New and Affordable ‘Active’ and ‘Advance’ Mingar Wearable Lines Turning Heads. . . Maybe Not All the Ones They Wanted

Who are these new customers? And do these racist complaints hold any merit?

Report prepared for MINGAR by Simplicity

2022-04-11

Contents

Executive summary

Our consultants at Simplicity, a company that makes consulting simple for its clients, will be tackling Mingar's most pressing questions today regarding their high-end fitness wearing tracking wearable devices. More specifically, we will be exploring their newest and most affordable lines of wearables, "Active" and "Advance", to uncover who their current audience is and what they find so attractive about these products.

Turns out, it is most likely the affordability. Customers of the new "Active" and "Advance" lines are between the ages of 60 or older and 30 or younger. One thing these age groups have are thier lower income levels. On average, incomes peak for individuals between 30 to 60. Moreover, household median income of the customer by neighbourhood also agrees with the finding that lower income earners prefer "Active" and "Advance" lines, while high income earners prefer others.

Secondly, there has been a trend of complaints that Mingar wearables are "racist", as they were reported to perform poorly for darker complexions compared to thier lighter counter parts. Our findings agreed with those complaints. On the spectrum of skin tones, we found that the number of quality flag raised during sleep increased as the complexion deepened. This may be due to some data recording quality issue or sensor incompetencies.

This report will utilize our strengths to uncover clever manipulations and wrangling of the data to best answer these questions with the limited information available.

Technical report

Introduction

In this report, Simplicity will tackle Mingar's most pressing questions to date, i.e. who are Mingar's customers of their newer and more cost-efficient wearable lines? And do the currently trending damaging claims of "racist" Mingar wearables hold any weight? Read on to find out.

Our company has gone through the exhaustive work of selecting the 2 best fitting models of the data that best answer each question by considering any predictors that may influence our variable of interest, within the realm of common sense and reason backed by statistical tests, handy plots and colourful illustrations. Due to racial and ethnic privacy issues that prevent Mingar from collecting such data, we have rolled up our sleeves, wrangled the data and gotten creative with our approach to solving our client's questions. And in the spirit of transparency, we will explain any limitations that we ran into in the process.

Research questions

Our report aims to address the 2 questions that Mingar has asked of us. More specifically, Question 1 targets primarily who Mingar's new customers of the affordable lines of wearables, "Active" and "Advance" are. We aim to highlight the specific significant factors that do and do not identify such customers.

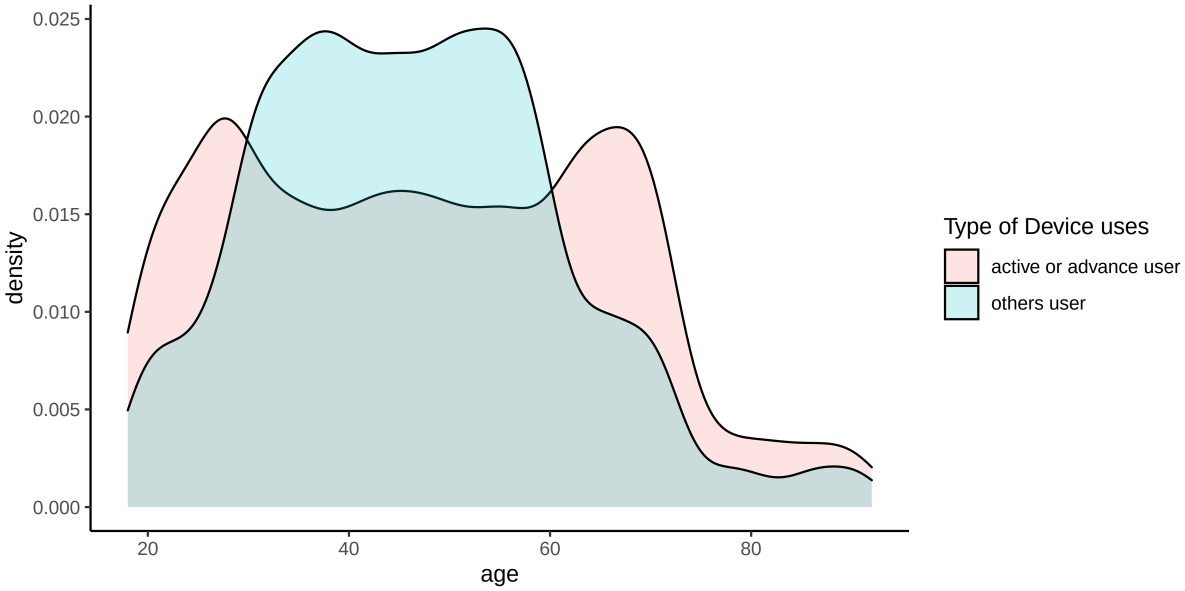
Question 2 asks to find any device incompetencies found during sleep score flag records due to user skin color likely stemming from sensor quality issue or other data quality issue to determine whether there is racial bias existing within Mingar's wearables.

Who are the "Active" and "Advance" Line Wearables Customers?

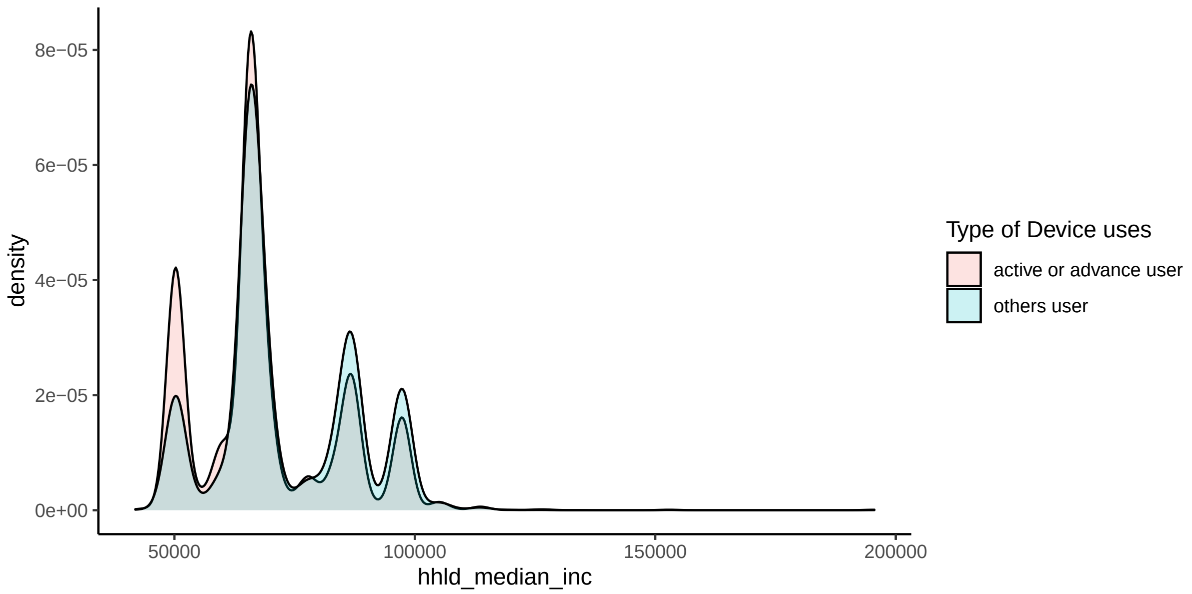
The data used for the purpose of this research question is based on the joint data of customer device, customer, and device data provided from the client Mingar. Then we further strengthen the data information by including living areas sorted by postal code, median income of areas, and population of the areas. Based on the existing variables, column age was created using Date of Birth information from the data. Column emoji color was also created based on the variable emoji modifier in the data. Users preferred emoji skin color can be used as an rough indicator of their skin color, if user uses no emoji modifier, the variable emoji color will default to yellow. However, yellow has zero indications on a user's potential skin color as default emoji face color without any alteration is yellow. Yellow can be treated as unkown potential skin color.

Let us begin by discussing the process surrounding the model selection for this question. We know that we want our response variable to be binomial, where 0 represents a user that does not own an active or advance wearable and 1 represents a user that does own anactive or advance wearable. Thus the underlying distribution of the generalized linear model(GLM) is binomial.

In the first stages of our analysis, we began by looking at the relationship different variables may have had on our response variable.

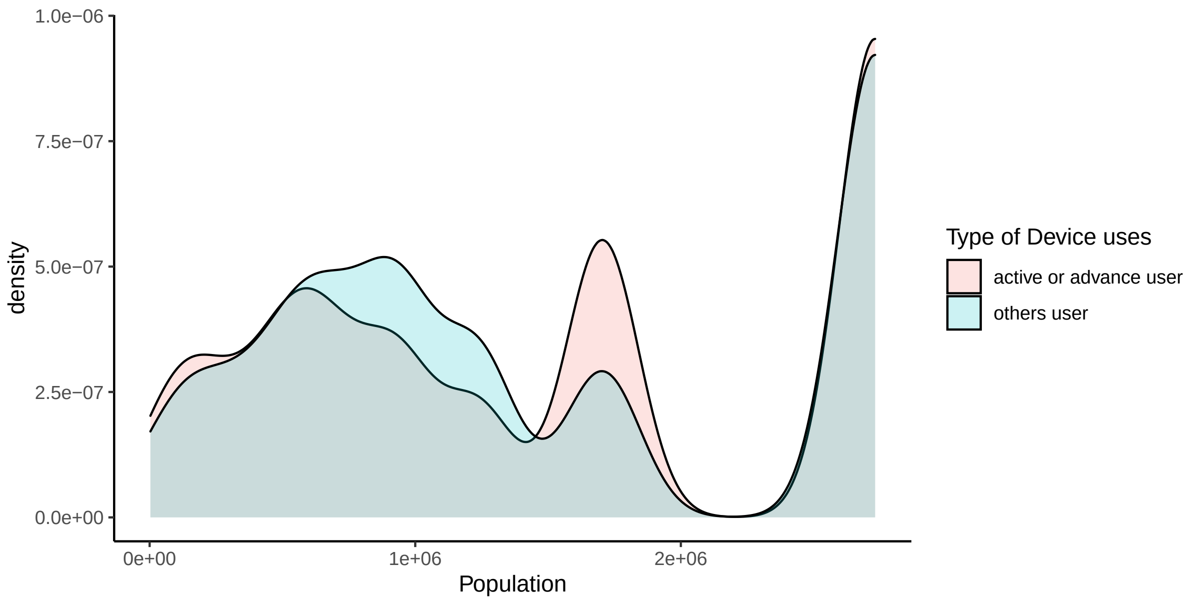


The density distribution figure above illustrates that users less than 30 and older than 60 own more active or advance line wearables than users of the same age group for other lines. This leads us to wonder whether age is a valid model predictor.

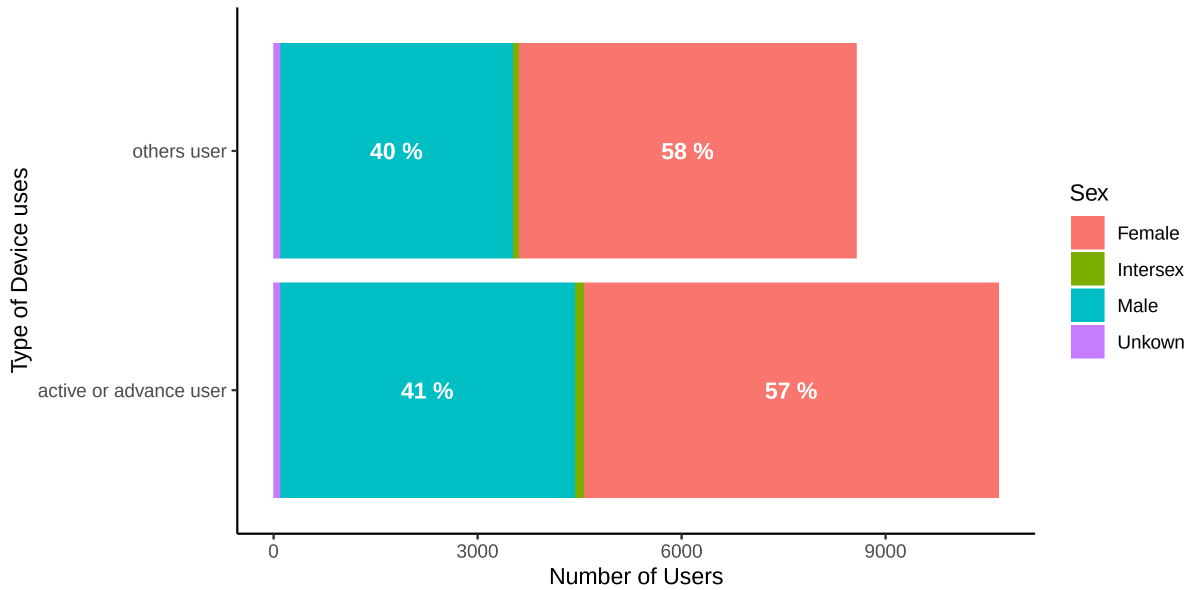


The next density distribution showcases the relationship between household median income for active or advance line users versus other.The figure above shows that customers with a household median income of approximately $70,000 or less own more of the active or advance line wearables than otherwise.

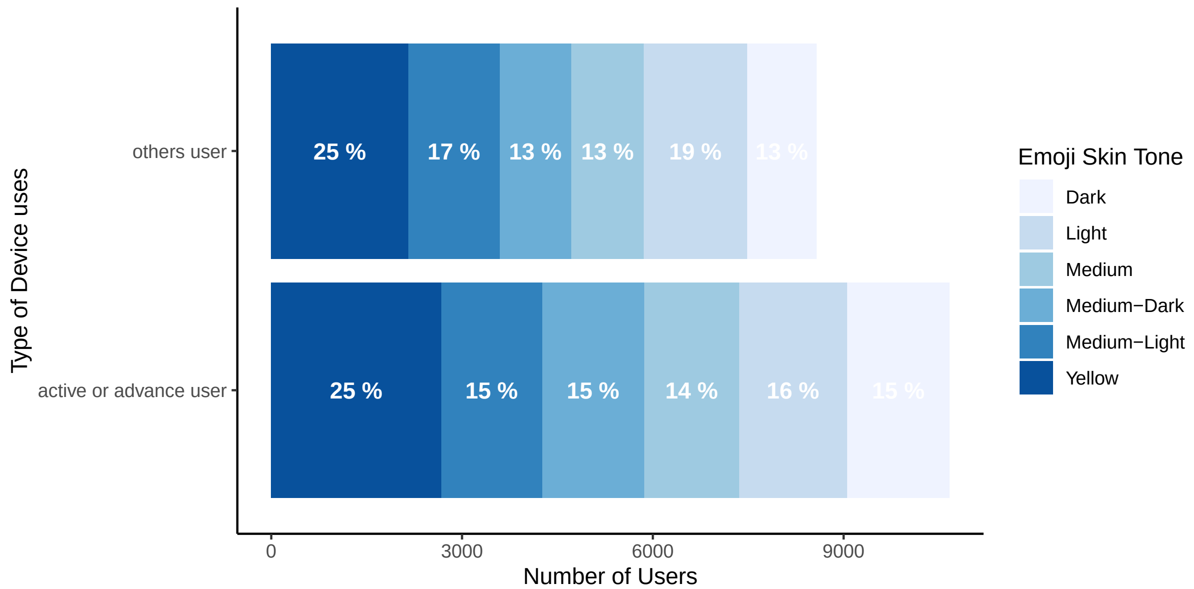
Once again suggesting that a potential product for determining active or advance customers from others is household median income.



Afterwards, we observe our third density distribution figure for this question, this time modeling the population variable, which represents the population of wearables users within a postal code/neighbourhood, against our response variable. However, no discernible pattern or findings can be drawn from its odd shape.



My partner and I were interested in whether the sex of a user differed based on the different lines of wearables. As you can see, the percentage of female, male, intersex and unknown barely altered when coming active or advance users against all others. We hypothesis at this point that sex will not be a predictor for our model.



We display another graph resembling the one above except this one compares emoji skin tone against lines of wearables. Similarly, no difference between active or advance users and other line wearable users in their percentage distribution

Model Selection

To identify the target audiences for Mingar's new budget line up of devices Active and Advance, Simplicity decides to a binomial generalized linear model. The reason to select such model is due to the nature of the variable we are to study, active advance users. It consists of 1 and 0 where 1 represents user uses active or advance devices, and 0 otherwise. We want to determine the demographic backgrounds of active and advance users by using active advance users as outcome and various demographic background as the predictors. And a generalized linear model is a perfect fit for such task. Through multiple nested likelihood ratio tests, testing on background variables like median income, age, sex, location, and potential skin colors, The most appropriate model is given by:

$ActiveAdvanceUsers \sim Binomial(N, p)$

$log(\frac{p}{1-p}) = \beta_0 + \beta_1 MedianIncome + \beta_2 Age + \beta_3 Population$

Where N represents the number of Device Users, and p represents the probability of a device user uses Active or Advance given the information on median income and population.

Table 1: Summary Table of The Generalized Linear Model

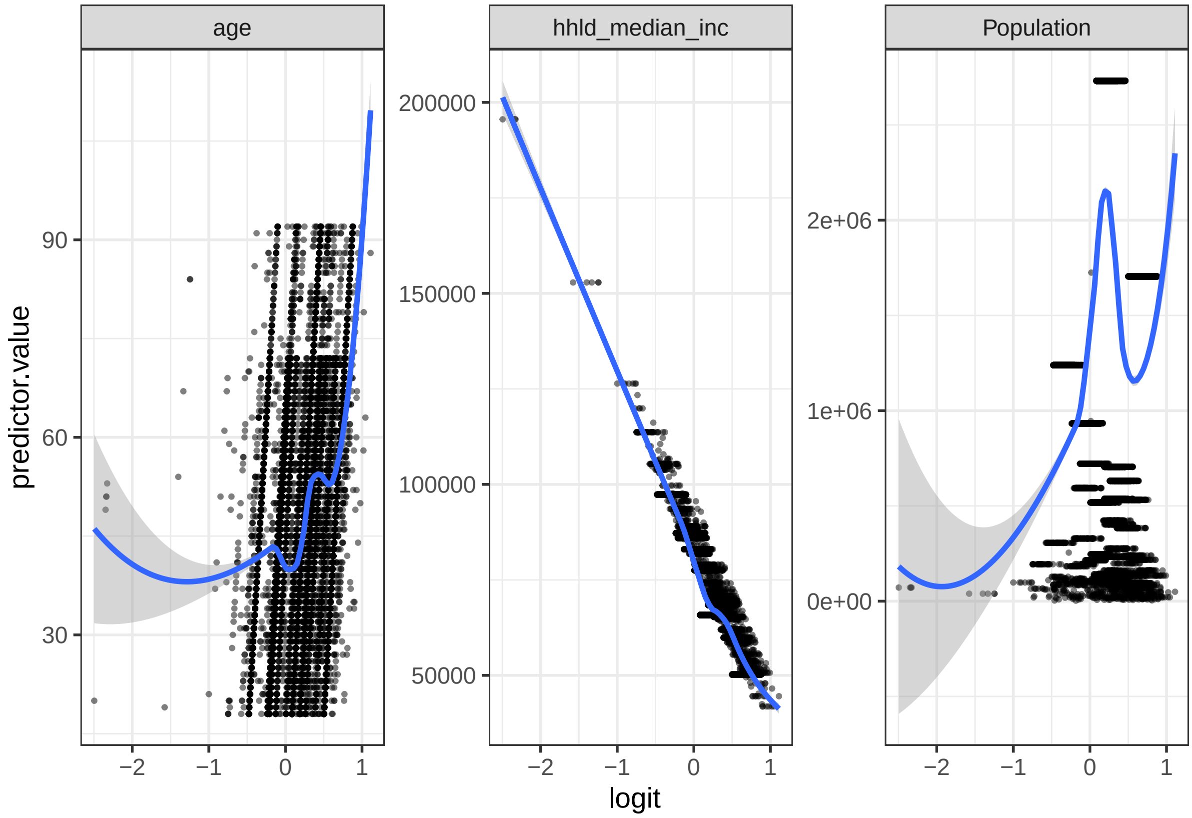
term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.817	0.063	12.934	0	0.693	0.941
scales::rescale(hhld_median_inc)	-3.318	0.192	-17.261	0	-3.695	-2.942
scales::rescale(age)	0.377	0.075	5.041	0	0.230	0.523
Population	0.000	0.000	-4.185	0	0.000	0.000

After conducting likelihood ratio tests, we found that the best suitable model for our binary response variable selected household median income, age and population as significant predictors, not surprisingly. Furthermore, we see that there is a negative relationship between income and our active and advance line users, which leads me to interpret that lower income earners prefer the active and advance wearable lines, while higher income earners prefer other wearables. Further to our findings from above regarding age, we do see that it does play a significant role in characterizing our active and advance line customers. The significance of the population variable leaves lots of rooms for questions as it is unclear as to how or why the number of people in a neighbourhood would play a role in determining a consumer's choice of active or advance line of wearables versus others. One theory I will pose is that perhaps the more individuals there are in a neighbourhood, the more likely those occupants will follow trends and purchase more afforable wearables.

Model Assumptions

For a generalized linear model to be valid, following assumptions needs to hold.

1. The model's response, the probability of Active Advance Users among the users should be independently distributed, and the errors are also assumed to be independent
2. The model assumes a linear relationship between the transformed response with the predictor variables. Which in this case is the $log(\frac{p}{1-p})$ and $age, median\ income$, and $population$. Where p represents the probability of a device user uses Active or Advance given the information on median income and population.



As seen from the figure, both age and median income have clear linear relationship with the transformed response, while population fails to reach the assumptions.

Although Population is significant predictor to the model, due to its low treatment effect on the probability of active or advance users, and its failure to reach the model assumptions. The predictor is removed from the model.

The new model is given by:

$ActiveAdvanceUsers \sim Binomial(N, p)$

$log(\frac{p}{1-p}) = \beta_0 + \beta_1 MedianIncome + \beta_2 Age$

Where N represents the number of Device Users, and p represents the probability of a device user uses Active or Advance given the information on median income and population.

Table 2: Summary Table of The Generalized Linear Model

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.647	0.048	13.450	0	0.553	0.741
scales::rescale(hhld_median_inc)	-3.043	0.180	-16.907	0	-3.396	-2.691
scales::rescale(age)	0.376	0.075	5.042	0	0.230	0.523

The model summary and confidence interval are almost identical to the previously proposed model, just without population as one of the model's predictor.

Are Mingar Wearables Racist?

The data used to determine the cause of poor sleep score performance of the devices is based on the previous data used for the first research question. Then we joint the data with customer sleep data provided by Mingar. The data allows us to explore and validate whether having darker skin color actually causes the devices to perform poorly on sleep functionalities.

The showcase of poor sleep performance of the wearable can be indicated by two variables, user's sleep duration, and the number of malfunction red flags that device records over user's sleep session. Thus, one of the variable will become the output of this study's model.

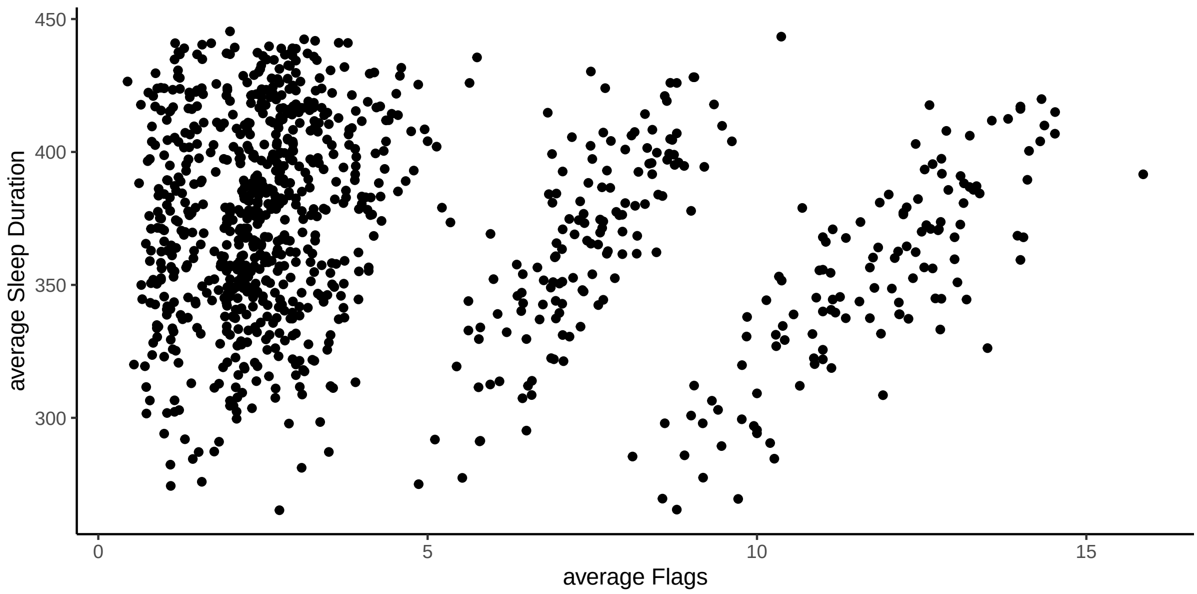


Figure 1: Relationship of average Sleep Duration and average Flags per User's Sleep

Figure 1 illustrates a positive linear relationships between average sleep duration and average red flags per user's sleep. This can be easily interpreted as when sleep duration last longer, wearable devices have more opportunity to malfunction and flags.

Now we wish to compare both sleep duration and flags to demographic backgrounds in order to determine further information for the construction of te model.

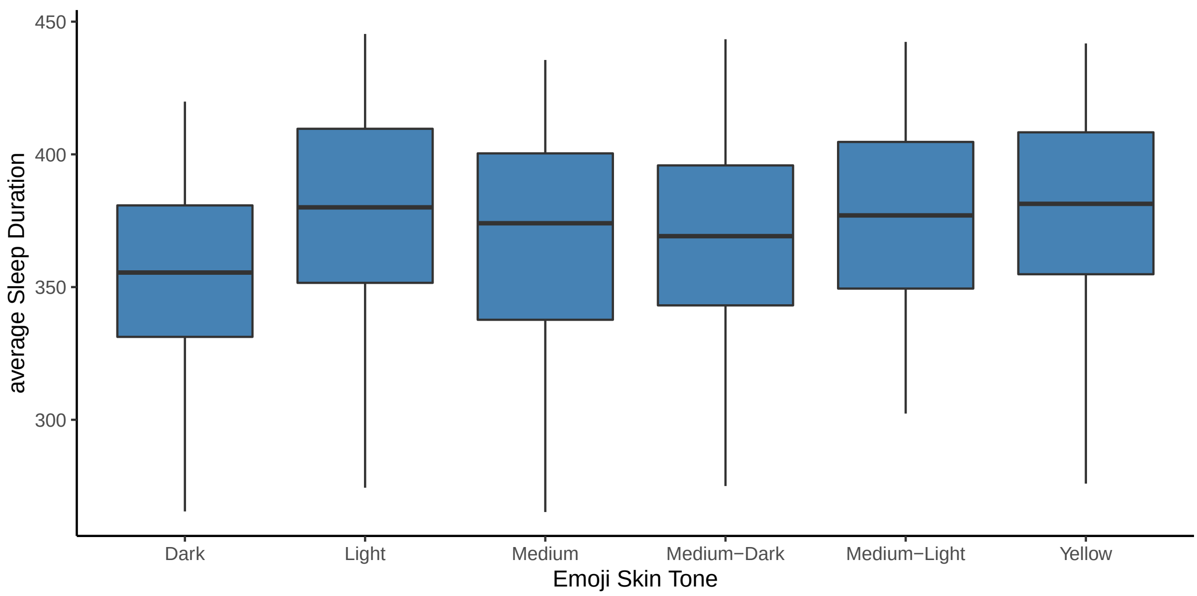


Figure 2 showcases a negative relation between sleep duration and darker skin color. As user's potential skin color gets darker, the sleep duration is expected to drop as well, resulting lower sleep qualities that's possibly due to malfunctions of the wearable.

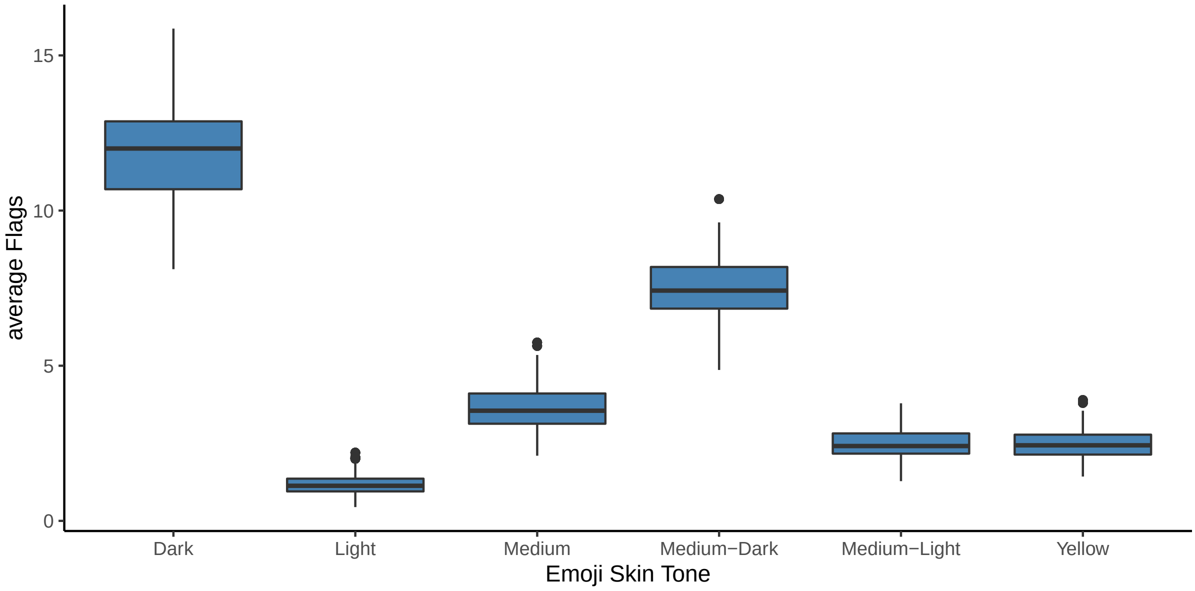


Figure 2: Distribution of average Flags per User's sleep by preferred Emoji Skin Tone

Figure 3 displayed an extremely positive trend of average flags with darker skin color. Reading from this chart, potential skin color is almost certainly a very significant predictor for the determining the average flags per user's sleep, as the number of flags gets higher for every darker change of skin tone.

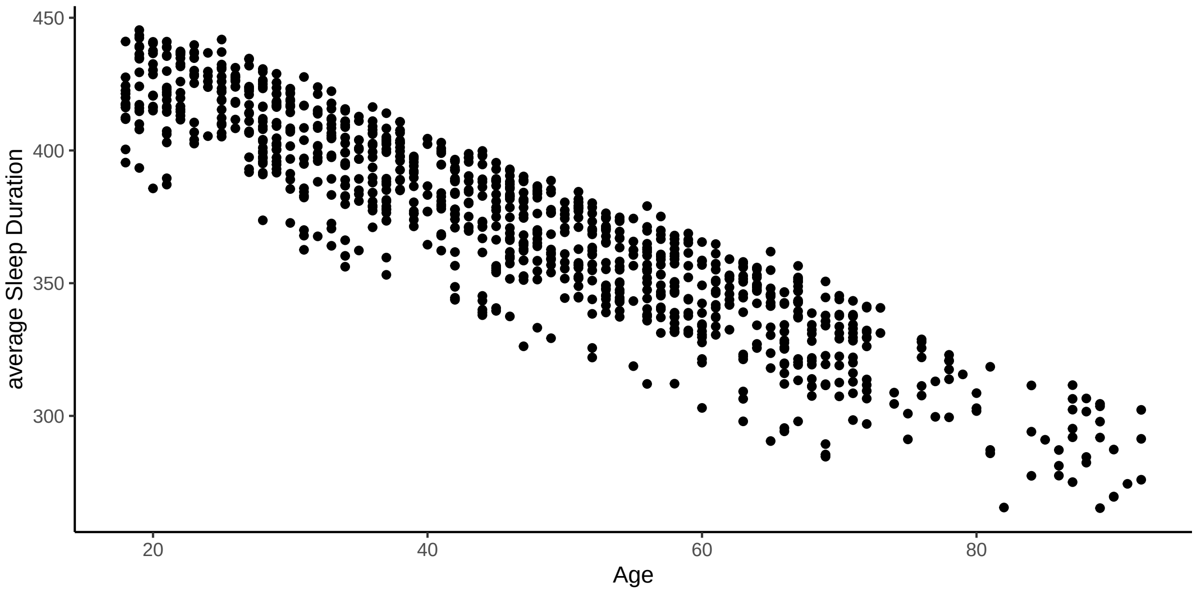


Figure 4 showcased a negative linear relationship between sleep durations and age. Although it is possible that age is a influential variable on Mingar wearable's underperformance on sleep score. It is also very reasonble to interpret the trend as user tends to have lower sleep quality and less sleep time when they get older, due to health and many other reasons.

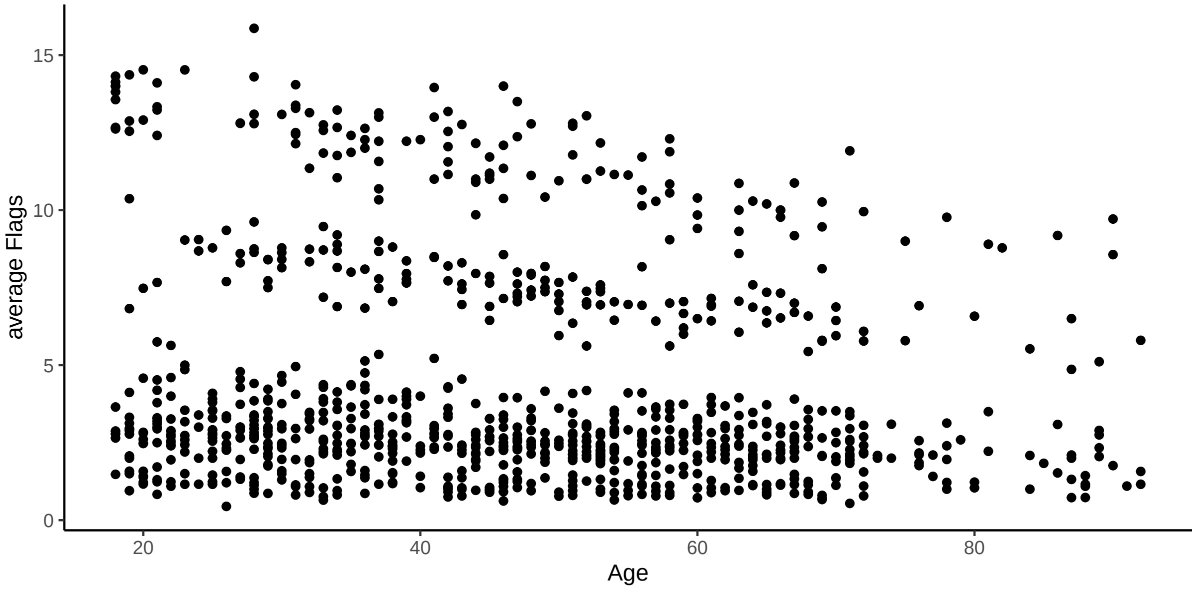
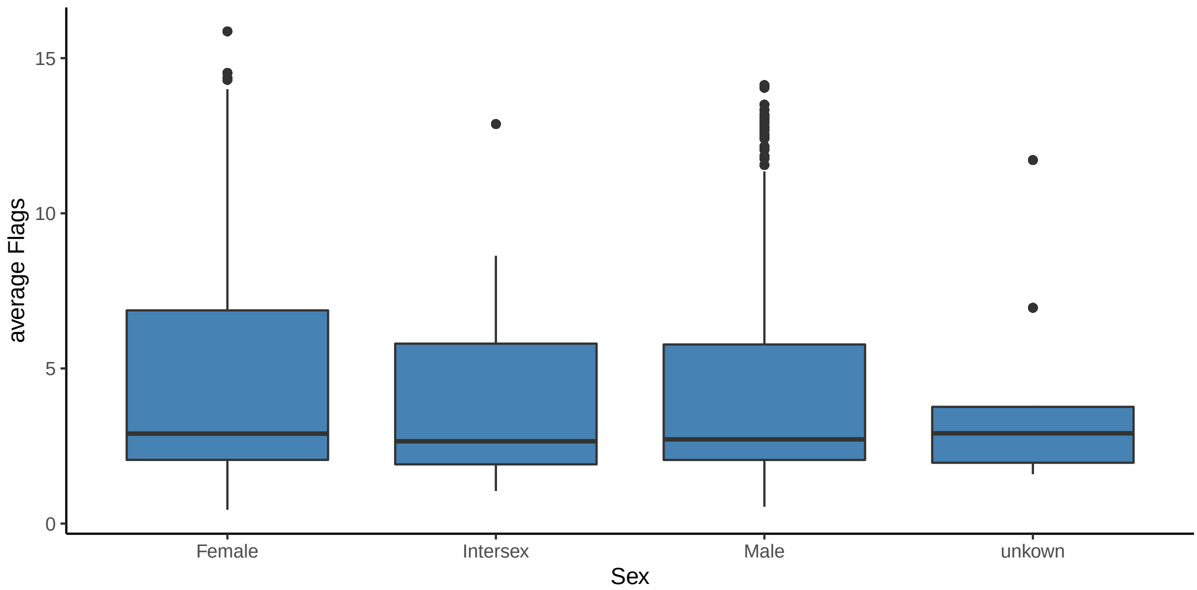


Figure 3: Relationship of average Flags and Age per User's sleep

Unlike figure 4, figure 5 actually almost shows no linear trends btween average number of flags and age. The distribution of the points are rather flat, indicating age actually have no influence on the number of malfunctions and poor performances of the devices. Thus, it is unlikely that the negative trend displayed in figure 4 is caused by wearable's underperformance on older users. This indicates flags is a better output for the model than sleep duration, as number of flags indicated the number errors caused by the machine, while sleep duration includes broader interpretations where a clear trend could possibily caused by other reason than device's fault.

Moving on, the report will only explore the relationship flags with the rest of the demographics.



No differences of average number of flags was displayed between genders. All three type of genders shares similar distributions of flags with the same mean. This indicates that sex have no influence on flags.

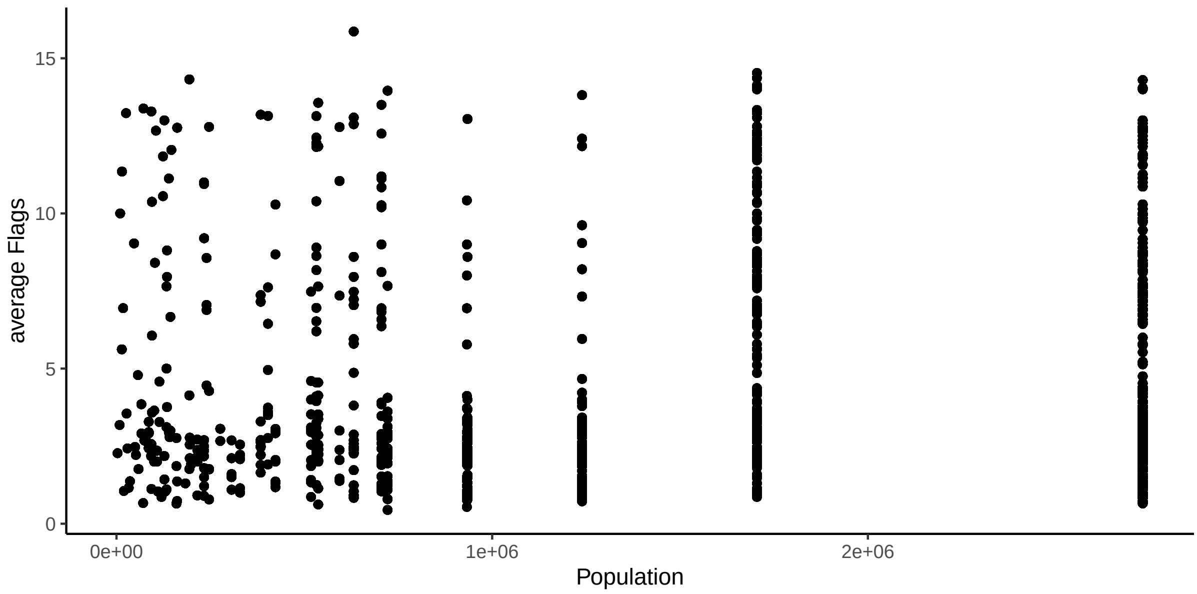


Figure 4: Relationship of average flags per User's Sleep and User's Postal Code Population

Similarly to sex, population also shows no clear trends with the number of average flags.

Model Selection

To identify the root cause for Mingar's some device performing poorly on sleep scores, Simplicity decides to use a Poisson generalized linear mixed model. The first reason to select such a model is due to the nature of the variable we are studying. Flags is a count variable that records the number of red flags of device malfunction over the period of a user's sleep, thus falls under Poisson distribution. Due to the collection method of the data, the data records large amount of observations of sleep from each users, therefore it is appropriate to include customer id as random effect, which makes the model a linear mixed model. From previous data discovery, we learned that flags and sleep duration have a positive linear relation, more flags can potentially appear as sleep lasts longer. To standardized number of flags within the same period of time, log sleep duration is selected to be the offset of the model.

Lastly, we want to determine whether skin color and other demographic backgrounds influences the performance of the devices,

Through multiple nested likelihood ratio tests, testing on background variables median income, age, sex, location, and potential skin colors.

The most appropriate model is given by:

$$log(flags) = log(duration) + \beta_1 emoji\_colorYellow + \beta_2 emoji\_colorLight + \beta_3 emoji\_colorMediumLight + \beta_4 emoji\_colorMedium + \beta_5 emoji\_colorMediumDark + \beta_6 emoji\_colorDark + U_i$$

Where U_i represents random effect for $Customer_i$. Because only dummy variables were used as predictors, the report decides not to use intercept for this model, and there will be no reference level.

Across all demographic backgrounds, Only potential skin colors are significant to the Poisson regression model.This validates the complaint from customers that devices are performing poorly for users with darker skin, particularly with respect to sleep scores.

Table 3: Summary Table of The Poisson Generalized Linear Mixed Model

effect	term	estimate	std.error	statistic	p.value	conf.low	conf.high
fixed	emoji_colorDark	0.033	0	-463.509	0	0.033	0.034
fixed	emoji_colorLight	0.003	0	-365.286	0	0.003	0.003
fixed	emoji_colorMedium	0.010	0	-435.055	0	0.010	0.010
fixed	emoji_colorMedium-Dark	0.020	0	-474.205	0	0.020	0.021
fixed	emoji_colorMedium-Light	0.007	0	-416.901	0	0.006	0.007

effect	term	estimate	std.error	statistic	p.value	conf.low	conf.high
fixed	emoji_colorYellow	0.007	0	-536.011	0	0.006	0.007

Examining the table of summary, all skin colors are extremely significant to the estimation of the model. As expected Dark have high positive influence on the number of flags, with an increase of 0.033 flags per minute in sleep. All skin colors coefficents are captured within the confidence interval as well.

Model Assumptions

For a Poisson generalized linear mixed model to be valid, following assumptions needs to hold.

1. The model's response needs to be under Poisson distribution. Flags is a count variable that records the number of red flags over the period of user's sleep time, thus falls under Poisson distribution.
2. Random effect, which in this case are the customers should be normally distributed.
3. The customer's random effects errors and residual errors within preferred emoji skin should have constant variance.
4. The model assumes a that the transformed response must have a linear function with predictor variables. Which in this case is $log(Flags)$ and $emoji$ $color$.

In conclusion, devices are definitely performing poorly for users with darker skin, particularly with respect to sleep scores. ## Discussion

After selecting the most appropriate model for each question, we can now answer the 2 questions at hand.

Who Are the "Active" and "Advance" Line Wearables Customers? First, we have to answer what the distinct characteristics of the "Active" and "Advance" customers are. Through the final selected generalized linear model, we see the most significant character traits to be income, age and number of residents in the same neighbourhood. Furthermore, we see that there is a negative relationship between the active and advance customers binary response variable and income, which reasons that wealthier individuals buy less of these 2 lines, meanwhile lower income individuals buy more of them. In one of our plots above, we see that customers less than 20 and over 60 buy more of the active and advance lines than the others. This suggests that once again adolescents or young adults and seniors are more the target audience for these lines. This is congruent to our prior conclusions regarding income, i.e. individuals in those age groups are generally regarded as having lower income looking for more affordable products. As for the last characteristic, population within a neighbourhood, conclusions are hard to draw as the relationship is unclear, though notably it is negative.

Are Mingar Wearables "Racist"? In short, yes. Let's talk about it… Through our generalized linear mixed model, it was evident that the number of flags, which are quality issues that occur during sleep, related to sensor error or other data quality issues, increased for darker skin tones of overall the emoji modifiers; in exactly the order of Light, Medium Light, Medium, Medium Dark to Dark (least flags to most flags respectively). Moreover, out of the six models we considered, the most suitable model found the emoji modifier predictor to be the only significant one for the flag response variable, leaving us with the belief that the only consequential factor of these flags is the individual's skin tone. Our recommendation is that further tests on the hardware of the wearables should be conducted to determine the exact cause behind the skin tone sensitivity and lack of quality as a result. ### Strengths and limitations

Simplicity believes in transparency and as such, we find it important to disclose the limitations of our research and findings in order for our clients to feel confident with what they are receiving. Since we do not have information on individual income, we had to improvise. We cleverly pulled income information based on postal code from the postal code conversion file (2016 census geography, August 2021 postal codes) provided to us as University of Toronto students. Thus, we only have a general sense of the wealth of a customer based on the neighbourhood they live in. It may not be exactly accurate per consumer considering that high income individuals may reside in low income neighbourhoods and the opposite is true too; however we believe that our analysis provides insights as to the average buying behaviours of your wearables customers, important when answering you first set of questions.

Another limitation we encountered was discovered when tackling your second set of questions regarding the racial complaints against Mingar's devices. Since, as mentioned, Mingar does not collect any data on the racial ethnicity of its users, we had to get creative with identifying darker complexion customers from lighter ones. We utilized the emoji modifier field populated by the customer. We must warn that some users have no data within the emoji_modifier field (labeled as NA) as they have presumably not selected one, so we had to work within a subset of the customer data. Moreover, it is not exactly accurate to suggest that all customers use the same skin tone emoji modifier as their own, because people have complex skin tones and may find it difficult to select the right shade. Or they may simply slect the emoji that they wish to express themselves with regardless of their own shades. However, we believe, once again, that the average consumer will selected a fairly similar emoji modifier to their own skin colour; and that was how we were able to proceed with our analysis and provide some meaningful results.

Not to toot our own horn, but Simplicity prides itself on being expert model selectors, whereby we consider numerous models, six, in fact, per question. Through an open and diverse lens, we consider various predictors when answering both questions of focus to ensure we do not exclude any important factors. Our strong foundational understanding of distributions assisted us in deciding what the underlying distributions of the datasets that we cleverly were able to wrangle, join together and explore for the most meaningful models. Most importantly, our company's strong core beliefs of accountability and integrity allows you to be assured that the final model selection was based on results that were authentic and reproducible with various illustrations, plots and tests to showcase the data and results.

Consultant information

Consultant profiles

Aaisha Eid. Aaisha is a senior consultant at Google. She specializes in reproducible analysis and statistical communication. Aaisha earned her Bachelor of Science, Specialist in Statistics Machine Learning and Data Mining, from the University of Toronto in 2022.

Charles Lu. Charles is a junior consultant with Facebook. He specializes in data wrangling and model selection. Charles earned their Bachelor of Science, Specializing in Computer Science from the University of Toronto in 2024.

Code of ethical conduct

Simplicity ensures that any information shared between itself and its clients remain confidential and protected against sale for personal gain or benefit. Conflicts of interest will surely be communicated to clients transparently and accordingly. Personal information provided to by clients will be treated anonymously and respectfully. Finally, Simplicity is operated by proud statisticians that uphold the professional statistical standards of procedure and analysis.

References

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL https://CRAN.R-project.org/doc/Rnews/

Hao Zhu (2021). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4. https://CRAN.R-project.org/package=kableExtra

David Robinson, Alex Hayes and Simon Couch (2021). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.6. https://CRAN.R-project.org/package=broom

Ben Bolker and David Robinson (2022). broom.mixed: Tidying Methods for Mixed Models. R package version 0.2.9.3. https://CRAN.R-project.org/package=broom.mixed

Chung-hong Chan, Geoffrey CH Chan, Thomas J. Leeper, and Jason Becker (2021). rio: A Swiss-army knife for data file I/O. R package version 0.5.26.

Hadley Wickham and Evan Miller (2020). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.3.1. https://CRAN.R-project.org/package=haven

Hadley Wickham and Jim Hester (2020). readr: Read Rectangular Text Data. R package version 1.4.0. https://CRAN.R-project.org/package=readr

Hadley Wickham (2021). rvest: Easily Harvest (Scrape) Web Pages. R package version 1.0.2. https://CRAN.R-project.org/package=rvest

von Bergmann, J., Dmitry Shkolnik, and Aaron Jacobs (2021). cancensus: R package to access, retrieve, and work with Canadian Census data and geography. v0.4.2.

Unicode, Inc. (2022). Unicode character table. Retrieved April 11, 2022, from https://unicode-table.com/en/

Appendix

Web scraping industry data on fitness tracker devices

The industry data on fitness tracker devices was web scrapped in R from https://fitnesstrackerinfohub.netlify.app/. the whole web scraping process utilizes the R packages "polite" and "rvest". informative user agents details including purpose of web scraping and contact information was provided to website before scraping. It is important to note that to respect the web scrapped contents, follow the terms and services and Robots.txt of the website and only take what is needed. Avoid web scraping at all if the target website provide public API to access the data. Respect the crawl limit suggested by the website. The web scraping industry data on fitness tracker devices are private data saved under the folder raw-data, and will never be shared and published on Github or any websites.

Accessing Census data on median household income

The Canada Census data on median household income was accessed from https://censusmapper.ca/ through the APIs of R package "cancencus" and the private API key kindly provided by CensusMapper. The data is governed by the Statistic Canada Open Data license and Statistic Canada, and acknowledgment of them is required use the data. The Census data are licensed data saved under the folder raw-data, and will never be shared and published on Github or any websites. The API key provided by CensusMapper should also not be shared anywhere as it is a personal private key.

Accessing postcode conversion files

The postal code conversion files was downloaded from https://mdl.library.utoronto.ca/collections/numeric-data/census-canada/postal-code-conversion-file through using University of Toronto student credentials. The file is governed by Statistic Canada and stored in University of Toronto online map and data library. Same as the Census data, the file is classified as licensed data saved under the folder raw-data, and will never be shared and published on Github or any websites.