

STATS5029P: Statistics Project and Dissertation

**#34: Does playing Pokémon Go increase physical activity?**

Steven Hoi Him Kwok

2640948k

2/12/2021

# Contents

**0. Abstract1**

**1. Introduction1**

**2. Literature Review 2**

**2.1 Background Information2**

**2.2 Research Objectives3**

**3. Methodology 4**

**3.1 Data Description and Processing4**

**3.2 Model Selection9**

**4. Analysis Results13**

**5. Discussion 16**

**6. Conclusion and Limitations18**

**7. References19**

**8. Appendix22**

# 0. Abstract

Pokémon Go is a well-known augmented mobile game, that potentially can be useful for increasing physical activity among users. However, some researchers have argued that the effects of Pokémon Go on the amount of exercise are neither direct nor stable. In this research, we examine the relationship between the Pokémon Go application usage and the amount of physical activity of users, as well as the players’ preferences. We also identify factors that affect the level of physical activity and the factors associated with participants' attitudes towards sports. The whole study is processed using a polynomial linear regression model, scatter plots and box plots. Finally, we discuss interpretations and the causes of those findings, as well as provide suggestions for future research and implications for the public.

# 1. Introduction

Pokémon Go is a globally popular augmented reality (AR) mobile game that affects the behaviour of a vast number of players. Following arguments presented by several researchers (like Gunther (2016)), it is hypothesized that Pokémon Go could increase levels of physical activity. To test this, I use a polynomial linear model to study the effects of Pokémon Go on physical activity. Several factors are under consideration, including the amount of app usage and; the age, social sharing level, education level, and attitude towards sports of 981 participants in America. Base on this research, the above hypothesis is not supported: players prefer participating in an app-related activity instead of general physical activity. However, the positive effects of app related activities are reduced with associated with the social sharing level. Due to the complex relationships between education level gender, age and attitude scores, attitude towards physical activity is a key factor contributing to the positive effect of Pokémon Go on levels of physical activity. The results of the analysis and implications for the public and future research are discussed in the following sections of the report.

The study is completed in the following steps: First, background information on sports, public health, AR technology and Pokémon Go are presented. Then, we study the effects of Pokémon Go on levels of physical activity, considering several aspects: Frequency of app usage, the participation preferences of players; other factors relate to the amount of physical activity and the effects of the attitudes towards physical activity (with the discovery of factors related to participants' attitudes). Before conducting the data analysis, we group highly-correlated variables by the mean of each record. The entire analysis is conducted using a polynomial linear regression model. The stepwise selection method (setting [AIC] as the variable selection criteria) is applied to obtain the best model, as well as drop insignificant variables. After presenting our observations, we discuss the causes and meaning of the results, provide suggestions to the public, and identify the limitations of the study and avenues for the future.

# 2. Literature Review

## 2.1 Background Information

Sports have been proven to improve the immunity system while training the muscles and respiratory system (Seippel, 2006). However, people living in urbanized societies are prone to living sedentary lives. By Buraimo et al. (2011), approximately half of those surveyed do not participate in any kind of sports for various reasons, including job characteristics, changes in entertainment preferences, and long working hours. With lower levels of physical activity, the risk of obesity, depression, high blood pressure, as well as a series of cardiovascular diseases rise rapidly, which is harmful to public health. Quinn (2021) find that a positive attitude encourages people to engage in sports more frequently. Augmented reality video games, which are perceived as a more interesting way of exercising, have been considered to solve health issues. Augmented reality is an extended version of virtual reality (VR). It uses visual devices, including eyeglasses, monitors, and smart devices, to combine virtual elements with the real world (Fisher. 2021). Therefore, people can see virtual objects appearing in the real world. Besides visualization, an AR system can also include sounds and tactile elements, providing a new version of the world. AR technology can be applied in various types of applications, such as maps and games, and mobile games are particularly common and attractive to users. In this study, we focus on Pokémon Go, a popular AR mobile game.

Pokémon Go, developed by Niantic Inc., is a well-known mobile game that was released in 2016 on both Apple Store and Google Play (Reilly, 2017). It combines an application with GPS signals and a real-world map. Players of Pokémon Go can locate, catch (also obtain ingredients for training Pokémon), hatch (players walk around 2 to 10 km to get a Pokémon), and train virtual creatures, Pokémon. Those Pokémon can be used for battling and gym controlling (Webster, 2015). Pokémon Go uses a map and camera to display the virtual spots, such as PokeStops, gyms, and activity location points. (Smith, 2017). Pokémon Go is a celebrated application, with 632 million downloads and 147 million monthly active players. Due to its popularity and attractiveness to users, it is plausible that Pokémon Go facilitates behavioural changes that might affect public health (Dillet, 2016). If a correlation is found between Pokémon Go and users’ levels of physical activity, this represent a new way of increasing engagement in sports, and fostering positive attitudes towards physical activity.

Multiple pieces of research have found that Pokémon Go affects the levels of physical activity of most participants (Gunther, 2016; Wong, 2017). One study discovers that users’ level of physical activities increased by approximately 25%, and physical activity levels increased across genders, ages, and weight statuses (Gunther, 2016). In general, the levels of physical activity of players who are inactive originally increase sharply (Gunther, 2016). Some studies, however, suggest that Pokémon Go does not directly improve physical health. One study finds that despite positive effects in the first period, players' physical activity levels drop sharply, indicating that the positive effects are not sustainable (Allana et al., 2016). This can be related to players’ motivation. Another study identifies three kinds of motivation among players: health, social, and immersion (Lukas et al., 2017). Although players motivated by health show a significant increase in physical activity levels, the effects of social motivation and immersion are limited. However, time spent outdoors increase significantly (Lukas et al., 2017). Base on the above arguments, it is hypothesized that Pokémon Go positively, but indirectly, affects physical activity levels. To test this assumption, we conduct a study on the relationships between Pokémon Go and physical activity. If Pokémon Go does increase levels of physical activity, more AR mobile games can be developed to improve public health. The objectives are presented in the "Research Objectives" section.

## 2.2 Research Objectives

Pokémon Go is designed as a game originally. Therefore, improving public health is not the initial objective. That is the reason for the confusion between the relation between Pokémon Go and the amount of physical activity. For examining the above confusion, the study is conducted based on four aspects: We first focus on the relationship between frequency of app usage and amount of physical activity. Theoretically, the more the app usage, the higher the opportunity for players to be physically active as most of the app-related activity requires walking (e.g. catching Pokémon, or turning Pokestop). Following the application usage aspect, we discuss the problem in players' characteristics. Since some researchers, like Alessandro (2017) argue that Pokémon Go players tend to join the game-related physical activity, instead of physical activity in general. The aim is to discover the existence of relations between Pokémon Go players and the amount of general physical activity; The level of physical activity can be affected by various factors, including motivation, education level, and gender. Therefore, this study locates variables associated with the number of physical activities. Last but not least, the effects of the attitude towards physical activity, caused by gender or educational level, are examined to explain the effects on the levels of physical activity based on participants' attitudes towards physical activity.

# 3. Methodology

# 3.1 Data Description and Processing

The data is obtained from a study in America, obeying the code of ethics of the world medical association (Declaration of Helsinki) for studies using humans as data. Amazon Mechanical Turk (MTurk), an internet-based platform offering an online participant pool, is selected for data collection (Buhrmester et al., 2011; Paolacci & Chandler, 2014). The original data contains 999 records, described by 31 variables. Before processing data, the number of missing values is examined (which is 0), as well as filtering out records by an attention filtering variable.

After primary data processing, a variable is applied for filtering out non-focus participants. Despite the convenience as well as limitless of time and location, the Mechanical Turk experiment cannot guarantee that participants are paying attention as the survey was completed online (Jennifer Jacquet, 2011). Ensuring only data from focusing-on-survey participants were collected, a variable that acted as an attention filter is used. If participants fail to choose "Disagree" in this question, the records will be removed due to being classified as non-focus records. The number of remaining records were 981. After primary data cleaning, we transform all columns into integer scores, according to the level of each variable (mentioned in the following paragraph). This subjective assigning method is plausible for applying interval scale and the concept of distance (Chaowei Yang, 2014). The identification number is only used for representing the number; the surveying date contains no information as just recording date of survey submission; the IP address of a computer do not affect participants’ behaviour; the last variable, representing participants’ behaviour, is marked as “cannot be used”. Thus, twenty-eight variables are applied for further grouping and analysis.

Except for age (discrete variable) and gender (nominal variable), all the others are ordinal variables. Both "Frequency of App Usage" (accessing the extent of players using Pokémon Go per month) and "How often sharing on social media" (demonstrates the frequency players share their achievements on social media) are ordinal data, fixed at a scale from 1 = "never" to 7 = "very often". Those variables mentioned above are treated as independent variables, while the remaining variables are grouped by row mean, under the result of Cronbach's alpha. Cronbach's alpha, also known as alpha reliability, is a measure for assessing the strength of internal consistency, of several items or variables (Glen, 2021; Cronbach Lee, 1951). The alpha score was calculated by correlating the score for every item with the total score for related observations, following the comparison of the variance of individual item scores (Cronbach Lee, 1951). The formula is shown in **formula 3.1**:

Where is the number of scale items, is the variance associated with item , and is denoted as the variance associated with the observed total scores (Chelsea Goforth, 2015). Following the ***Rule of Thumb***, if the alpha score is between 0.7 and 0.8, the grouping process is plausible (Stephanie Glen, 2021). With the alpha score mentioned below (format: ), we group variables having strong internal consistency by mean of each instance. Cortina (1992) defined that grouping is acceptable if the alpha score is larger than 0.7, recommended in many pieces of research as it is uninterrupted. In supplement, the group-records-by-mean method is applied as we prefer grouping variables (Underhill L.G, 1998), without missing much information or altering the scale. However, only one score exist if we group records my median (Ganti, 2021). There can be a big difference between the median and the recorded scores larger than the median. Using this, we cannot reflect participants' real attitudes towards physical activity. Moreover, every new variable is grouped from a various number of original variables, we do not want to observe scaling differences between all-new variables (McNeish & Wolf, 2020). The detailed variables grouping process, as well as the alpha score, are mentioned below.

Formula 3.1: Alpha Score

There are 12 variables, grouped as one variable denoted as the attitude of participants towards general physical activity (). (Scale of all questions were from 1 = "completely disagree" to 7 = "completely agree"). Players' physical behaviour is assessed in two aspects, recency, and frequency (). The first three items for measuring recency of participants’ physical activity are “When was the last time you had (1) a walk for more than 30 min/ (2) had a run/ (3) had a bike ride to get some exercise?”. (The scale for those questions is 1= “more than one month ago”, 2= “about four weeks ago”, 3= “about three weeks ago”, 4= “about two weeks ago”, 5= “about one week ago”, 6= “during the last week” and 7= “yesterday”.) For measuring frequency, the following three questions are adopted: “How many times have you had (1) a walk for more than 30 min/ (2) had a run/ (3) had a bike ride to get some exercise during the last month?” with the scale from 1=” never" to 7=" every day". Both the former and the latter are transformed as one variable representing Pokémon Go related behaviour. The remaining three variables, representing participants behaviour related to Pokémon Go, are used for assessing participants’ physical behaviour relating to Pokémon Go (). Questions represented by these three variables are “How many times have you walked more than 30 min/ had a run/ had a bike ride with the intent of searching for Pokémon Go during the last month?” (anchored with 1= “never”, 2= “two times”,3= “from three to five times”,4= “from six to eight times”, 5= “from nine to eleven times”, 6= “from twelve to fourteen times” and 7= “every day”). Those variables are grouped as one variable representing participants’ general physical activity. After variables grouping, the new data set contains eight variables and 981 records. Before model selection, we look at the summary (**table 3.2**) for detailed information.

According to **Table 3.2**, there were 981 records and 8 variables. By the pattern demonstrated in **Figure 3.2**, we observed that Pokémon Go related behaviour was unusually correlated with age, attitude towards physical activity, and the amount of physical activity. There are curve-linear relations between variables possibly, with themselves or others. **Figure 3.3**, additionally, showed that relations exist between gender and three variables, like education, attitude towards sports, and the amount of physical activity. **Figure 3.4** also proved that education level has positive correlations to age and attitude towards physical activity. For modelling the interactions between variables, we apply the polynomial regression model, mentioning more details in the following session.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| (n=981) | Min | 1st Qu. | Median | Mean | 3rd Qu. | Max |
| age | 18.00 | 25.00 | 30.00 | 32.55 | 37.00 | 74.00 |
| education | 1.000 | 5.000 | 8.000 | 6.889 | 8.000 | 11.000 |
| Gender | 1.000 | 1.000 | 1.000 | 1.378 | 2.000 | 2.000 |
| Attitude | 3.000 | 5.167 | 5.417 | 5.362 | 5.667 | 6.500 |
| PhysicalActivity | 1.000 | 2.333 | 3.000 | 3.244 | 4.167 | 7.000 |
| PokemonGo\_AppUsage | 1.00 | 1.00 | 1.00 | 2.45 | 4.00 | 7.00 |
| social\_sharing | 1.000 | 1.000 | 1.000 | 1.611 | 1.000 | 7.000 |
| PokemonG0\_Relate.Behaviour | 1.000 | 1.000 | 1.000 | 1.492 | 1.667 | 7.000 |

Table 3.2: Summary of Grouped Data



Figure 3.3: Scatter plot of data distribution

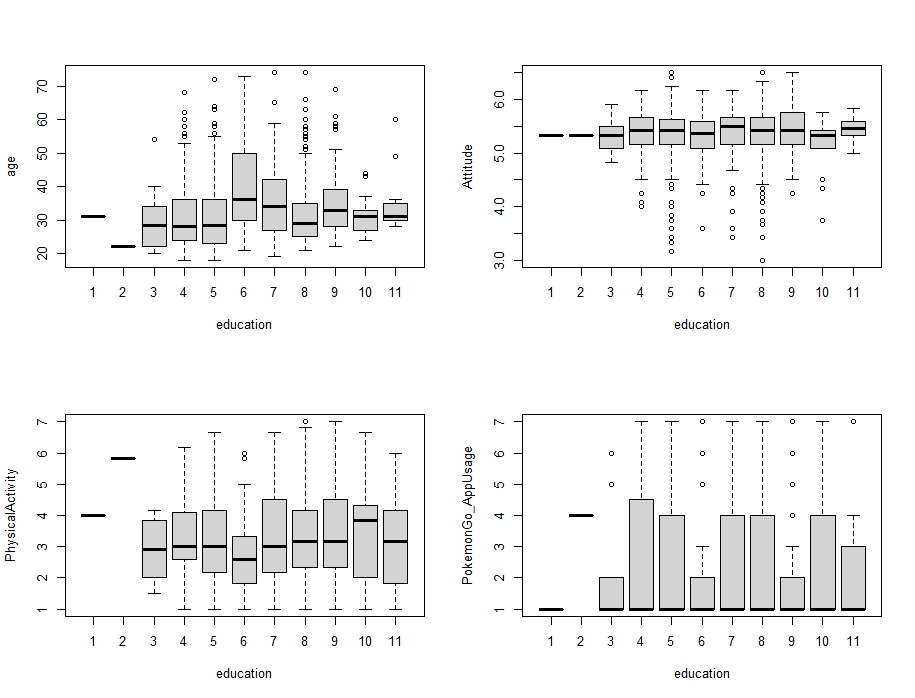


Figure 3.5: Boxplots of Education level vs 4 other variables

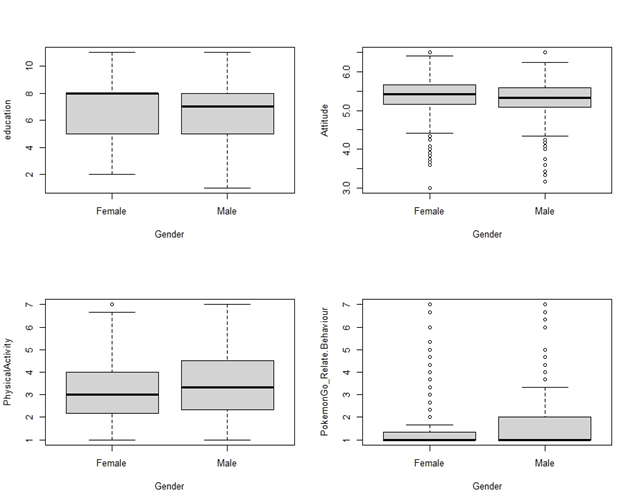


Figure 3.4: Boxplots of Gender vs 4 other variables

# 3.2 Model Selection

Polynomial linear regression model, a linear regression model with a degree of coefficient of more than one, is a model combining interactions between variables (Abhigyan, 2020). Polynomial regression can model the non-linear relationship between dependent and independent variables by adding polynomial terms to linear regression, such as the square of a variable (Agrawal. 2021). A polynomial regression model is selected as the most suitable approximation of relations between variables can be provided. Some variables, in reality, have correlations with others or themselves. Simple linear regression, however, cannot model these interactions, leading to large error and inaccurate relationship estimation. With polynomial terms, we build up a better model for better observations of these latent relations. A vast range of models, also, can be applied in the polynomial regression model, including linear, Gamma, and Poisson. The curvature of a polynomial regression model, moreover, is flexible (Pant. 2019). We can fit a wide range of positive values. Due to accuracy and flexibility, we select the polynomial linear regression model. A linear model is applied as it is easy for interpretation and explanation. Moreover, there is no evidence to apply another model. There are four assumptions for the polynomial linear model. The behaviour of a response variable, to begin with, can be explained by an additive relationship (both linear and curvilinear are plausible) between a response variable and several explanatory variables. It is assumed that the relations between a response variable and explanatory variables are linear or curvilinear. The independent variables, additionally, should be independent of each other. The residuals must be independent and normally distributed, following a mean zero and constant variance (Abhigyan. 2020). We examine the following assumption after obtaining the best model.

After developing a full linear model (showed and explained in **Formula 3.11**), stepwise selection, using AIC as criteria, was applied to select the best model. Stepwise regression is a step-by-step iterative and automatic model selection approach (Adam Hayes, 2021), based on backward regression and combining with forward. With this approach, we can re-examine the importance of variables, as well as correct misleadings caused by backward selection. For instance, it is plausible that a variable, removed in the backward selection, is included in the first stage of the forward selection method. Applying the stepwise selection method, we can include that variable again, for obtaining the best model (R. R. Hocking, 1967). McElreath (2016) revealed that Akaike Information Criteria (AIC) is a well-known information criterion, for evaluating the data-fitting performance of a model. The formula was shown in **Formula 3.6**, where **k** represented the number of parameters in a model; L denoted as the likelihood of a model. The model with the smallest AIC is the best as describing the greatest amount of information with the smallest amount of variables (Bevans, 2021). However, the p-value (listed in **Table 3.7**) suggest that the interaction between attitude towards physical activity, and Pokémon Go related behaviour should be rejected. In addition, the interval of the above interaction term is between -0.112 and 0.114, containing zero, following the drop of AIC from 3244.34 to 3242.341. Thus, this interaction term should be rejected; According to the p-value in **Table 3.7,** the interaction between social sharing level and Pokémon Go related behaviour is doubtful. However, **Figure 3.12** showed that the relationships between the amount of physical activity and Pokémon Go related behaviour were differed by the levels of social sharing. This interaction term, thus, is not rejected. The assumptions of the above model are examined in the following paragraph.

According to the plot, “Residuals vs Fitted Values” in **Figure 3.9**, the pattern of residuals is not obvious. In accordance with **Figure 3.9**, the red line is approximately horizontal across the plot, proving that the residuals contain zero mean. Finally, **Figure 3.9** show residuals spread equally around the red line, proved that the error terms have the constant variance. These evidences suggest that the assumption of linear or curvilinear is acceptable. Outliers, additionally do not exist as no residual standing away from the pattern (Department of Statistics Online Programs, 2018); Although having a light tail, Normal Q-Q plot suggested that the errors of the model have normality concerns (Ford, 2015), but still generally follow the normal distribution (Zach, 2020). Last but not least, as the majority of players are American (Clement, 2021), and the surveying processing were not restricted to species, religions, or researchers’ favour. Thus, we assumed all records are independent. Based on the assumptions plots in **Figure 3.9**, the polynomial linear regression model, demonstrated in the summary, was applied for studying the relations between Pokémon Go and the amount of physical activity. According to “Residuals vs Leverage” plot in **Figure 3.9**, there are some points far away from the main group. However, they both lie inside the Cook distance. Therefore, we do not reject any data points. The detailed final model was demonstrated in **Table 3.8 and Formula 3.10**.

Formula 3.6: formula of AIC score

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 2.5790759 | 0.1960665 | 13.154 | < 2e-16 |
| Age | -0.0738051 | 0.0190248 | -3.879 | 0.000112 |
| PokemonGo\_AppUsage | -0.1623928 | 0.0293782 | -5.528 | 4.17e-08 |
| PokemonGo\_Relate.Behaviour | 0.7551409 | 0.3127692 | 2.414 | 0.015946 |
| Education \* Gender | 0.0387994 | 0.0099712 | 3.891 | 0.000107 |
| Age \* Attitude | 0.0116992 | 0.0034941 | 3.348 | 0.000844 |
| Attitude \* PokemonGo\_Relate.Behaviour | 0.0008241 | 0.0575328 | 0.014 | 0.988575 |
| Social\_sharing \* PokemonGo\_Relate.Behaviour | -0.0233682 | 0.0131280 | -1.780 | 0.075383 |

Table 3.7: Summary of Final Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 2.579123 | 0.195938 | 13.163 | < 2e-16 |
| Age | -0.073988 | 0.014109 | -5.244 | 1.93e-07 |
| PokemonGo\_AppUsage | -0.162347 | 0.029188 | -5.562 | 3.44e-08 |
| PokemonGo\_Relate.Behaviour | 0.759422 | 0.092167 | 8.240 | 5.54e-16 |
| Education \* Gender | 0.038794 | 0.009958 | 3.896 | 0.000105 |
| Age \* Attitude | 0.011734 | 0.002523 | 4.650 | 3.77e-06 |
| Social\_sharing \* PokemonGo\_Relate.Behaviour | -0.023376 | 0.013111 | -1.783 | 0.074902 |

Table 3.8: Summary of Final Model

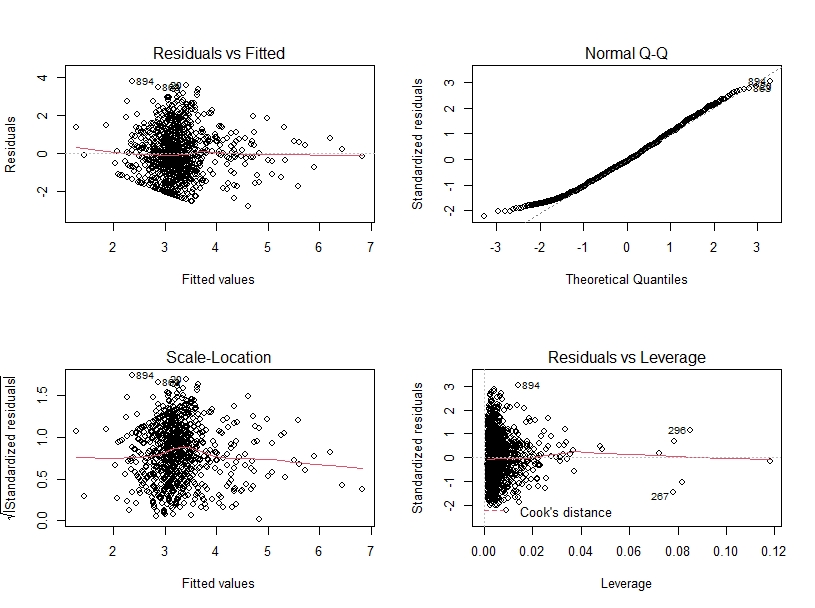
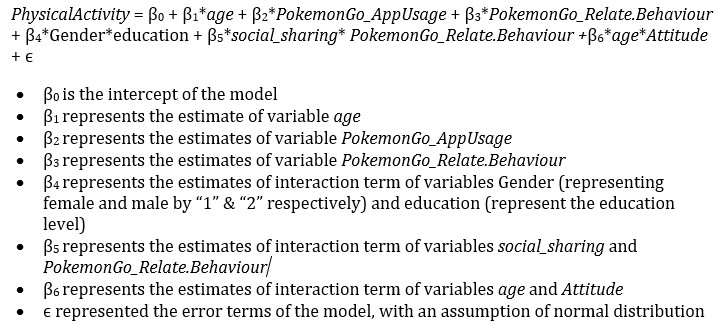


Figure 3.9: Model Assumption Plots



Formula 3.10: Equation of Final Model with Explanation

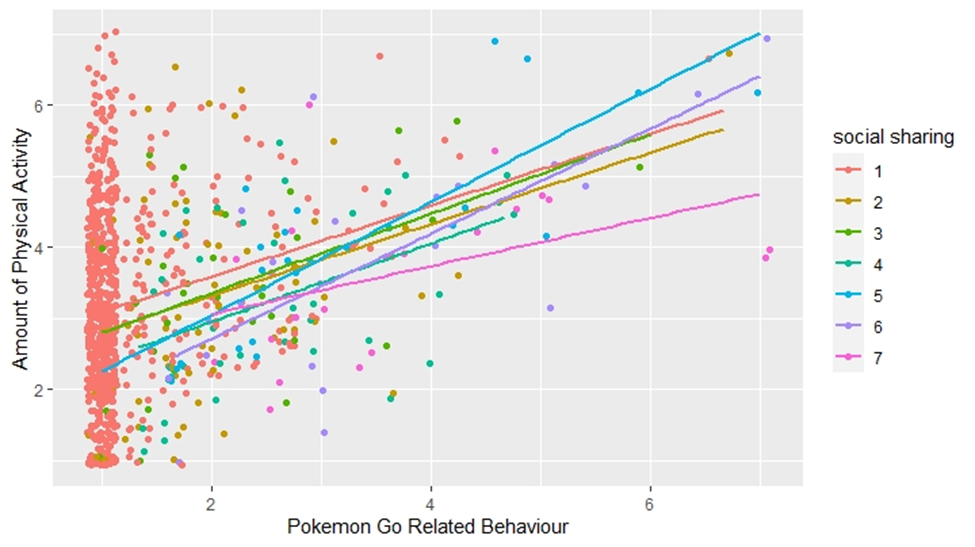


Figure 3.12: Amount of Physical Activity by Age and Education Level

# 

Formula 3.11: Full Model with Explanation (to be continued on next page)



Formula 3.11: Full Model with Explanation

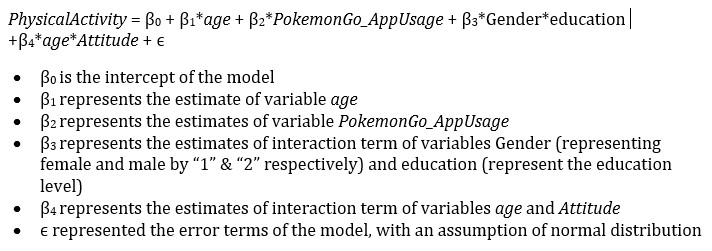
# 4. Analysis Results

With the polynomial regression model constructed before, we answered the questions mentioned in the “research objectives” session. The first is the relationship between the frequency of app usage and the amount of physical activity. Under the summary in **Table 3.8**, the estimate for the number of app usage is -0.162347, meaning that a 1% increase in App usage lowered around 0.16 units the level of the amount of general physical activity. In contrast with what is expected, playing Pokémon Go negatively affect the amount of physical activity. (Considering the levels of sharing achievements on social media, the effects of Pokémon Go related behaviour become negative) This phenomenon can be related to the participation preference of players, the second research aspect. Unlike the variable represented the usage of the Pokémon Go application, **Table 3.8** show that a 1% increase of Pokémon Go related behaviour swells the levels of physical activity by 0.759422 unit. However, the effects become negative if the increase of attitude scores is squared. For explaining and examining this result, the factor of Pokémon Go related behaviour was removed from the model, and the summary is shown in **Table 4.1.** According to **Table 4.1**,theestimated values of Pokémon Go application usage become positive (0.049172) after eliminating the effects of Pokémon Go related behaviour. This represented that the activities associated with Pokémon Goact as a suppressor of the amount of physical activity. Therefore, it is plausible that participants focus on the Pokémon Go related activity, rather than the general physical activity. Despite the variables mentioned before, there are more variables related to the amount of physical activity.

Besides the variables mentioned above, there are more variables affecting players’ physical behaviour. However, **Table 3.8** presents that only age was individually (and negatively) affecting the amount of physical activity. The remaining variables are correlated. The level of social sharing, additionally, was negatively associated with the levels of Pokémon Go related behaviour (indicated by **Table 3.8)**. Despite age, the interaction term between gender and education level individually affects the amount of physical activity. In **Table 3.8**, education level and gender have enough correlations to form an interaction variable between them. If both education level and age increased by 1%, the amount of physical activity would increase around 0.0388 units. Indeed, education level is related to the attitude of participants towards physical activity, but the relationship between gender and attitude scores is questionable***.*** Under **Figure 4.3**, the average attitude score of participants, with the first three education levels, were lower than the participants accepting higher education levels. This indicate that participants with higher education hold a more active attitude towards physical activity, which show the correlation between education level and attitude towards physical activity. Besides education level, **figure 4.3** represented that females have a more positive attitude towards physical activity (with a large area of overlapping), comparing with males. (Note: in the variabledenoting gender of participants’, a female was labelled as "1", while the male was denoted as “2”.) Despite gender and education level, **Table 3.8** shows that there is an interaction term age and attitude towards physical activity, proving these two factors are strongly correlated. Per **Table 3.8**, age reduces the amount of physical activity. However, if both age and attitude increased by 1%, the amount of physical activity increased by approximately 0.01174 units. Therefore, age and education level were correlated with participants’ attitudes towards sports. With the above findings and interaction terms between education level and gender, we conclude that attitude towards physical activity is an important factor.

Table 4.1: Summary of Model without Effects of Pokémon Go Related Behaviour

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 3.072985 | 0.193880 | 15.850 | < 2e-16 |
| age | -0.072152 | 0.014839 | -4.862 | 1.35e-06 |
| PokemonGo\_AppUsage | 0.049172 | 0.021195 | 2.320 | 0.0205 |
| education \* Gender | 0.051381 | 0.010349 | 4.965 | 8.11e-07 |
| age \* Attitude | 0.010964 | 0.002653 | 4.133 | 3.89e-05 |



Formula 4.2: Model without Effects of Pokémon Go Related Behaviour, with Explanations



Figure 4.3: Boxplot for Relationships between Education Level and Attitude towards Sports



Figure 4.4: Boxplot for Relations between Gender and Attitude towards Sports

# 5. Discussion

Pokémon Go is a popular AR mobile game, causing the revolution of mobile games by combing AR technology and mobile game (Lopez German 2016). Players of this game can catch and hatch Pokémon, a virtual creature. Those activities require walking for a certain distance or arriving at a specific location. With this characteristic, some researchers like Kamboj and Krishna (2016) claim that Pokémon Go is an effective mobile game for rising the walking time, following the improvement of public health, including obesity. However, Gabbiadini and Greitemeyer (2018) declare that the effects caused by Pokémon Go are restricted to the activities related to the application, similar to my opinion mentioned before. As mentioned in the “Analysis Results” session, the amount of app usage (with the estimate of -0.162347) reduces the levels of physical activity. The interpretation is Pokémon Go do not lead players more favour in physical activity. There is no obvious relationship between the frequency of playing video games and the amount of general physical health (Baranowski et. al, 2012). There are several possible reasons for this phenomenon. The first one is the game design. For instance, some players hatched the Pokémon eggs while driving (with a speed lower than 10km per hour.) (Ayers et al., 2016), or took public transport for catching a Pokémon. In this way, walking can be unnecessary. Another reason is the preference of participants. Unfortunately, "analysis results" suggest that players are more willing to join the app-related activity, instead of general physical activity. Comparing **Table 3.8**with **Table 4.1**, we have discovered that the estimated value of the amount of app usage becomes positive if the factor of Pokémon Go related activity is removed, demonstrating that app-related activity causes huge effects on the amount of physical activity. We surmise with confidence that players were willing to join an application-related activity, instead of general physical activity. In this way, we discovered that Pokémon Go cannot increase the amount of physical activity directly as the effects are limited. The effects will disappear due to the altering of playing methods, as well as the reduction of players.

The effects of Pokémon Go on the amount of physical activity, however, are uncertain, unstable and inconsistent. **Table 3.8** presented the effects of Pokémon Go related behaviour became negative if interacted with the levels of social sharing, amount of sharing their achievements in social media. If Pokémon Go players are motivated by social purposes, they are less active when playing Pokémon Go (Kaczmarek et. al. 2017). The format of activity, additionally, can be altered with accidents. Due to the COVID-19 pandemic, the entire game is changed for indoor playing (Maher 2020). For example, players have not required to hatch Pokémon through walking. Players, indeed, buy tools-in-game for hatching Pokémon automatically. Players, additionally, used “Incense", a tool for attracting Pokémon, to catch Pokémon without travelling. Last but not least, players can join Raid Battle (activity for catching rare Pokémon) without reaching a Gym. Those changes lower the requirements for walking outside. Despite the company's policy, it is commonly known that most mobile games cannot attract a large number of users forever. Bratuskins (2018) also claim that the lifespan of the mobile game is short. For the above reasons, it is indirect, unstable and unsustainable to use mobile games for public health enhancement. Others, including attitude towards, age, education level and gender are in consideration for discovering solutions optimizing public health.

Besides app usage, social sharing and the amount of participation in app-related activities, both attitudes towards sports, education level, gender and age are factors related to the amount of physical activity. In accordance with **Table 3.8**, the interaction between gender and education level is correlated, increasing the levels of physical activity by approximately 0.039. Nevertheless, **Figure 3.5** indicates there is no positive relationship between education level and the amount of physical activity, unlike the claim that higher education level is related to the lower probability of physical inactivity (Piirtola, et. al., 2016). In my opinion, the increase of estimates is more dependent on the gender of participants. **Figure 3.4** presents that males have a higher amount of physical activity, same as the conclusion from the University of Exeter (2009). This is the reason why the interaction term between gender and education level raised the amount of physical activity. Although this information is not useful as the gender of a person cannot be changed easily, we could find that both gender and education level are related to the attitude towards physical activity.

By **Figure 4.4**, we prove that females have a more positive attitude towards physical activity on average, maximum and minimum. Tomik (2008) suggests females had a more positive attitude towards various areas of sports, compared with males. This result, however, is not absolute. In **Figure 4.4**, we also perceive a considerable area of overlapping when comparing the box plots of males and females. In some cases, males have a more positive towards sports (Koca & Demirhan, 2004). Given this, we focus on the education level. Following **Figure 4.3,** education level is positively associated with the attitude towards physical activity, different from the result obtained by Tomik (2008). Most of the education system include sports subject, allowing students to know more about sports. Moreover, the positive impact of sports has been taught in school. According to Mere-Exposure Effect, people usually have a more positive attitude to an item if they do or contact it (Lee. 2011). Besides education level, the age of a participant is associated with the amount of physical activity, due to the formation of the interaction term between age and levels of physical activity (shown in **Table 3.8**). With the above findings, we conclude that attitude towards sports is a key factor affecting the amount of physical activity, similar to the opinions from Araújo and Dosil (2015). Since from **Table 3.8,** the interaction between age and attitude increase the amount of physical, we believed a positive attitude can help increase the amount of sport, with the rise of age. As the age and gender of a person cannot be controlled by a human, we decided the solution improve people's attitude towards education. In my opinion, the government can provide more funding for tertiary education, as well as increase the free-education year to twelve. Moreover, a teacher can instil a positive attitude towards sports in students. In my opinion, schools can design AR mobile games for instilling a positive attitude in students.

**6. Conclusion and Limitations**

Sport is essential for public health. However, the proportion having exercise regularly in the United Kingdom was limited. Pokémon Go is a well-known AR mobile game, with a huge number of players. A hypothesis was drawn, stating that Pokémon Go is serviceable for public health due to the claims of multiple pieces of research, whereas some researchers argued that the effects caused by Pokémon Go were indirect and unsustainable (Gabbiadini et. al., 2018; Kaczmarek et. al., 2017). Discovering methods for public health improvement, we studied the relationships between Pokémon Go and the amount of general physical activity. The entire study was processed in four aspects: The relations between amount of app usage and the amount of physical activity; the preference of Pokémon Go players; other variables related to the amount of physical activity, as well as the relationships between the attitude and two variables, gender and education level. Constructing required variables, we grouped some series of variables by the mean value of each instance, applying Cronbach's alpha as an internal correlation observation method, as well as evidence for the variables-grouping process. After variables grouping, the polynomial linear model, a linear model that allows interactions between variables, was applied for discovering relations between variables. Stepwise selection methods using AIC as selection criteria were used for model selection. The final model was written in **Formula 3.10**. With this model, we found some interesting facts.

To begin with, Pokémon Go cannot directly increase the amount of physical activity because participants focused on app-related activities. Since players with social motivation tend to be inactive, as well as a mobile game can be changed or collapse. Therefore, Pokémon Go is not a good method of improving public health due to the uncertainty, instability and unsustainability of mobile games. Attitude towards physical activity is a key factor as both education level and age are related to the attitude towards physical activity, with a high correlation between education level and gender. A positive attitude towards sports improves the amount of physical activity with the increasing age. Focusing on the positive between education level and attitude towards sport, we suggested reducing education fees for public health improvement, extending free-education year, and instilling a positive attitude towards sports in students through AR mobile games. There are several limitations during the study, being improved in future. To begin with, the population of the dataset was from America. The statistics from Clement (2021), however, shows that there is a considerable number of players in Great Britain, Japan, Sweden and Canada. It is well known that there are many social differences between countries, including culture and education system. Biases possibly exist if only observing data from players in America. The study should also be conducted in more countries in the future. 981 records were used in our project. However, there are more than eight hundred thousand of active users in America, let alone the whole world. For future study, increasing the population was recommended. This dataset only contains values from questions inside the survey. There is a risk that participants forget the number of times playing Pokémon, or lie on the survey due to shame. The future study is suggested including participants’ data inside the application. Finally, we can try other models, including Poisson, negative-binomial or neural network model.

**7. References**

* Agrawal Raghav. (2021). All you need to know about Polynomial Regression. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/07/all-you-need-to-know-about-polynomial-regression/
* Araújo, A. T., & Dosil, J. (2015). The influence of attitudes toward physical activity and sports. Motriz: Revista de Educação Física, 21, 344-351.
* Ayers, J. W., Leas, E. C., Dredze, M., Allem, J. P., Grabowski, J. G., & Hill, L. (2016). Pokémon GO—a new distraction for drivers and pedestrians. JAMA internal medicine, 176(12), 1865-1866.
* Baranowski, T., Abdelsamad, D., Baranowski, J., O’Connor, T. M., Thompson, D., Barnett, A., ... & Chen, T. A. (2012). Impact of an active video game on healthy children’s physical activity. Pediatrics, 129(3), e636-e642.
* Bradfield & Underhill. (2004). IntroSTAT. 2nd. Department of Statistical Sciences, University of Cape Town. p.181-183
* Bratuskins Martins. (2018). Mobile game lifespan is shortening, but there’s a way to extend it.. Medium. https://medium.com/monetizr/mobile-game-lifespan-is-shortening-but-theres-a-way-to-extend-it-d7e96561b008
* Buraimo Babatunde, Jones Helen & Millward Peter. (2011). "Adult participation in sport: Analysis of the Taking Part Survey". department for culture, media, and sport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/137986/tp-adult-participation-sport-analysis.pdf
* Clement J.. (2021). Number of daily active users (DAU) of Pokémon Go via iPhone in selected countries in January 2021. https://www.statista.com/statistics/604551/pokemon-go-daily-active-users-in-europe/
* Cortina Jose. (1992). What Is Coefficient Alpha? An Examination of Theory and Applications. American Psychological Association, Inc. p.101
* Cronbach, Lee J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika. Springer Science and Business Media LLC. 16 (3): 297-334
* Department of Statistics Online Programs. (2018). “4.2 - Residuals vs. Fits Plot”. The Pennsylvania State University. https://online.stat.psu.edu/stat462/node/117/
* Eysenbach Gunther. (2016). Influence of Pokémon Go on Physical Activity: Study and Implications". Journal of Medical Internet Research. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5174727/
* Fisher Tim. (2021). What Is Augmented Reality?. Lifewire. https://www.lifewire.com/augmented-reality-ar-definition-4155104
* Ford Clay. (2015). Understanding Q-Q Plots. The University of Virginia. https://data.library.virginia.edu/understanding-q-q-plots/
* Gabbiadini, A., Sagioglou, C., & Greitemeyer, T. (2018). Does Pokémon Go lead to a more physically active life style? . Computers in Human Behavior, 84, 258-263.

# Ganti Akhilesh. (2021). "Median". Investopedia. https://www.investopedia.com/terms/m/median.asp

# Glen Stephanie. (2021). Cronbachâ’s Alpha: Simple Definition, Use and Interpretation. Statistics How to. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cronbachs-alpha-spss/

# Goforth Chelsea. (2015). Using and Interpreting Cronbachâ’s Alpha. The University of Virginia Library. https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/

# Hayes Adam. (2021). Stepwise Regression. Investopedia. https://www.investopedia.com/terms/s/stepwise-regression.asp

# Hocking, R. R. (1976). A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32(1), 1â“49. https://doi.org/10.2307/2529336

# Jacquet Jennifer. (2011). The pros & cons of Amazon Mechanical Turk for scientific surveys. Scientific American. https://blogs.scientificamerican.com/guilty-planet/httpblogsscientificamericancomguilty-planet20110707the-pros-cons-of-amazon-mechanical-turk-for-scientific-surveys/

# Kaczmarek, L. D., Misiak, M., Behnke, M., Dziekan, M., & Guzik, P. (2017). The Pikachu effect: Social and health gaming motivations lead to greater benefits of Pokémon GO use. Computers in Human Behavior, 75, 356-363.

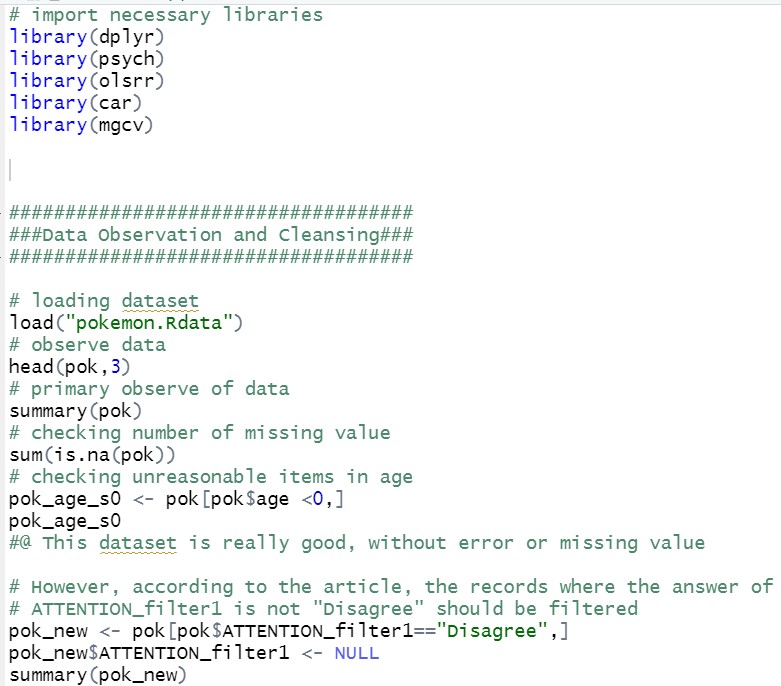
# Kamboj, A. K., & Krishna, S. G. (2017). Pokémon GO: An innovative smartphone gaming application with health benefits. Primary care diabetes, 11(4), 397-399.

# Koca, C., & Demirhan, G. (2004). An examination of high school students' attitudes toward physical education with regard to sex and sport participation. Perceptual and motor skills, 98(3), 754-758.

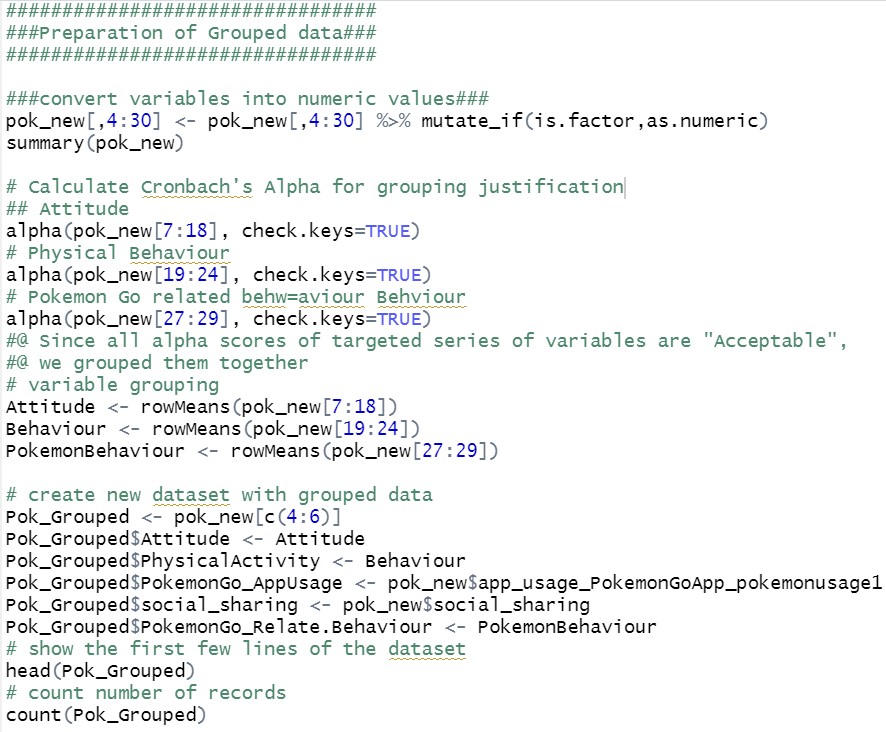
# Kostoulas Achilleas. (2013). On Likert scales, ordinal data and mean values. Achilleas Kostoulas. https://achilleaskostoulas.com/2013/02/13/on-likert-scales-ordinal-data-and-mean-values/

* Lee, A. Y. (2001). The mere exposure effect: An uncertainty reduction explanation revisited. Personality and Social Psychology Bulletin, 27(10), 1255-1266.
* Lopez German. (2016). Pokémon Go, explained. Vox. https://www.vox.com/2016/7/11/12129162/pokemon-go-android-ios-game
* Maher, C. (2020). How COVID-19 transformed Pokémon GO into “Pokémon stay-at-home”. ArsTechnica. Last modified July, 13.
* McElreath Richard. (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press. p.189
* McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. Behavior research methods, 52(6), 2287-2305.
* Piirtola, M., Kaprio, J., Kujala, U. M., Heikkilä, K., Koskenvuo, M., Svedberg, P., ... & Ropponen, A. (2016). Association between education and future leisure-time physical inactivity: a study of Finnish twins over a 35-year follow-up. BMC Public Health, 16(1), 1-11.
* Poplinski, A., Tüttelmann, F., Kanber, D., Horsthemke, B., & Gromoll, J. (2010). Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. International journal of andrology, 33(4), 642-649.
* Reilly Luke. (2017). Pokémon GO Coming to Smartphones. IGN Entertainment. https://www.ign.com/articles/2015/09/10/pokemon-go-coming-to-smartphones
* Seippel Ornulf. (2006). The meanings of sport: fun, health, beauty or community?. Sport in Society, 9(1), 51-70.
* Sengupta Somak. (2020). Gamma Distribution Explained | What is Gamma Distribution? . Great Learning. https://www.mygreatlearning.com/blog/gamma-distribution/
* Smith, C. (2017). 80 amazing Pokemon Go statistics. RetrievedNovember 17, 2017, from https://expandedramblings.com/index.php/pokemon-go-statistics/.
* Stacy, E. W. (1962). A generalization of the gamma distribution. The Annals of mathematical statistics, 1187-1192.
* Tomik, R. (2008). Adolescent attitudes towards sport depending on school level, gender and school sports club membership. Journal of Human kinetics, 20(1), 121-130.
* University of Exeter. (2009). Lifelong Gender Difference In Physical Activity Revealed. Science Daily. www.sciencedaily.com/releases/2009/01/090105190740.htm
* Webster Andrew. (2015). "With Pokémon Go, Nintendo is showing that it takes mobile seriously". The Verge. https://www.theverge.com/2015/9/10/9300101/pokemon-go-nintendo-mobile-games
* Wong, F. Y. (2017). Influence of Pokémon Go on physical activity levels of university players: a cross-sectional study. International journal of health geographics, 16(1), 1-12.
* Zach. (2020). How to Interpret a Scale-Location Plot (With Examples). Scatology. https://www.statology.org/scale-location-plot/

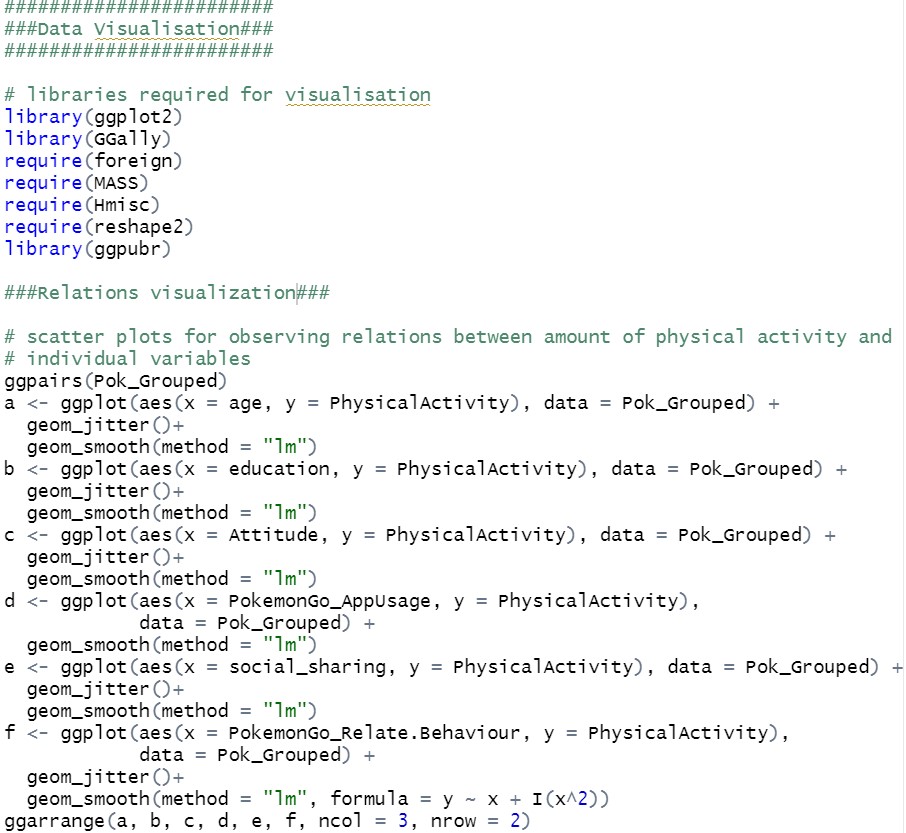
**8. Appendix**

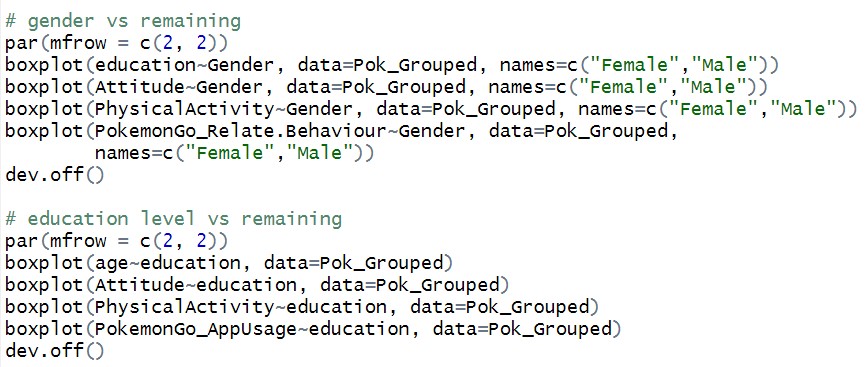


We first discover some unreasonable data, missing value or outliers, following the filtering of records which have been classified as non-focus



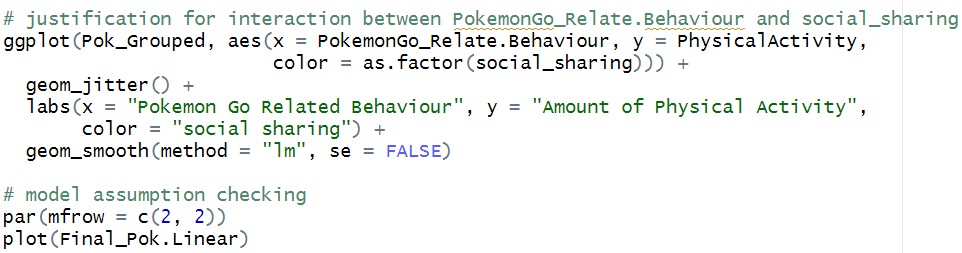
We first use Alpha Score to justify the grouping, then group variables together.



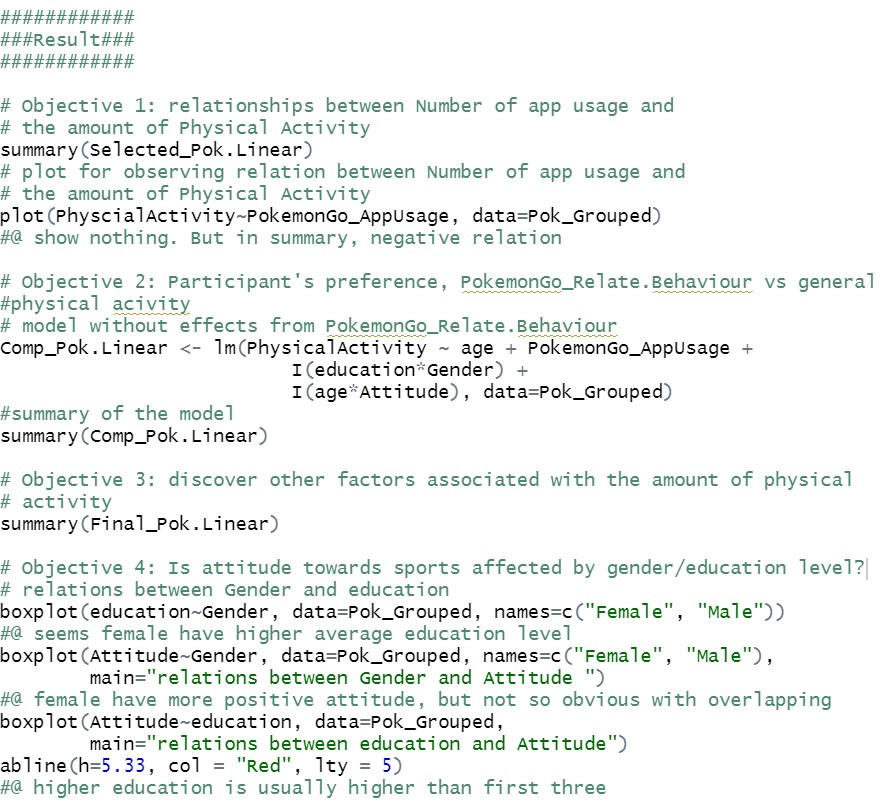


Visualise data for data discovery





Construct the full model, following the variable selection. (Using Stepwise selection method, and AIC as model selection criteria)



Assign model summaries and plots to different sessions, for answering all research objectives mentioned before.