A Space Odyssey

Fernando Finotto Visani

7/16/2020

What's the proportion of Sex and Occupation?
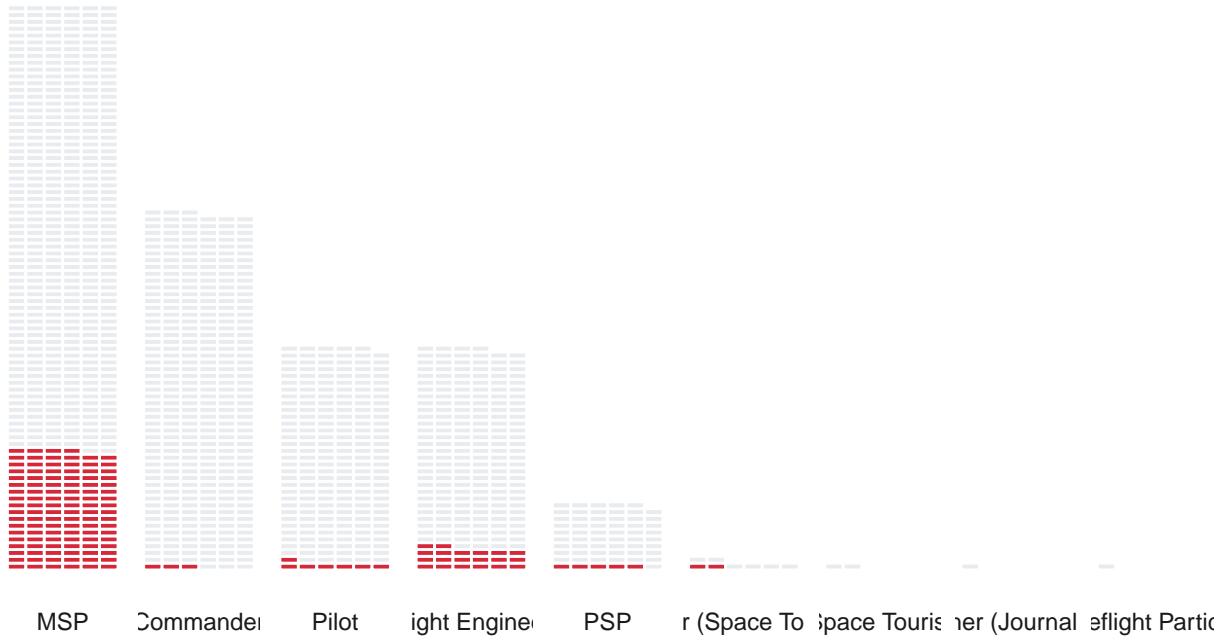
```r
prop_astronauts <- astronauts %>%
  mutate(occupation = str_to_title(occupation)) %>%
  count(sex, occupation, sort = TRUE) %>%
  group_by(sex) %>%
  mutate(prop = n / sum(n),
         occupation = case_when(
            str_detect(occupation,"Msp") ~ "MSP",
            str_detect(occupation, "Psp") ~ "PSP",
            TRUE ~ occupation)) %>%
  arrange(sex) %>%
  mutate(sex = str_to_title(sex))

occupation_rank <- prop_astronauts %>%
  group_by(occupation) %>%
  summarise(occupation_rank = sum(n)) %>%
  arrange(desc(occupation_rank))

prop_astronauts %>%
  left_join(occupation_rank, by = "occupation") %>%
  mutate(occupation = fct_reorder(occupation, occupation_rank, .desc = TRUE)) %>%
  ggplot(aes(fill = sex, values = n)) +
  geom_waffle(color = "white", size = .5, n_rows = 6, flip = TRUE) +
  facet_wrap(~occupation, nrow = 1, strip.position = "bottom") +
  labs(title = "Space mission jobs by <span style='color:#D62839;'>**female**</span> astronauts",
       subtitle = "74% of the female astronauts take part in the The Maximizing Student Potential (MSP)
  theme(panel.grid = element_blank(),
        axis.text = element_blank(),
        legend.title = element_blank(),
        legend.position = "none",
        legend.justification = c(1, 1),
        plot.title = element_markdown(family = "sans",
                                      hjust = .5,
                                      size = 20),
        plot.subtitle = element_text(family = "sans",
                                      hjust = .5,
                                      size = 15)) +
  scale_fill_manual(values = c("#D62839", "#E9ECEF"))
```

Space mission jobs by **female** astronauts

emale astronauts take part in the The Maximizing Student Potential (MS
t Engineer – in second place – represents only 14% of the female popu

| MSP | Commander | Pilot | ight Engine | PSP | r (Space To | Space Touris | her (Journal | eflight Partic |

What's the age proportion for male/female astronauts?

```r
age_mission <- astronauts %>%
  filter(mission_number == 1) %>%
  group_by(sex, year_of_birth, year_of_selection, year_of_mission) %>%
  transmute(age_selection = year_of_selection - year_of_birth,
            age_mission = year_of_mission - year_of_birth)

age_mission %>%
  ungroup() %>%
  group_by(sex) %>%
  summarise(mean_age_selection = mean(age_selection, na.rm = TRUE),
            mean_age_mission = mean(age_mission, na.rm = TRUE),
            diff_age = mean_age_mission - mean_age_selection)
```

```
## # A tibble: 2 x 4
##   sex    mean_age_selection mean_age_mission diff_age
##   <chr>              <dbl>            <dbl>    <dbl>
## 1 female              32.5             38.1     5.62
## 2 male                34.6             40.6     6.03
```

```r
age_mission %>%
  ggplot(aes(age_selection, age_mission)) +
  geom_jitter(color = "#D62839",
              size = 3,
              alpha = 0.75,
              position = position_jitter(seed = 1)) +
  gghighlight(sex == "female",
              unhighlighted_params = list(color = "#E9ECEF")) +
  labs(title = "How long it takes for <span style='color:#D62839;'>**female**</span> astronauts to be i
       subtitle = "Female astronauts take, in average, *5,6 years* to be in space for their first missi
       x = "Age of Selection",
       y = "Age of Mission") +
  theme(panel.grid = element_blank(),
        legend.position = "none",
        plot.title = element_markdown(family = "sans",
                                      hjust = .05,
                                      size = 20),
        plot.subtitle = element_markdown(family = "sans",
                                         hjust = .08,
                                         size = 15)) +
  annotate("curve", x = 50, xend = 47.2, y = 62, yend = 56.2, curvature = 0.4,
           size = .75, arrow = arrow(length = unit(2, "mm")), color = "#938ca1", na.rm = TRUE) +
  geom_textbox(aes(x = 52.5, y = 62, label = "**Morgan, Barbara R.** was the oldest female astronaut at
  annotate("curve", x = 35, xend = 28.7, y = 30, yend = 39.7, curvature = -0.4,
           size = .75, arrow = arrow(length = unit(2, "mm")), color = "#938ca1", na.rm = TRUE) +
  geom_textbox(aes(x = 36.7, y = 30, label = "**Yamazaki, Naoko** took the longest to be in space - 11 y
```

How long it takes for **female** astronauts to be in sp

Female astronauts take, in average, *5,6 years* to be in space for their f

the oldest female astronaut at the age of 56

Yamazaki, Naoko took the longest to be in space – 11 years!

Age of Mission

60

50

40

30

Age of Selection

30   40   50   60
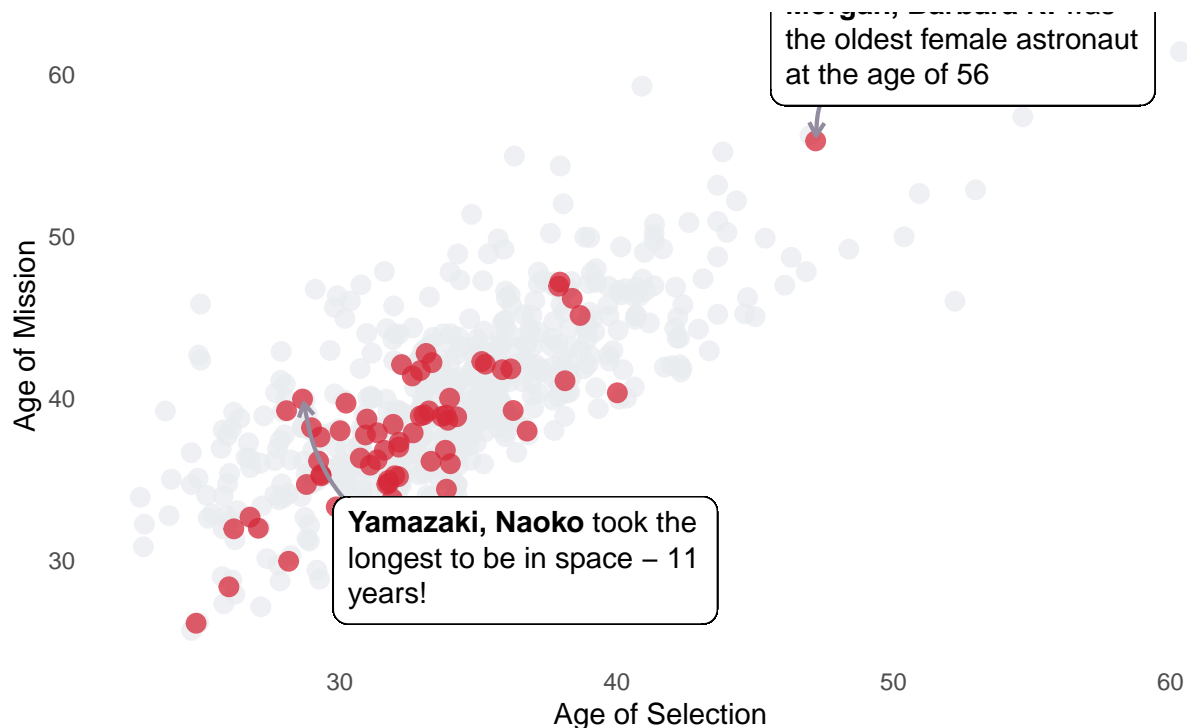
```r
theme_finotto <- function() {
  theme_minimal() +
    theme(panel.grid = element_blank(),
          legend.position = "none",
          plot.background = element_rect(color = "#212529",
                                         fill = "#212529"),
          axis.text = element_text(color = "#F8F9FA",
                                   size = 13),
          axis.title = element_text(color = "#F8F9FA",
                                    size = 13),
          strip.text.x = element_text(color = "#F8F9FA",
                                      size = 13),
          plot.title = element_markdown(family = "sans",
                                        hjust = .5,
                                        size = 20,
                                        color = "#F8F9FA"),
          plot.subtitle = element_markdown(family = "sans",
                                           hjust = .5,
                                           size = 17,
                                           color = "#F8F9FA"))
}

waiting_for_misson <- astronauts %>%
  select(number, sex, mission_number, year_of_birth, year_of_selection, year_of_mission) %>%
  group_by(number, sex, mission_number) %>%
  arrange(number) %>%
```

```r
    mutate(age_selection = year_of_selection - year_of_birth,
           age_mission = year_of_mission - year_of_birth,
           selection_to_mission = year_of_mission - year_of_selection,
           sex = str_to_sentence(sex)) %>%
    group_by(number) %>%
    mutate(second_mission = selection_to_mission - lag(selection_to_mission),
           waiting_for_misson = coalesce(selection_to_mission, second_mission)) %>%
    filter(waiting_for_misson > 0) %>%
    select(-c(selection_to_mission, second_mission)) %>%
    ungroup() %>%
    group_by(sex) %>%
    mutate(waiting_mean = mean(waiting_for_misson, na.rm = TRUE)) %>%
    ungroup()

annotation <- tibble(mission_number = c(7, 7),
                     waiting_for_misson = c(9, 9.8),
                     sex = c("Female", "Male"),
                     label = c("Mean: 8.48", "Mean: 9.28"))

ggplot(waiting_for_misson, aes(mission_number, waiting_for_misson)) +
  # geom_jitter(position = position_jitter(seed = 2), alpha = .2) +
  geom_boxplot(aes(color = factor(sex), group = factor(mission_number)), alpha = .75, fill = NA) +
  # geom_hline(aes(yintercept = waiting_mean),
  #            linetype = "dashed",
  #            color = "#F8F9FA",
  #            size = 1) +
  facet_wrap(~sex) +
  scale_x_continuous(breaks = seq(0, 7, by = 1)) +
  scale_y_continuous(breaks = seq(0, 40, by = 5)) +
  labs(title = "How long astronauts wait for a mission after the previous one?",
       x = "Mission Number",
       y = "Waiting Years for a Mission") +
  scale_color_manual(values = c("#D62839", "#98C1D9")) +
  # geom_text(data = annotation, aes(x = mission_number,
  #                                  y = waiting_for_misson,
  #                                  label = label),
  #           color = "#F8F9FA",
  #           size = 5) +
  theme_finotto()
```
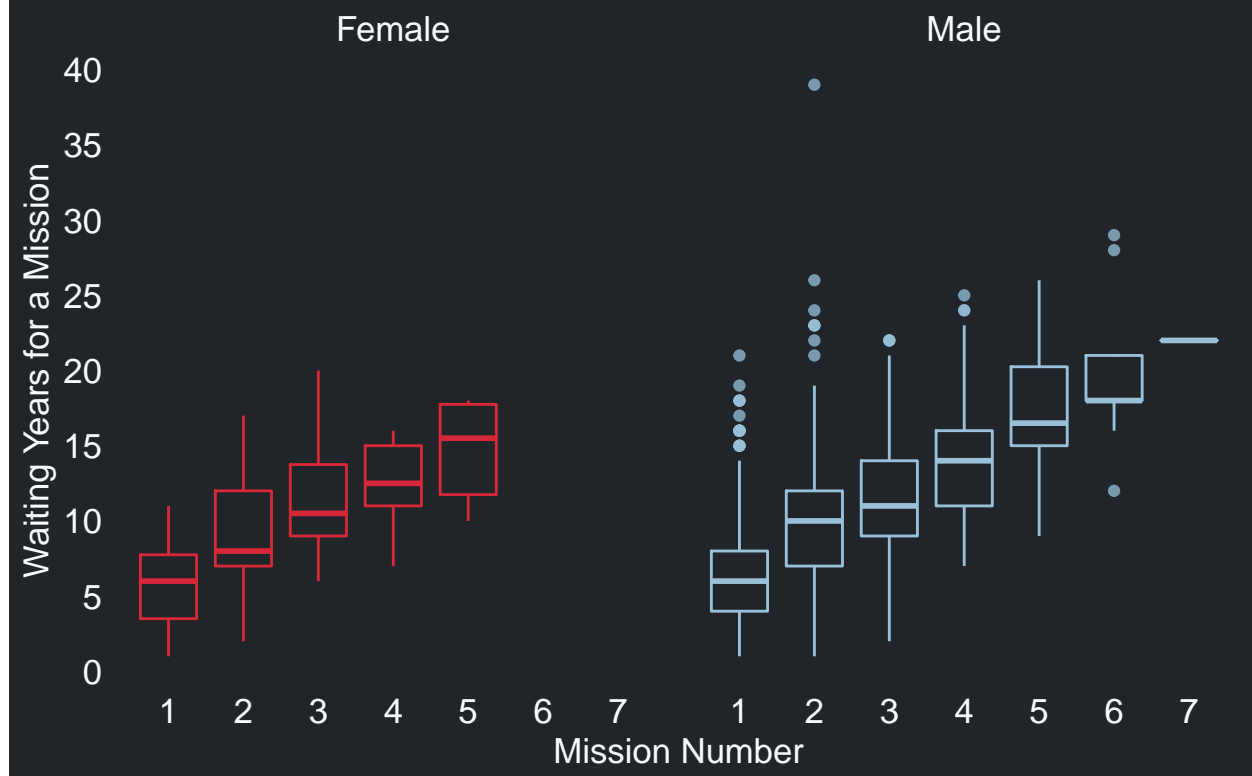
```
waiting_for_misson %>%
  # mutate(sex = ifelse(sex == "Male", 1, 2)) %>%
  ggplot(aes(factor(mission_number),
             waiting_for_misson,
             color = factor(mission_number),
             fill = sex)) +
  geom_violin(color = "#F8F9FA",
              alpha = .5,
              draw_quantiles = 0.5) +
  scale_fill_manual(values = c("#D62839", "#98C1D9")) +
  labs(title = "How long astronauts wait for a mission after the previous one?",
       x = "Mission Number",
       y = "Waiting Years for a Mission") +
  theme_finotto()
```
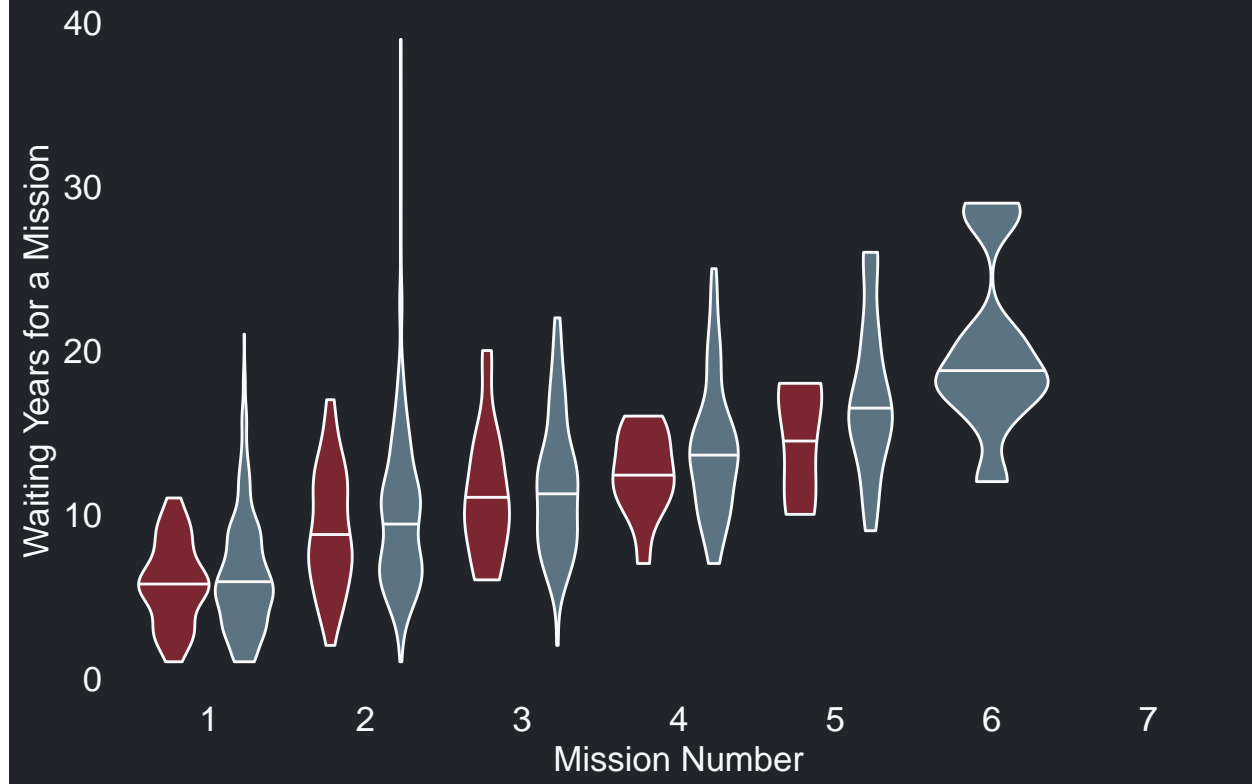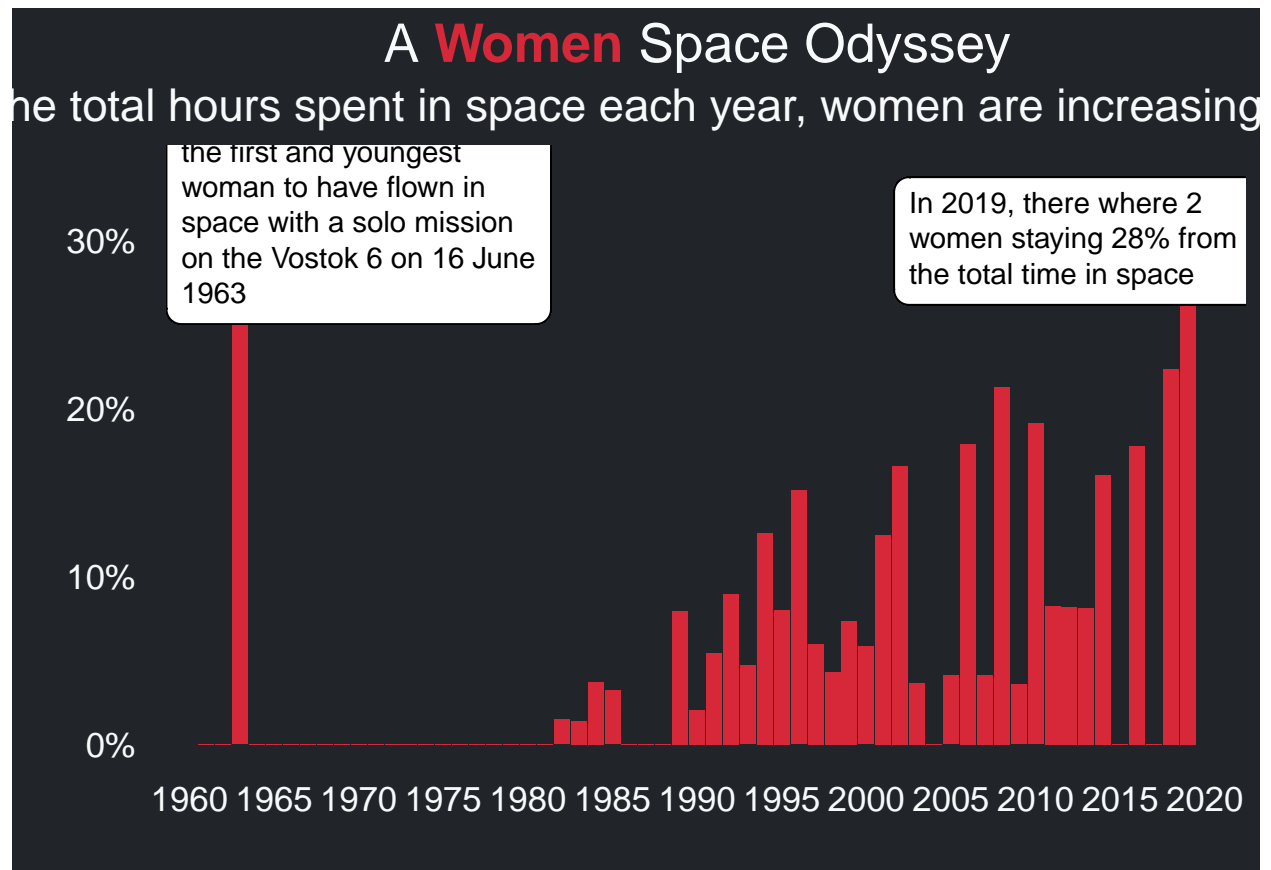
```
mission_time <- astronauts %>%
  group_by(year_of_mission, sex) %>%
  summarise(hours_mission = sum(hours_mission, na.rm = TRUE)) %>%
  group_by(year_of_mission) %>%
  mutate(pct_sex = hours_mission / sum(hours_mission),
         pct_sex = ifelse(sex == "male", 0, pct_sex))

mission_time %>%
  ggplot(aes(year_of_mission, pct_sex)) +
  geom_bar(stat = "identity", fill = "#D62839") +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1),
                     breaks = seq(0, 0.4, by = 0.1)) +
  scale_x_continuous(breaks = seq(1960, 2020, by = 5)) +
  labs(title = "A <span style='color:#D62839;'>**Women**</span> Space Odyssey",
       subtitle = "From the total hours spent in space each year, women are increasing their staying",
       x = "",
       y = "") +
  annotate("curve", x = 1966, xend = 1963, y = 0.34, yend = 0.32, curvature = 0.4,
           size = .75, arrow = arrow(length = unit(2, "mm")), color = "#F8F9FA", na.rm = TRUE) +
  geom_textbox(aes(x = 1970, y = 0.32, label = "**Valentina Tereshkova** is the first and youngest woma
  annotate("curve", x = 2017, xend = 2019, y = 0.3, yend = 0.285, curvature = -0.4,
           size = .75, arrow = arrow(length = unit(2, "mm")), color = "#F8F9FA", na.rm = TRUE) +
  geom_textbox(aes(x = 2013, y = 0.3, label = "In 2019, there where 2 women staying 28% from the total
  theme_finotto()
```

A **Women** Space Odyssey

he total hours spent in space each year, women are increasing

the first and youngest woman to have flown in space with a solo mission on the Vostok 6 on 16 June 1963

In 2019, there where 2 women staying 28% from the total time in space

```r
# ggsave("2020_29_women_odyssey.jpg", width = 40, height = 20, units = "cm")
```

```r
astronauts %>%
  filter(nationality %in% c("U.S.", "U.S.S.R/Russia")) %>%
  group_by(year_of_mission, nationality) %>%
  summarise(hours_mission = sum(hours_mission, na.rm = TRUE)) %>%
  mutate(hours_mission = ifelse(nationality == "U.S.S.R/Russia", hours_mission * -1, hours_mission)) %>%
  ggplot(aes(year_of_mission, hours_mission, fill = nationality, color = nationality)) +
  geom_area(alpha = 0.75) +
  scale_x_continuous(breaks = seq(1960, 2020, by = 5)) +
  scale_y_continuous(labels = scales::number_format()) +
  labs(title = "<span style='color:#E63946;'>**USA**</span> and <span style='color:#457B9D;'>**Russia**
      x = "",
      y = "Total Mission Hours") +
  scale_fill_manual(values = c("#E63946", "#457B9D")) +
  scale_color_manual(values = c("#E63946", "#457B9D")) +
  theme_finotto()
```

USA and Russia are still in the Space Race