Final_Project

Songyi Ahn, Nakyung Lee, Nishat Parvez

Common Work

Loading Packages

Loading and Renaming the dataset

```
# Load the file
file_path <- here("data", "hints6_public.rda")
load(file_path)

# Change the file name to 'hints'
hints <- public
```

RQ1. What is the distribution of demographic/socio-economic categories, including age, gender, ethnicity, education, and income level, among individuals who do and do not use Telehealth services?

RQ1_1. Data Cleaning

```
#this is for Nishat's work
```

RQ2. How do social media usage frequency and health information engagement levels on social media (low, medium, high) differ between individuals who use and do not use Telehealth services?

RQ2_1. Data Cleaning

```
# Selecting relevant columns for RQ2

data_nl <- hints %>%
  select(HHID, ReceiveTelehealthCare, SocMed_Visited,
         SocMed_WatchedVid, SocMed_Interacted, SocMed_SharedPers)

# Renaming each column

data_nl <- data_nl %>%
  rename(telehealth = ReceiveTelehealthCare,
         sm_usage_freq = SocMed_Visited,
         sm_watching = SocMed_WatchedVid,
         sm_interact =
        SocMed_Interacted,
        sm_sharing = SocMed_SharedPers)

# Recoding responses

data_nl <- data_nl %>%
  mutate(telehealth =
           recode(telehealth,
                  "No telehealth visits in the past 12 months" = "No",
                  "Yes, by video" = "Yes",
                  "Yes, by phone call (voice only with no video)" = "Yes",
                  "Yes, some by video and some by phone call" = "Yes",
                  "Missing data (Web partial - Question Never Seen)" =
                  "Missing", "Missing data (Not Ascertained)" =
                  "Missing"))

# Creating a dataset for 5 point Likert scale

data_nl_2 <- data_nl %>%
  mutate(sm_usage_freq =
           recode(sm_usage_freq,
                  "Never" = 1,
```

```
                  "Less than once a month" = 2,
                  "A few times a month" = 3,
                  "At least once a week" = 4,
                  "Almost every day" = 5,
                  "Missing data (Not Ascertained)" = NA_real_,
                  "Missing data (Web partial - Question Never Seen)" = NA_real_,
                  "Multiple responses selected in error" = NA_real_),
        sm_watching = recode(sm_watching,
                     "Missing data (Not Ascertained)" = NA_real_,
                     "Missing data (Web partial - Question Never Seen)" = NA_real_,
                     "Never" = 1,
                     "Less than once a month" = 2,
                     "A few times a month" = 3,
                     "At least once a week" = 4,
                     "Almost every day" = 5),
        sm_interact = recode(sm_interact,
                     "Missing data (Not Ascertained)" = NA_real_,
                     "Missing data (Web partial - Question Never Seen)" = NA_real_,
                     "Never" = 1,
                     "Less than once a month" = 2,
                     "A few times a month" = 3,
                     "At least once a week" = 4,
                     "Almost every day" = 5),
        sm_sharing = recode(sm_sharing,
                    "Missing data (Not Ascertained)" = NA_real_,
                    "Missing data (Web partial - Question Never Seen)" = NA_real_,
                    "Multiple responses selected in error" = NA_real_,
                    "Never" = 1,
                    "Less than once a month" = 2,
                    "A few times a month" = 3,
                    "At least once a week" = 4,
                    "Almost every day" = 5)) %>%
  filter(!is.na(sm_usage_freq)) %>%
  filter(!is.na(sm_watching)) %>%
  filter(!is.na(sm_interact)) %>%
  filter(!is.na(sm_sharing)) %>%
  filter(telehealth !="Missing")
```

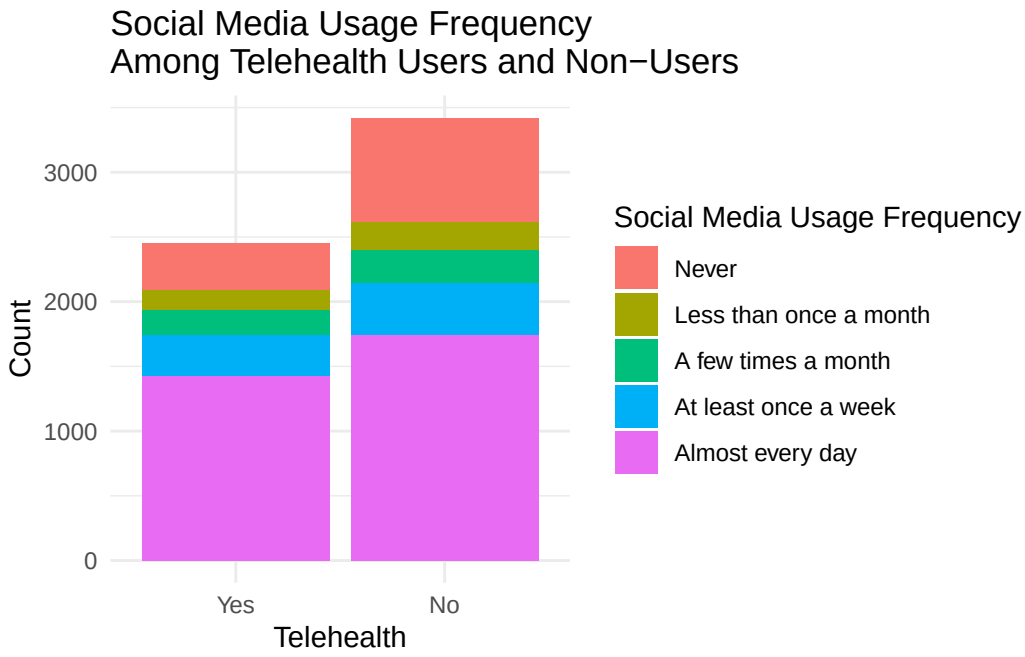
RQ2_2. Grouping by Health Information Engagement Levels on Social Media: pivot_longer()

```
# Reshape health-information-engagement-on-social-media categories into one longer engagement

data_long <- data_nl_2 %>%
  pivot_longer(cols = c(sm_watching, sm_interact, sm_sharing),
               names_to = "engagement_type", values_to = "frequency")

# Low, medium, high grouping

data_long <- data_long %>%
  group_by(HHID) %>%
  mutate(total_sm_health_eng_lev = sum(frequency))

data_freq_eng_group <- data_long %>%
  distinct(HHID, telehealth, sm_usage_freq, total_sm_health_eng_lev) %>%
  mutate(sm_health_eng_lev_group =
           case_when(
             total_sm_health_eng_lev >= 3 &
               total_sm_health_eng_lev <= 6 ~ "Low",
             total_sm_health_eng_lev >= 7 &
               total_sm_health_eng_lev <= 10 ~ "Medium",
             total_sm_health_eng_lev >= 11 &
               total_sm_health_eng_lev <= 15 ~ "High"))
```

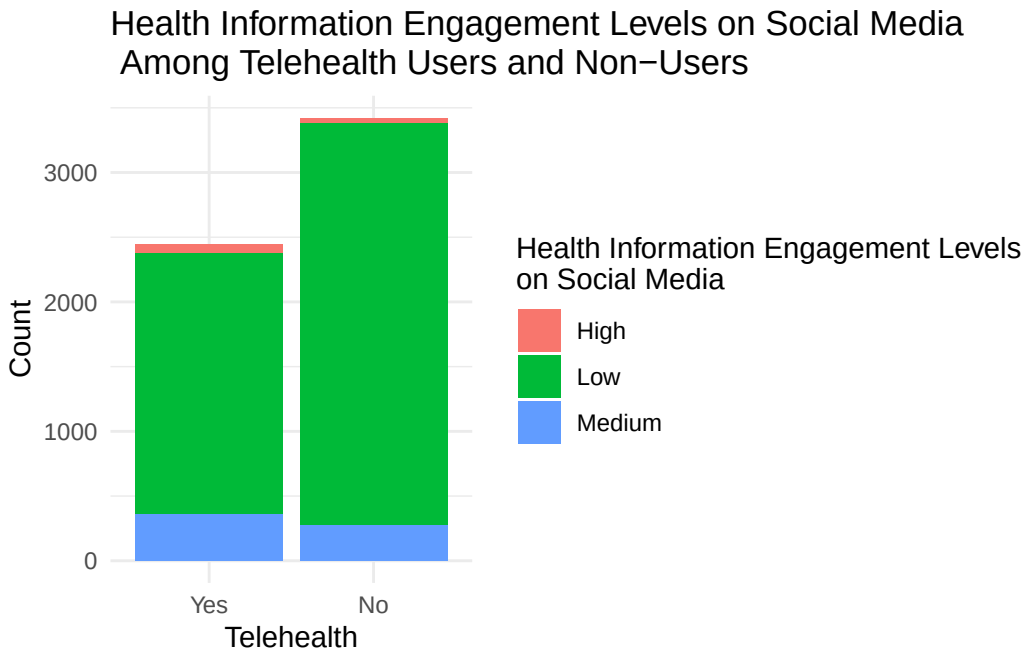
RQ2_3. Data Visualization

```
# Social Media Frequency x Telehealth

ggplot(data_freq_eng_group, aes(x = telehealth, fill = as.factor(sm_usage_freq)))+
         geom_bar() +
  scale_fill_discrete(
    labels = c("Never", "Less than once a month",
               "A few times a month", "At least once a week",
               "Almost every day")) +
         theme_minimal() +
  labs(
    title = "Social Media Usage Frequency\nAmong Telehealth Users and Non-Users",
```

```
       x = "Telehealth", y = "Count",
       fill = "Social Media Usage Frequency")
```




```
# Health Information Engagement Levels on Social Media X Telehealth

ggplot(data_freq_eng_group, aes(x = telehealth, fill = sm_health_eng_lev_group)) +
  geom_bar() +
  theme_minimal() +
  labs(
    title = "Health Information Engagement Levels on Social Media\n Among Telehealth Users a
    x = "Telehealth", y = "Count",
    fill = "Health Information Engagement Levels\non Social Media")
```



RQ2_4. Summary Statistics

```
# Social Media Usage Frequency X Telehealth

summary_table_sm_freq <- data_freq_eng_group %>%
  mutate(sm_usage_freq = case_when(
      sm_usage_freq == 1 ~ "1 (Never)",
      sm_usage_freq == 2 ~ "2 (Less than once a month)",
      sm_usage_freq == 3 ~ "3 (A few times a month)",
      sm_usage_freq == 4 ~ "4 (At least once a week)",
      sm_usage_freq == 5 ~ "5 (Almost every day)")) %>%
  group_by(telehealth, sm_usage_freq) %>%
  summarize(Count = n()) %>%
  mutate(Percentage = round(Count / sum(Count) * 100, 2)) %>%
  rename(Telehealth = telehealth, SM_Usage_Freqency = sm_usage_freq)

kable(summary_table_sm_freq)
```

Telehealth	SM_Usage_Freqency	Count	Percentage
Yes	1 (Never)	358	14.62
Yes	2 (Less than once a month)	152	6.21
Yes	3 (A few times a month)	200	8.17
Yes	4 (At least once a week)	314	12.83
Yes	5 (Almost every day)	1424	58.17
No	1 (Never)	811	23.70
No	2 (Less than once a month)	214	6.25
No	3 (A few times a month)	254	7.42
No	4 (At least once a week)	398	11.63
No	5 (Almost every day)	1745	50.99

```r
# Mean, Median, SD for Social Media Usage Frequency X Telehealth

summary_table_sm_freq_stats <- data_freq_eng_group %>%
  group_by(telehealth) %>%
  summarize(SM_Usage_Freq_Mean = round(mean(sm_usage_freq), 2),
            Median = median(sm_usage_freq),
            SD = round(sd(sm_usage_freq), 2),
            Count = n()) %>%
  rename(Telehealth = telehealth)

kable(summary_table_sm_freq_stats)
```

Telehealth	SM_Usage_Freq_Mean	Median	SD	Count
Yes	3.94	5	1.49	2448
No	3.60	5	1.68	3422

```r
# Health Information Engagement Levels on Social Media X Telehealth

summary_table_eng_lev_group <- data_freq_eng_group %>%
  group_by(telehealth, sm_health_eng_lev_group) %>%
  summarize(Count = n()) %>%
  mutate(Percentage = round(Count / sum(Count) * 100, 2)) %>%
  rename(
    Telehealth = telehealth,
    SM_Health_Info_Eng = sm_health_eng_lev_group)

kable(summary_table_eng_lev_group)
```

Telehealth	SM_Health_Info_Eng	Count	Percentage
Yes	High	69	2.82
Yes	Low	2020	82.52
Yes	Medium	359	14.67
No	High	38	1.11
No	Low	3106	90.77
No	Medium	278	8.12

```
# Mean, Median, SD for Health Information Engagement Levels on Social Media X Telehealth

summary_table_eng_lev_stats <- data_long %>%
  group_by(telehealth) %>%
  summarize(SM_Health_Info_Eng_Mean = round(mean(total_sm_health_eng_lev), 2),
            Median = median(total_sm_health_eng_lev),
            SD = round(sd(total_sm_health_eng_lev), 2)) %>%
  rename(Telehealth = telehealth)

kable(summary_table_eng_lev_stats)
```

Telehealth	SM_Health_Info_Eng_Mean	Median	SD
Yes	4.90	4	2.16
No	4.26	4	1.71

RQ3. How does the distribution of cancer-related risk perception groups (low, medium, high) compare between individuals who use and do not use Telehealth services?

RQ3_1. Data Cleaning

```
# Selecting relevant columns for RQ3
cancer <- hints %>%
  select(HHID, ReceiveTelehealthCare, InterestedCaScreening, FreqWorryCancer, ChanceGetCancer

# Renaming each column
cancer <- cancer %>%
  rename(telehealth = ReceiveTelehealthCare,
```

```
          screening = InterestedCaScreening,
          worry = FreqWorryCancer,
          chance = ChanceGetCancer2,
          cause = EverythingCauseCancer,
          prevent = PreventNotPossible)

# Recoding responses
cancer <- cancer %>%
  mutate(telehealth =
           recode(telehealth,
                  "No telehealth visits in the past 12 months" = "No",
                  "Yes, by video" = "Yes",
                  "Yes, by phone call (voice only with no video)" = "Yes",
                  "Yes, some by video and some by phone call" = "Yes",
                  "Missing data (Web partial - Question Never Seen)" = "Missing",
                  "Missing data (Not Ascertained)" = "Missing"))

# Creating a dataset for 5 point Likert scale
cancer_2 <- cancer %>%
  mutate(screening =
           recode(screening,
                  "Missing data (Not Ascertained)" = NA_real_,
                  "Missing data (Web partial - Question Never Seen)" = NA_real_,
                  "Multiple responses selected in error" = NA_real_,
                  "Not at all" = 1,
                  "A little" = 2,
                  "Somewhat" = 3,
                  "Very" = 4,
                  "Not applicable/I am up-to-date with screening tests" = 5),
         worry =
           recode(worry,
                  "Missing data (Not Ascertained)" = NA_real_,
                  "Missing data (Web partial - Question Never Seen)" = NA_real_,
                  "Multiple responses selected in error" = NA_real_,
                  "Not at all" = 1,
                  "Slightly" = 2,
                  "Somewhat" = 3,
                  "Moderately" = 4,
                  "Extremely" = 5),
         chance =
           recode(chance,
                  "Missing data (Not Ascertained)" = NA_real_,
```

```
                  "Missing data (Web partial - Question Never Seen)" = NA_real_,
                  "Multiple responses selected in error" = NA_real_,
                  "I already had cancer" = NA_real_,
                  "I don't know" = NA_real_,
                  "Very unlikely" = 1,
                  "Unlikely" = 2,
                  "Neither likely nor unlikely" = 3,
                  "Likely" = 4,
                  "Very likely" = 5),
         cause =
           recode(cause,
                  "Missing data (Not Ascertained)" = NA_real_,
                  "Missing data (Web partial - Question Never Seen)" = NA_real_,
                  "Multiple responses selected in error" = NA_real_,
                  "Strongly disagree" = 1,
                  "Somewhat disagree" = 2,
                  "Somewhat agree" = 4,
                  "Strongly agree" = 5),
         prevent =
           recode(prevent,
                  "Missing data (Not Ascertained)" = NA_real_,
                  "Missing data (Web partial - Question Never Seen)" = NA_real_,
                  "Multiple responses selected in error" = NA_real_,
                  "Strongly disagree" = 1,
                  "Somewhat disagree" = 2,
                  "Somewhat agree" = 4,
                  "Strongly agree" = 5)) %>%
  filter(!is.na(screening)) %>%
  filter(!is.na(worry)) %>%
  filter(!is.na(chance)) %>%
  filter(!is.na(cause)) %>%
  filter(!is.na(prevent)) %>%
  filter(telehealth !="Missing")
```

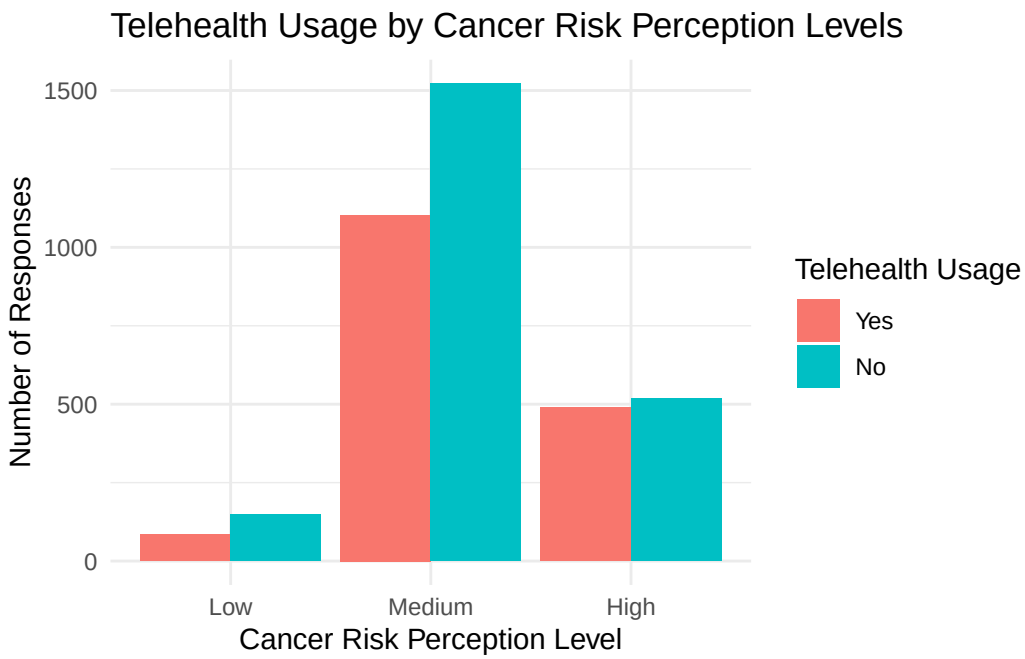
RQ3_2. Grouping by cancer risk perception groups

```
# Reshape cancer risk perceptions into one longer column
cancer_long <- cancer_2 %>%
  pivot_longer(
    cols = c(screening, worry, chance, cause, prevent),
```

```
    names_to = "riskperceptiontype",
    values_to = "riskperceptiondegree"
  )

# Low, medium, high grouping
cancer_long <- cancer_long %>%
  group_by(HHID) %>%
  mutate(riskperceptionlevel = mean(riskperceptiondegree, na.rm = TRUE))

# Keeping only one risk perception degree value per household id
cancer_risk_group <- cancer_long %>%
  distinct(HHID, telehealth, riskperceptionlevel) %>%
  mutate(risk_group =
           case_when(
             riskperceptionlevel <= 1.67 ~ "Low",
             riskperceptionlevel <= 3.33 ~ "Medium",
             riskperceptionlevel <= 5 ~ "High"))
```

RQ3_3. Data Visualization

```
# Cancer Risk Levels X Telehealth
cancer_risk_group %>%
  mutate(risk_group = factor(risk_group, levels = c("Low", "Medium", "High"))) %>%
  ggplot(aes(x = risk_group, fill = telehealth)) +
  geom_bar(position = "dodge") +
  theme_minimal() +
  labs(title = "Telehealth Usage by Cancer Risk Perception Levels",
       x = "Cancer Risk Perception Level",
       y = "Number of Responses",
       fill = "Telehealth Usage")
```



RQ3_4. Summary Statistics

```
# Cancer Risk Levels X Telehealth
summary_table_cancer_risk_group <- cancer_risk_group %>%
  group_by(telehealth, risk_group) %>%
  summarize(count = n()) %>%
  mutate(percentage = round(count/sum(count)*100, 2))
kable(summary_table_cancer_risk_group)
```

telehealth	risk_group	count	percentage
Yes	High	489	29.18
Yes	Low	84	5.01
Yes	Medium	1103	65.81
No	High	518	23.67
No	Low	148	6.76
No	Medium	1522	69.56

```
# Mean, Median, SD for SME X Telehealth
summary_table_cancer_risk_stats <- cancer_long %>%
```

```
  group_by(telehealth) %>%
  summarize(riskperception_mean = round(mean(riskperceptiondegree), 2),
            riskperception_median = median(riskperceptiondegree),
            riskperception_sd = round(sd(riskperceptiondegree), 2))
kable(summary_table_cancer_risk_stats)
```

telehealth	riskperception_mean	riskperception_median	riskperception_sd
Yes	2.90	3	1.29
No	2.79	3	1.30