

BigIdeas_Analysis

Kylie Balotin

2025-12-11

Analysis of Big Ideas Lab Survey Data

This file is to document the analysis process for the Big Ideas Lab survey data.

First I will load the necessary packages for analysis.

Then I will load the dataset to be analyzed.

```
big_ideas_data <- read_excel("/Users/kyliebalotin/Github/Coursera-Case-Study---Bellabeat/Analysis/Big_Ideas_Lab_Survey_Data.xlsx")
```

In the following code section, I'm combining the separate columns of the different phone and wearable types into a single column (each) that contains a string of the name of the device manufacturer.

```
#Changing how the data is organized
#initiate some of the new column names
phone_type <- rep("phone", (length(big_ideas_data$Gender)))
wearable_type <- rep("wearable", (length(big_ideas_data$Gender)))

#Creating a new data attribute listing what kind of phone each submission has
for (i in 1:length(big_ideas_data$Gender)) {
  if (big_ideas_data[i, 1]==1) {
    phone_type[i] = "iPhone"
  }
  else if (big_ideas_data[i, 2]==1) {
    phone_type[i] = "Android"
  }
  else if (big_ideas_data[i, 3]==1) {
    phone_type[i] = "Other"
  }
  else if (big_ideas_data[i, 4]==1) {
    phone_type[i] = "None"
  }
}

#Checking that the only for phone type are: None, iPhone, Android, and Other
unique(phone_type)
```

```
## [1] "Android" "iPhone" "None" "Other"
```

```

#Creating a new data attribute listing what kind of wearable each submission has
for (i in 1:length(big_ideas_data$Gender)) {
  if (big_ideas_data[i, 12]==1) {
    wearable_type[i] = "None"
  }
  else if (big_ideas_data[i, 13]==1) {
    wearable_type[i] = "Fitbit"
  }
  else if (big_ideas_data[i, 14]==1) {
    wearable_type[i] = "Apple Watch"
  }
  else if (big_ideas_data[i, 15]==1) {
    wearable_type[i] = "Garmin"
  }
  else if (big_ideas_data[i, 16]==1) {
    wearable_type[i] = "Samsung"
  }
  else if (big_ideas_data[i, 17]==1) {
    wearable_type[i] = "Other"
  }
}

```

```

#Checking that the options for wearable type are: None, Fitbit, Apple Watch, Garmin, Samsung, and Other
unique(wearable_type)

```

```

## [1] "None"          "Apple Watch" "Fitbit"      "Samsung"     "Other"
## [6] "Garmin"

```

```

#Merge these new columns back into original dataframe
big_ideas_data_org <- data.frame(big_ideas_data, phone_type, wearable_type)
#View(big_ideas_data_org)

```

Next, I'm creating some subsets of the data (based on gender). Bellabeat's primary consumer demographic is women, so I want to make sure I am able to capture the results specifically for survey participants who identify as female.

The next coding section documents the different calculations I performed on the dataset. I first look at the demographic information about the survey participants to have a better idea about how the sample population might relate to the larger population.

Then I looked at the survey's responses regarding how participants are using their smart phones and wearables (ex: fitness tracking, sleep monitoring, etc) and reasons why the participants might not own a wearable. I also calculated how many of the participants own phones and wearables. I have performed these calculations for the whole survey sample population and the subset of the sample population that identify as female.

```

#Find out some demographic information about the survey participants

#Finding number of participants who identify as different genders
unique_g <- unique(big_ideas_data_org$Gender)
count_unique_g <- rep(0, length(unique_g))
for (i in 1:length(unique_g)) {
  count_unique_g[i] <- sum(big_ideas_data_org$Gender==unique_g[i], na.rm=TRUE)
}

```

```
count_g <- data.frame(unique_g, count_unique_g)
count_g
```

```
##           unique_g count_unique_g
## 1           Female           871
## 2            Male           478
## 3   Gender Fluid/Queer           3
## 4 Other Gender - Not Disclosed           8
## 5           Non-binary           4
## 6   Other Gender - Disclosed           3
## 7   Transgender Male           1
```

#Finding out number of participants in different age groups

```
unique_age <- unique(big_ideas_data_org$generation_age_group)
count_unique_age <- rep(0, length(unique_age))
```

```
count_unique_age_f <- rep(0, length(unique_age))
```

```
for (i in 1:length(unique_age)) {
  count_unique_age[i] <- sum(big_ideas_data_org$generation_age_group==unique_age[i])
  count_unique_age_f[i] <- sum(big_ideas_data_org_f$generation_age_group==unique_age[i])
}
count_age <- data.frame(unique_age, count_unique_age, count_unique_age_f)
count_age
```

```
##   unique_age count_unique_age count_unique_age_f
## 1    42_57           460           334
## 2    26_41           176           116
## 3    58_76           579           355
## 4     77+           129            49
## 5   18_25            24            17
```

#Level of Education

```
unique_edu <- unique(big_ideas_data_org$Highest.level.of.education)
count_unique_edu <- rep(0, length(unique_edu))
```

```
count_unique_edu_f <- rep(0, length(unique_edu))
```

```
for (i in 1:length(unique_edu)) {
  count_unique_edu[i] <- sum(big_ideas_data_org$Highest.level.of.education==unique_edu[i])
  count_unique_edu_f[i] <- sum(big_ideas_data_org_f$Highest.level.of.education==unique_edu[i])
}
count_edu <- data.frame(unique_edu, count_unique_edu, count_unique_edu_f)
count_edu
```

```
##           unique_edu count_unique_edu count_unique_edu_f
## 1   Graduate degree           515           314
## 2 Some college but no degree           231           158
## 3   College graduate           545           353
## 4   High school graduate            74            45
## 5   Less than high school            3             1
```

```

#Employment Status
unique_emp <- unique(big_ideas_data_org$Employment.Status)
count_unique_emp <- rep(0, length(unique_emp))

count_unique_emp_f <- rep(0, length(unique_emp))

for (i in 1:length(unique_emp)) {
  count_unique_emp[i] <- sum(big_ideas_data_org$Employment.Status==unique_emp[i])
  count_unique_emp_f[i] <- sum(big_ideas_data_org_f$Employment.Status==unique_emp[i])
}
count_emp <- data.frame(unique_emp, count_unique_emp, count_unique_emp_f)
count_emp

```

```

##                unique_emp count_unique_emp count_unique_emp_f
## 1      Employed full-time           630           417
## 2    Retired, not looking for work           400           207
## 3    Disabled, not able to work           120            85
## 4      Employed part-time           108            76
## 5 Not employed, but looking for work            41            27
## 6 Not employed, not looking for work            69            59

```

```

#Race/Ethnicity
eth_simp <- c("Black/African American", "Asian/Asian American", "Hispanic", "White/Caucasian", "Other")
count_eth <- rep(0, length(eth_simp))
count_eth_f <- rep(0, length(eth_simp))

for (i in 1:length(eth_simp)) {
  x <- i+45
  count_eth[i] <- sum(big_ideas_data_org[x], na.rm=TRUE)
  count_eth_f[i] <- sum(big_ideas_data_org_f[x], na.rm=TRUE)
}

count_ethnicity <- data.frame(eth_simp, count_eth, count_eth_f)
count_ethnicity

```

```

##                eth_simp count_eth count_eth_f
## 1 Black/African American           390           306
## 2   Asian/Asian American            60            32
## 3             Hispanic            78            45
## 4   White/Caucasian           826           477
## 5              Other            77            55

```

#Looking at reasons why participants use smartphones/wearables

#Use of Smart Phones

```

activity_phone <- colnames(big_ideas_data_org[5:9])
activity_phone_simp <- c("Not tracking", "Fitness and workout monitoring", "Health tracking", "Sleep monitoring")
count_activity_phone <- rep(0, length(activity_phone))
count_activity_phone_f <- rep(0, length(activity_phone))
for (i in 1:length(activity_phone)) {
  count_activity_phone[i] <- sum(big_ideas_data_org[i+4], na.rm=TRUE)
  count_activity_phone_f[i] <- sum(big_ideas_data_org_f[i+4], na.rm=TRUE)
}

```

```
count_activity_ph <- data.frame(activity_phone_simp, count_activity_phone, count_activity_phone_f)
count_activity_ph
```

```
##           activity_phone_simp count_activity_phone count_activity_phone_f
## 1           Not tracking           551           343
## 2 Fitness and workout monitoring           627           405
## 3           Health tracking           284           180
## 4           Sleep monitoring           269           171
## 5                Other           143            90
```

#Reasons for not owning a wearable

```
reason_not_simp <- c("Don't own one yet", "No particular reason", "Too expensive", "Too hard to read", "
count_reason_not <- rep(0, length(reason_not_simp))
count_reason_not_f <- rep(0, length(reason_not_simp))
for (i in 1:length(reason_not_simp)) {
  count_reason_not[i] <- sum(big_ideas_data_org[i+17], na.rm=TRUE)
  count_reason_not_f[i] <- sum(big_ideas_data_org_f[i+17], na.rm=TRUE)
}
count_reason_no <- data.frame(reason_not_simp, count_reason_not, count_reason_not_f)
count_reason_no
```

```
##           reason_not_simp count_reason_not count_reason_not_f
## 1 Don't own one yet           58           36
## 2 No particular reason        125           63
## 3 Too expensive              178          123
## 4 Too hard to read            17            9
## 5 Don't trust they work correctly           38           22
## 6 Don't know enough           54           34
## 7 Not interested in tracking        127           77
## 8 Privacy                     68           33
## 9 Other                       85           48
```

#Use of wearable

```
activity_wear_simp <- c("Apps", "Fitness and workout monitoring", "Sleep monitoring", "Health tracking"
unique_reasons <- unique(big_ideas_data_org$Wearable.device.usage.by.reason_Apps..social.media..news..e
unique_reasons
```

```
## [1] NA "Secondary reason"
## [3] "Not a reason" "Not applicable to my device(s)"
## [5] "Main reason"
```

```
count_main <- rep(0, length(activity_wear_simp))
count_sec <- rep(0, length(activity_wear_simp))
count_notreason <- rep(0, length(activity_wear_simp))
count_na <- rep(0, length(activity_wear_simp))

count_main_f <- rep(0, length(activity_wear_simp))
count_sec_f <- rep(0, length(activity_wear_simp))
count_notreason_f <- rep(0, length(activity_wear_simp))
count_na_f <- rep(0, length(activity_wear_simp))
for (i in 1:length(activity_wear_simp)) {
```

```

count_main[i] <- sum(big_ideas_data_org[i+26]=="Main reason", na.rm=TRUE)
count_sec[i] <- sum(big_ideas_data_org[i+26]=="Secondary reason", na.rm=TRUE)
count_notreason[i] <- sum(big_ideas_data_org[i+26]=="Not a reason", na.rm=TRUE)
count_na[i] <- sum(big_ideas_data_org[i+26]=="Not applicable to my device(s)", na.rm=TRUE)

count_main_f[i] <- sum(big_ideas_data_org_f[i+26]=="Main reason", na.rm=TRUE)
count_sec_f[i] <- sum(big_ideas_data_org_f[i+26]=="Secondary reason", na.rm=TRUE)
count_notreason_f[i] <- sum(big_ideas_data_org_f[i+26]=="Not a reason", na.rm=TRUE)
count_na_f[i] <- sum(big_ideas_data_org_f[i+26]=="Not applicable to my device(s)", na.rm=TRUE)
}

count_activity_wear <- data.frame(activity_wear_simp, count_main, count_sec, count_notreason, count_na)
count_activity_wear

```

```

##           activity_wear_simp count_main count_sec count_notreason count_na
## 1                Apps          75      157          430          140
## 2 Fitness and workout monitoring    522      188           75           17
## 3                Sleep monitoring    126      265          338           73
## 4                Health tracking    228      272          210           92
## 5                Communication    291      229          200           82
## 6 Music/audiobooks/podcasts         50      161          433          158
## 7                Navigation         72      187          385          158
## 8                Fashion          36      108          568           90

```

```

count_activity_wear_f <- data.frame(activity_wear_simp, count_main_f, count_sec_f, count_notreason_f, count_na_f)
count_activity_wear_f

```

```

##           activity_wear_simp count_main_f count_sec_f count_notreason_f
## 1                Apps          46        97          293
## 2 Fitness and workout monitoring    358       125           44
## 3                Sleep monitoring     69       187          232
## 4                Health tracking    150       185          140
## 5                Communication    196       142          144
## 6 Music/audiobooks/podcasts         30       104          285
## 7                Navigation         49       108          266
## 8                Fashion          28        74          378
## count_na_f
## 1          102
## 2           11
## 3           50
## 4           63
## 5           56
## 6          119
## 7          115
## 8           58

```

```

#Counting Phone and Wearable Ownership
phones <- unique(phone_type)
count_phones <- rep(0, length(phones))
count_phones_f <- rep(0, length(phones))

for (i in 1:length(phones)) {

```

```

count_phones[i] <- sum(big_ideas_data_org$phone_type==phones[i], na.rm=TRUE)
count_phones_f[i] <- sum(big_ideas_data_org_f$phone_type==phones[i], na.rm=TRUE)
}
count_ph <- data.frame(phones, count_phones, count_phones_f)
count_ph

```

```

##   phones count_phones count_phones_f
## 1 Android         436           290
## 2 iPhone          894           559
## 3   None           25            14
## 4   Other           13             8

```

```

wearables <- unique(wearable_type)
count_wearables <- rep(0, length(wearables))
count_wearables_f <- rep(0, length(wearables))

for (i in 1:length(wearables)) {
  count_wearables[i] <- sum(big_ideas_data_org$wearable_type==wearables[i], na.rm=TRUE)
  count_wearables_f[i] <- sum(big_ideas_data_org_f$wearable_type==wearables[i], na.rm=TRUE)
}
count_w <- data.frame(wearables, count_wearables, count_wearables_f)
count_w

```

```

##   wearables count_wearables count_wearables_f
## 1   None          566           333
## 2 Apple Watch     312           200
## 3   Fitbit        337           246
## 4   Samsung        54            35
## 5   Other         50            33
## 6   Garmin        49            24

```

Finally, I plot the different calculations in order to show my findings visually. The following code checks that the demographic information does not change too much when the participant population is subset to look at only female-identifying participants.

```

#Demographic information
#Pie Chart of Participants' Genders
demographic_g <- ggplot(count_g, aes(x="", y=count_unique_g, fill=unique_g)) + geom_bar(stat="identity")

#Pie Chart of Participants' Ages
demographic_age <- ggplot(count_age, aes(x="", y=count_unique_age, fill=unique_age)) + geom_bar(stat="identity")

#Female population age demographics
demographic_age_f <- ggplot(count_age, aes(x="", y=count_unique_age_f, fill=unique_age)) + geom_bar(stat="identity")

#Level of Education demographics
demographic_edu <- ggplot(count_edu, aes(x="", y=count_unique_edu, fill=unique_edu)) + geom_bar(stat="identity")

#Female population edu demographics
demographic_edu_f <- ggplot(count_edu, aes(x="", y=count_unique_edu_f, fill=unique_edu)) + geom_bar(stat="identity")

#Employment Status demographics

```

```

demographic_emp <- ggplot(count_emp, aes(x="", y=count_unique_emp, fill=unique_emp)) + geom_bar(stat="identity")

#Female population employment status demographics
demographic_emp_f <- ggplot(count_emp, aes(x="", y=count_unique_emp_f, fill=unique_emp)) + geom_bar(stat="identity")

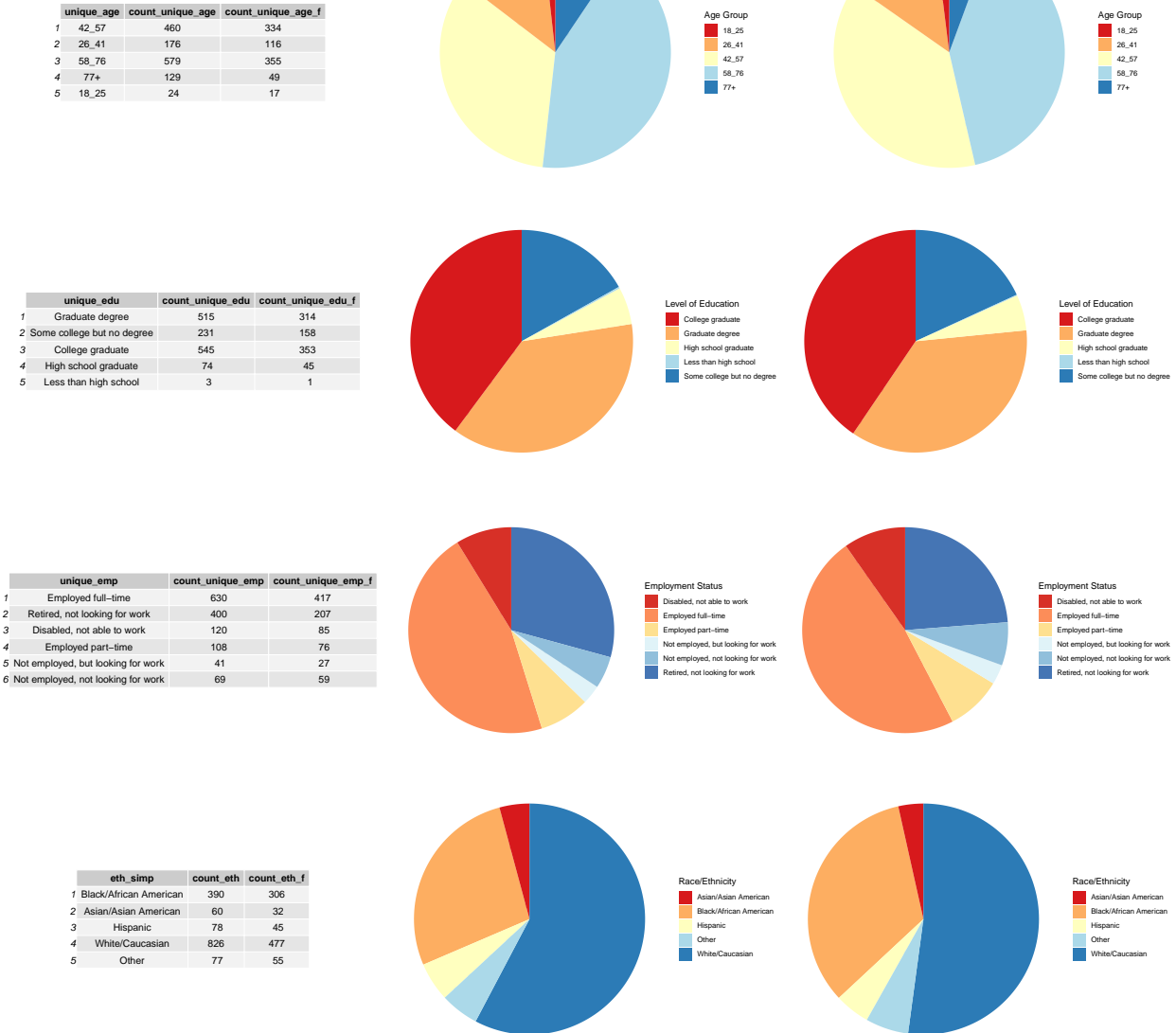
#Race/ethnicity demographics
demographic_eth <- ggplot(count_ethnicity, aes(x="", y=count_eth, fill=eth_simp)) + geom_bar(stat="identity")

#Female population employment status demographics
demographic_eth_f <- ggplot(count_ethnicity, aes(x="", y=count_eth_f, fill=eth_simp)) + geom_bar(stat="identity")

grid.arrange(tableGrob(count_age), demographic_age, demographic_age_f,
              tableGrob(count_edu), demographic_edu, demographic_edu_f,
              tableGrob(count_emp), demographic_emp, demographic_emp_f,
              tableGrob(count_ethnicity), demographic_eth, demographic_eth_f,
              nrow=4, ncol=3,
              top="Demographic Information About Survey Participants", bottom = "Plots in middle column")

```


Demographic Information About Survey Participants



Plots in middle column represent the whole survey participant population; plots in the right column represent survey participants who identify as female

Then I looked at how the participants are using their phones and wearables, as well as why some participants do not own a wearable device. The majority of the participants own a smart phone, but about half of the participants do not own a wearable. The majority of phone owning participants use their phone for fitness and workout monitoring, followed closely by participants who do not use their phones to track anything. Some participants will use their phones for sleep and health monitoring.

Among the whole survey participant population, primary reason why participants do not own a wearable is the price (i.e. wearables are too expensive), followed closely by participants not being interested in tracking activity and participants do not have a specific reason why they don't own one.

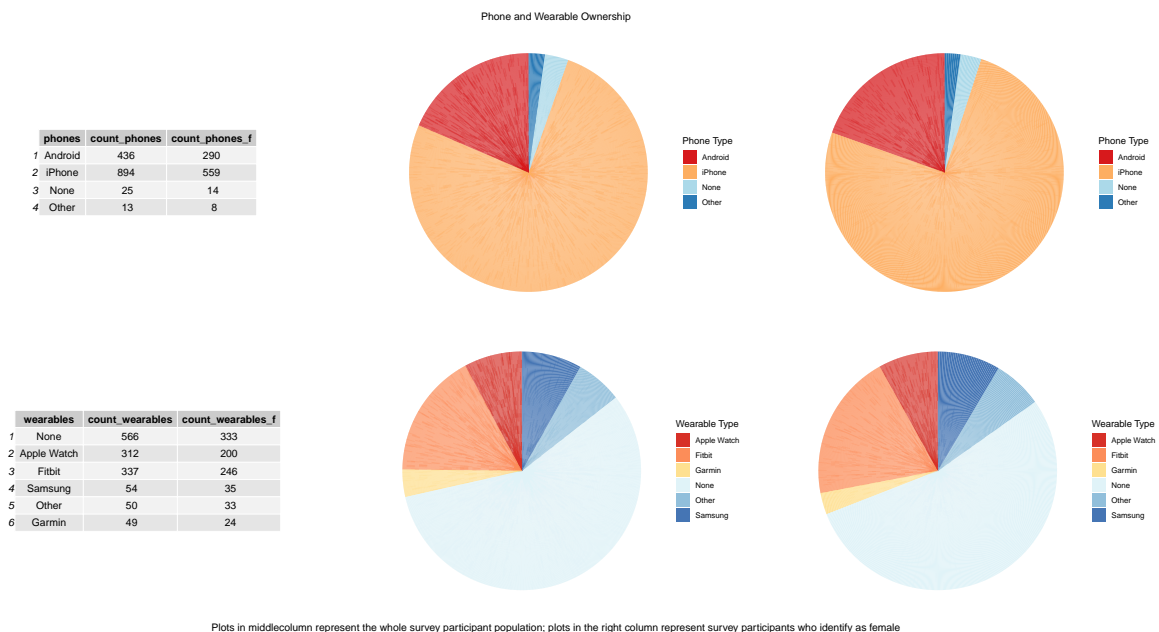
The main reason why participants use wearable devices is for fitness and workout monitoring, followed by communication and health tracking, respectively. Sleep monitoring, health tracking, and communication are the top three secondary uses of wearables, respectively. In terms of not being considered a main use of a wearable, fashion was ranked the highest. Music/audiobooks/podcasts, navigation, and other apps are not common features of wearables.

All of these results do not change when the survey population is filtered to only female-identifying participants.

#Breakdown of Phone and Wearable Ownership

```
own_phone <- ggplot(big_ideas_data_org, aes(x="", y=phone_type, fill=phone_type)) + geom_bar(stat="identity")
own_phone_f <- ggplot(big_ideas_data_org_f, aes(x="", y=phone_type, fill=phone_type)) + geom_bar(stat="identity")
own_wear <- ggplot(big_ideas_data_org, aes(x="", y=wearable_type, fill=wearable_type)) + geom_bar(stat="identity")
own_wear_f <- ggplot(big_ideas_data_org_f, aes(x="", y=wearable_type, fill=wearable_type)) + geom_bar(stat="identity")

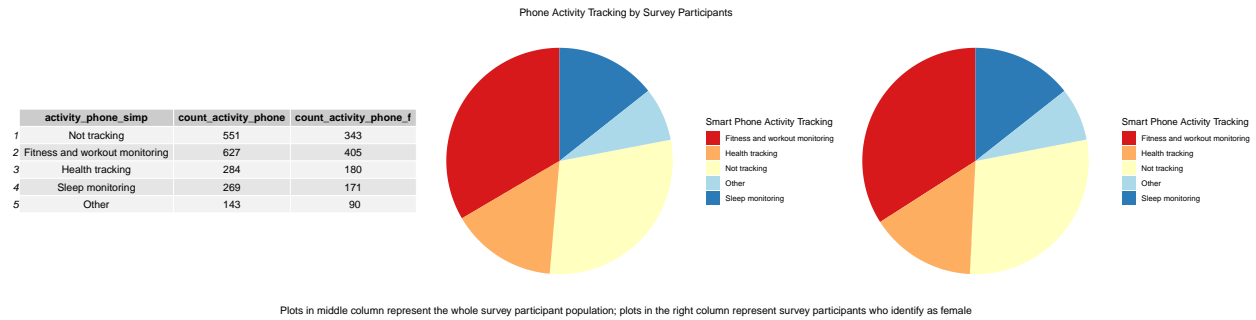
grid.arrange(tableGrob(count_ph), own_phone, own_phone_f,
              tableGrob(count_w), own_wear, own_wear_f,
              nrow=2, ncol=3,
              top = "Phone and Wearable Ownership",
              bottom = "Plots in middlecolumn represent the whole survey participant population; plots in right column represent female-identifying participants")
```



#Breakdown of Phone Usage

```
phone_use <- ggplot(count_activity_ph, aes(x="", y=count_activity_phone, fill=activity_phone_simp)) + geom_bar(stat="identity")
phone_use_f <- ggplot(count_activity_ph, aes(x="", y=count_activity_phone_f, fill=activity_phone_simp)) + geom_bar(stat="identity")

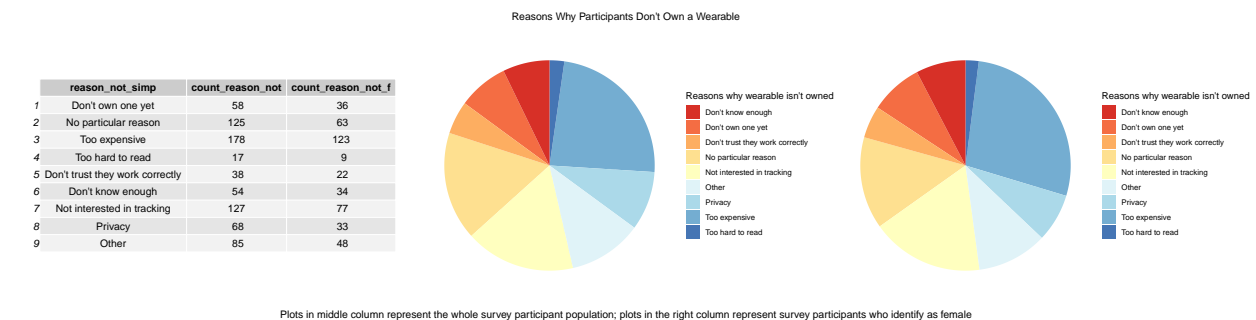
grid.arrange(tableGrob(count_activity_ph), phone_use, phone_use_f,
              nrow=1, ncol=3,
              top="Phone Activity Tracking by Survey Participants", bottom = "Plots in middle column represent the whole survey participant population; plots in right column represent female-identifying participants")
```



```
#Breakdown of Reasons why participants don't own wearables
wearable_no <- ggplot(count_reason_no, aes(x="", y=count_reason_not, fill=reason_not_simp)) + geom_bar()

wearable_no_f <- ggplot(count_reason_no, aes(x="", y=count_reason_not_f, fill=reason_not_simp)) + geom_bar()

grid.arrange(tableGrob(count_reason_no), wearable_no, wearable_no_f,
              nrow=1, ncol=3,
              top="Reasons Why Participants Don't Own a Wearable", bottom = "Plots in middle column repr
```



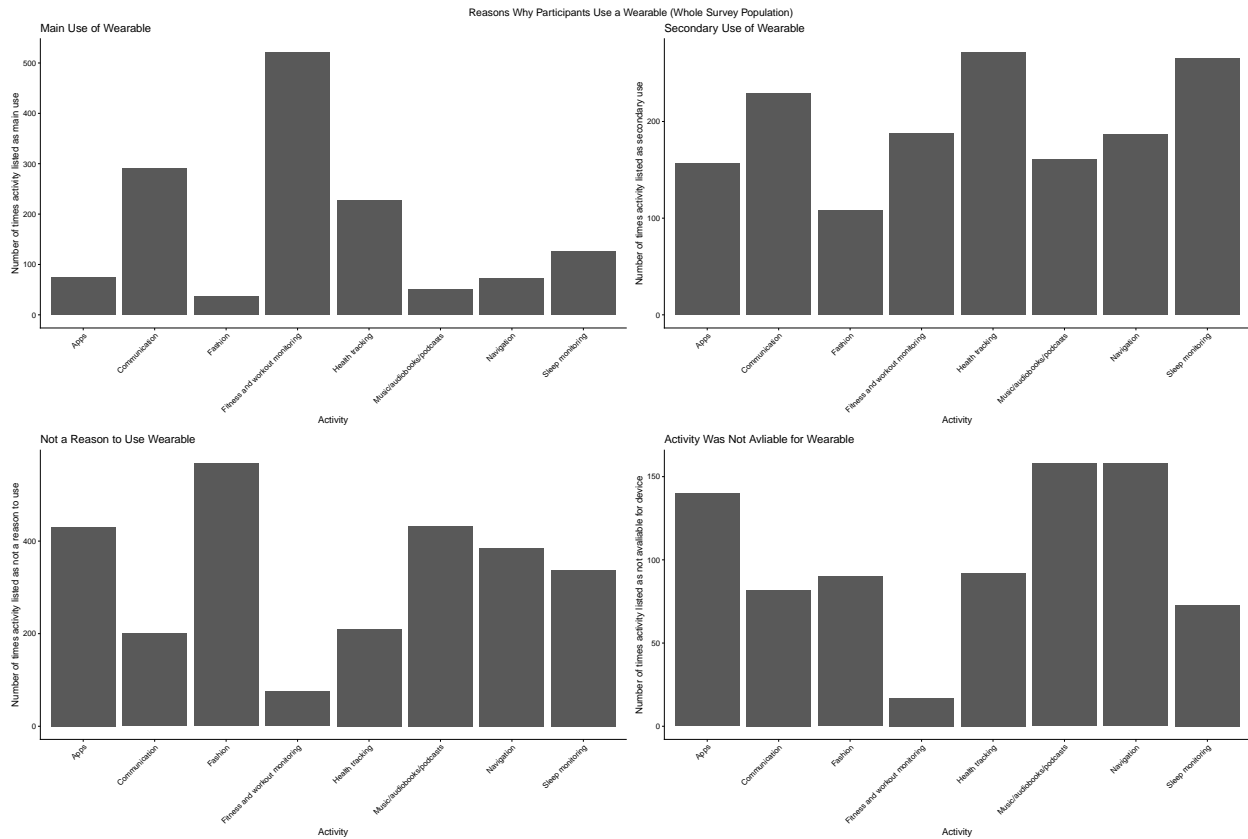
```
#Breakdown of Wearable Usage
wear_use_main <- ggplot(count_activity_wear, aes(x=activity_wear_simp, y=count_main)) + geom_bar(stat="identity")

wear_use_sec <- ggplot(count_activity_wear, aes(x=activity_wear_simp, count_sec)) + geom_bar(stat="identity")

wear_use_not <- ggplot(count_activity_wear, aes(x=activity_wear_simp, count_notreason)) + geom_bar(stat="identity")

wear_use_na <- ggplot(count_activity_wear, aes(x=activity_wear_simp, count_na)) + geom_bar(stat="identity")

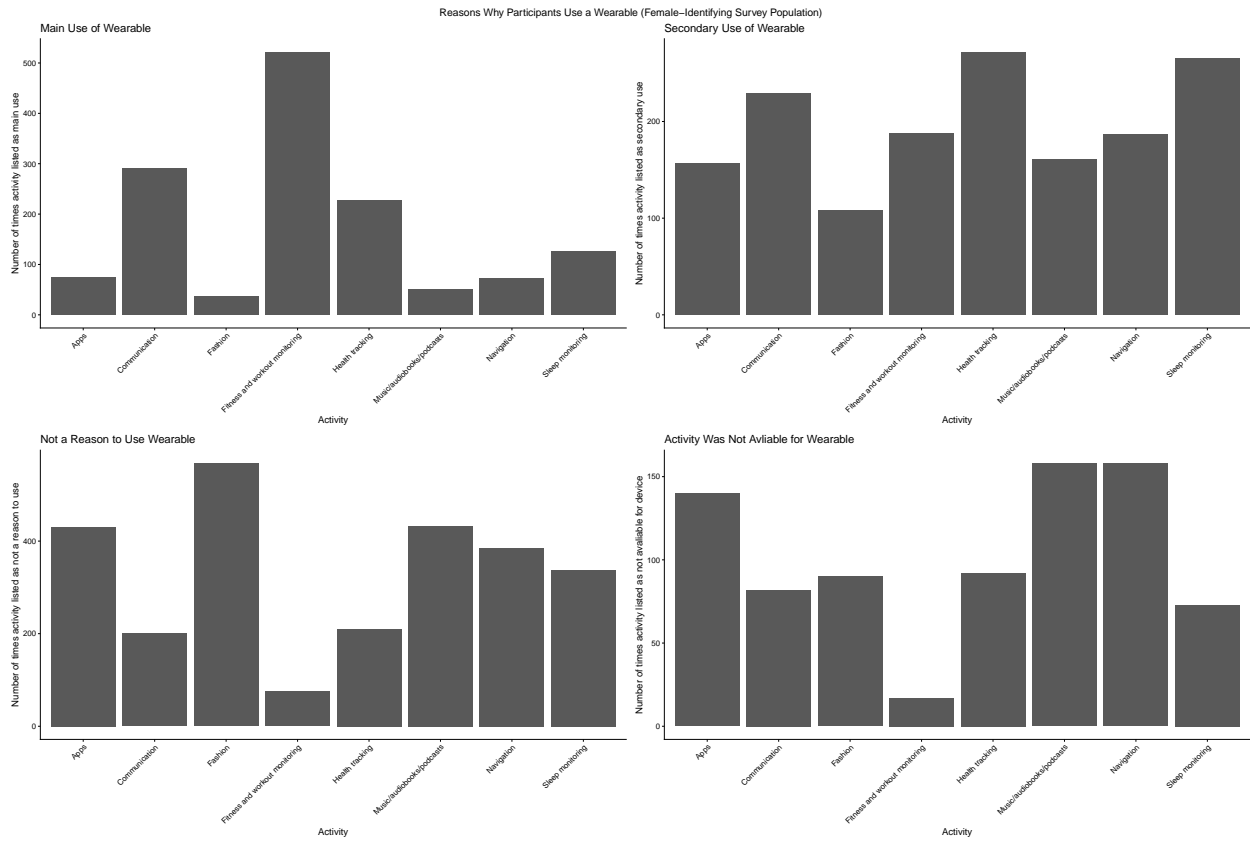
grid.arrange(wear_use_main, wear_use_sec,
              wear_use_not, wear_use_na,
              tableGrob(count_activity_wear),
              nrow=3, ncol=2,
              top="Reasons Why Participants Use a Wearable (Whole Survey Population)")
```



	activity_wear_simp	count_main	count_sec	count_notreason	count_na
1	Apps	75	157	430	140
2	Fitness and workout monitoring	522	188	75	17
3	Sleep monitoring	126	265	338	73
4	Health tracking	228	272	210	92
5	Communication	291	229	200	82
6	Music/audiobooks/podcasts	50	161	433	158
7	Navigation	72	187	385	158
8	Fashion	36	108	568	90

```
wear_use_main_f <- ggplot(count_activity_wear_f, aes(x=activity_wear_simp, y=count_main)) + geom_bar(stat="count")
wear_use_sec_f <- ggplot(count_activity_wear_f, aes(x=activity_wear_simp, count_sec)) + geom_bar(stat="count")
wear_use_not_f <- ggplot(count_activity_wear_f, aes(x=activity_wear_simp, count_notreason)) + geom_bar(stat="count")
wear_use_na_f <- ggplot(count_activity_wear_f, aes(x=activity_wear_simp, count_na)) + geom_bar(stat="count")

grid.arrange(wear_use_main_f, wear_use_sec_f,
              wear_use_not_f, wear_use_na_f,
              tableGrob(count_activity_wear_f),
              nrow=3, ncol=2,
              top="Reasons Why Participants Use a Wearable (Female-Identifying Survey Population)")
```



	activity_wear_simp	count_main_f	count_sec_f	count_notreason_f	count_na_f
1	Apps	46	97	293	102
2	Fitness and workout monitoring	358	125	44	11
3	Sleep monitoring	69	187	232	50
4	Health tracking	150	185	140	63
5	Communication	196	142	144	56
6	Music/audiobooks/podcasts	30	104	285	119
7	Navigation	49	108	266	115
8	Fashion	28	74	378	58