Encyclopedia of Big Data Technologies. Springer, Cham. <https://doi.org/10.1007/978-3-319-63962-8_9-1>