

Data Science Exam

Precious Nhamo

Abstract

This document comprises the responses to **Questions 1 to 5** of the 2025 Data Science examination, along with separate analyses (e.g., a PowerPoint presentation) completed in accordance with the exam instructions and organised within the designated folder.

Data Science Exam

Question1

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Data analysis

We used R (Version 4.4.3; R Core Team, 2025) and the R-packages *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *fastDummies* (Version 1.7.5; Kaplan, 2025), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *ggrepel* (Version 0.9.6; Slowikowski, 2024), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.5; Barth, 2025) for all our analyses.

Results

##Discussion

#Question2 ## Data analysis We used R (Version 4.4.3; R Core Team, 2025) and the R-packages *dplyr* (Version 1.1.4; Wickham et al., 2023), *fastDummies* (Version 1.7.5; Kaplan, 2025), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *ggrepel* (Version 0.9.6; Slowikowski, 2024), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5;

Wickham, Hester, et al., 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, et al., 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.5; Barth, 2025) for all our analyses.

##Results

##Discussion

#Question3 ## Data analysis We used R (Version 4.4.3; R Core Team, 2025) and the R-packages *dplyr* (Version 1.1.4; Wickham et al., 2023), *fastDummies* (Version 1.7.5; Kaplan, 2025), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *ggrepel* (Version 0.9.6; Slowikowski, 2024), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5; Wickham, Hester, et al., 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, et al., 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.5; Barth, 2025) for all our analyses. ##Results

##Discussion

#Question4 ## Data analysis We used R (Version 4.4.3; R Core Team, 2025) and the R-packages *dplyr* (Version 1.1.4; Wickham et al., 2023), *fastDummies* (Version 1.7.5; Kaplan, 2025), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *ggrepel* (Version 0.9.6; Slowikowski, 2024), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5; Wickham, Hester, et al., 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, et al., 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.5; Barth, 2025)

for all our analyses. ##Results

##Discussion

#Question5 ## Data analysis We used R (Version 4.4.3; R Core Team, 2025) and the R-packages *dplyr* (Version 1.1.4; Wickham et al., 2023), *fastDummies* (Version 1.7.5; Kaplan, 2025), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *ggrepel* (Version 0.9.6; Slowikowski, 2024), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2024), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5; Wickham, Hester, et al., 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, et al., 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.5; Barth, 2025) for all our analyses. ##Results

##Discussion

References

- Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- Barth, M. (2025). *tinylabels: Lightweight variable labels*.
<https://doi.org/10.32614/CRAN.package.tinylabels>
- Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>
- Kaplan, J. (2025). *fastDummies: Fast creation of dummy (binary) columns and rows from categorical variables*. Retrieved from <https://CRAN.R-project.org/package=fastDummies>
- Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from <https://CRAN.R-project.org/package=patchwork>
- R Core Team. (2025). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Slowikowski, K. (2024). *Ggrepel: Automatically position non-overlapping text labels with 'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=ggrepel>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . .

- Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2025). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>
- Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>