Rich Ramsey | 2022/02/18 | ST 540 | Midterm 1

Introduction

To examine whether a host country advantage exists for the Summer Olympics, I examined data from host years 1952 to 2021 with the number of participants and medals won for the host country in their host year and in the previous Summer Olympics.

		Previous Olympics		Host Year	
	Year	Medals Won	Participating Athletes	Medals Won	Participating Athletes
Finland	1952	24	129	22	258
Australia	1956	11	81	35	294
Italy	1960	25	135	36	280
Japan	1964	18	162	29	328
Mexico	1968	1	94	9	275
West Germany	1972	26	275	40	423
Canada	1976	5	208	11	385
Soviet Union	1980	125	410	195	489
United States	1984	94	396	174	522
South Korea	1988	19	175	33	401
Spain	1992	4	229	22	422
United States	1996	108	545	101	647
Australia	2000	41	417	58	617
Greece	2004	13	140	16	426
China	2008	63	384	100	599
Great Britain	2012	47	304	65	530
Brazil	2016	17	236	19	462
Japan	2021	41	395	51	621
Aggregate		682	4,715	1,016	7,979

Part I: Aggregate Analysis

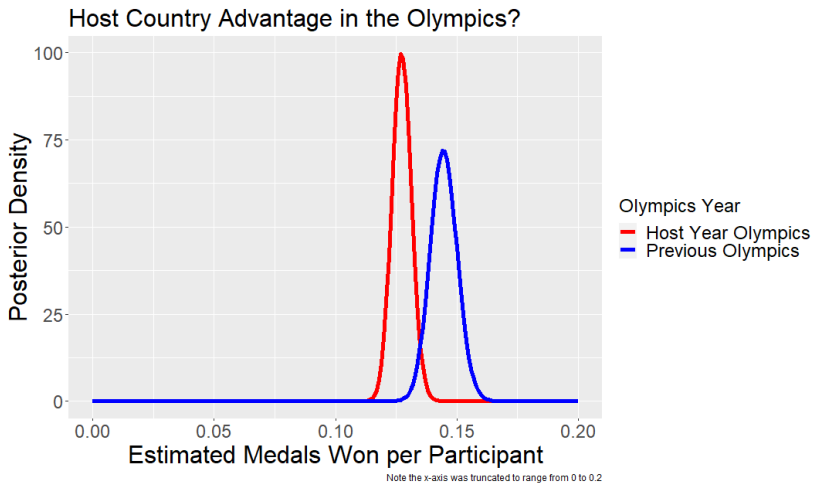
Initially, I looked at the participants and medals won data combined across all the host countries and years to look for evidence of a host country advantage over all the countries.

Year	**Y**	**N**
Previous	682	4715
Host	1016	7979

The number of medals won is a discrete number of events, and the number of medals won per participant from the country can be seen as a rate. I used a Poisson distribution as the likelihood for the previous Olympics (rate λ_0) and the host year Olympics (rate λ_1) with Y as the number of medals won and N as number of participants. I am choosing to use this distribution because an athlete can participate in multiple events (and win multiple medals), so I think it's a reasonable approach to model it as the rate of medals won per participant.

This will have a likelihood modeled as: $Y \mid \lambda \sim Poisson(N * \lambda)$

The uninformative conjugate prior distribution for both will be Gamma(a = 0.1, b = 0.1). This yields a posterior distribution of Gamma(A = Y + a, B = N + b) based on the likelihood and prior.



One of the assumptions here is that there is not a significant difference in host country advantage by country, given that all the countries are combined for this section. It's illustrated in more detail in part IV, but I think this is fairly reasonable assumption. Another assumption is that host country advantage has not changed over time. I'm less sure of this one, but have not thought of a compelling reason why it would have changed noticeably over time. The final big assumption is that the pools of athletes competing for a given country in the previous Olympics and host year Olympics are of similar quality – this assumption may not hold.

Part II: Hypothesis Test

The estimated posterior distributions above seem to indicate that there is not a host country advantage. To test the hypothesis that there is a host country advantage ($\lambda_1 > \lambda_0$), I used Monte Carlo sampling to generate 100,000 samples from the posterior distribution of host year and previous Olympics estimated λ. The percentage of those where λ_1 was greater than λ_0 was 0.5%. This does not offer enough evidence to reject the null hypothesis that λ, the medals won per participant, is equal in the host year and the previous year. In fact, this would seem to indicate that the medals won per participant in the previous Olympics is higher than that of the host year, given 99.5% of the Monte Carlo samples had a higher estimated λ_0 (medals won per participant rate for the previous Olympics).

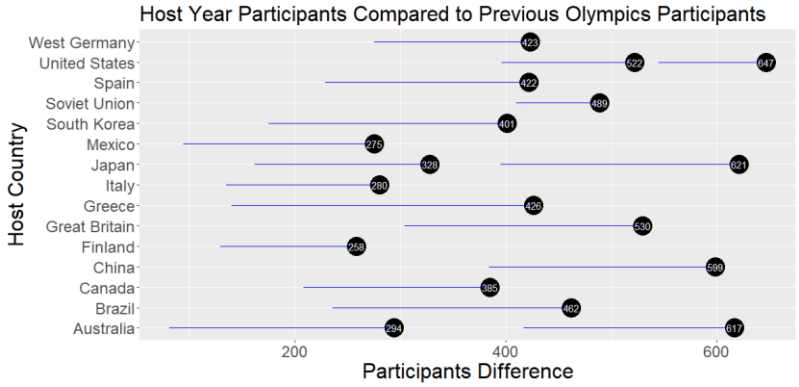
Prior	**% with estimated λ_1 > estimated λ_0**
Gamma(0.1, 0.1)	0.52%
Gamma(0.5, 0.5)	0.51%
Gamma(1, 1)	0.51%
Gamma(2, 2)	0.52%
Gamma(3, 3)	0.49%
Gamma(4, 4)	0.49%
Gamma(5, 5)	0.41%
Gamma(10, 10)	0.42%

There are some slight changes based on the prior, but the result is the same – I don't see evidence of a host country advantage as measured by estimated ($\lambda_1 > \lambda_0$).

Part III: Prediction

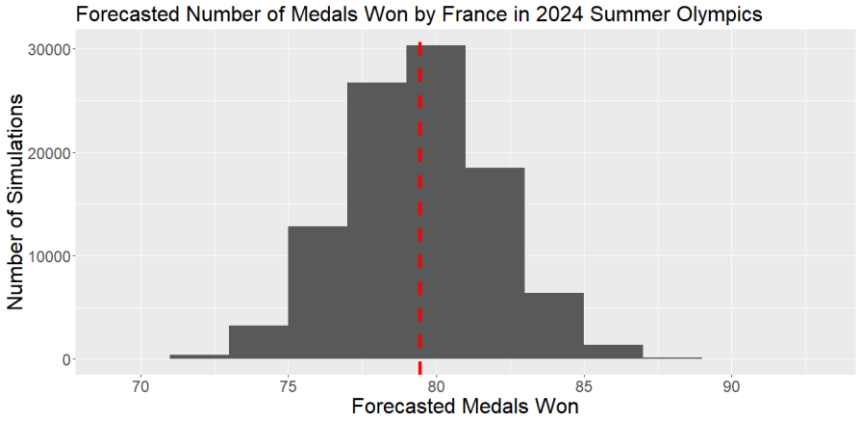
The next host of the Summer Olympics will be France in 2024. Based on the aggregate estimations of host country advantage and forecasted number of athletes competing for France in 2024, we will predict the number of medals France will win in the 2024 Olympics. First, we need to predict the number of athletes from France will participate in the next Summer Olympics. In the 2021 (originally 2020) Summer Olympics, France had 398 participants. However, it's important to

note that the host countries are guaranteed spots in each of the team sports and athletes from the host country have easier roads to qualification for the individual sports[1]. Below, I've shown the differences in host year participants compared to previous Olympics participants for the years in our data set.



The average ratio of host year participants to previous year participants in this data set is 1.957 (rounded). However, not all host countries are equal in this respect – some send far more participants to the Olympics each time than others. France is traditionally one of the more well-represented countries at the Olympics, so I think limiting the sample to similarly well-represented countries will provide a slightly more accurate estimate. Looking only at host countries with at least 200 participants in the previous Olympics, the average ratio is 1.568. Using that, I arrived at an estimate of 624 participants (rounded to nearest whole number) for France in 2024.

Based on that number and the posterior distribution for medals per participant rate during the host year, I can estimate the number of medals France will win in 2024. To arrive at that estimate, I generated a random number from that posterior distribution 100,000 times and applied that rate to the 624 estimated participants.



Based on these assumptions and samples from the host year posterior distribution, I estimate that France will win a mean of 79.3 medals in 2024. The 2.5th percentile of the estimated medals won is 74.6, while the 97.5th percentile is 84.6. Rounding those with more assumed uncertainty, I will make that estimated range from 74 to 85.

Part IV: Country-specific Analysis

Perhaps some countries have had stronger home country advantage than others. To investigate this, I adapted the Poisson-Gamma model to estimate the posterior distribution of host year and previous year medals per participant rate for each country separately. The likelihood for each country uses Y_i medals won and N_i participants for the i^{th} country. I maintained the uninformative conjugate prior distribution of Gamma(a = 0.1, b = 0.1). This leads to a posterior distribution of Gamma(a + Y_i, b + N_i) for each country. Using those posterior distributions, I took 100,000 samples from the host year posterior and previous Olympics posterior for each country. I took the results of those samples and calculated the percentage of those

with estimated host year rates greater than estimated previous Olympics rates, as well as calculating the ratio of estimated host year rate to estimated previous Olympics rate.

	Host Year Olympics		Previous Olympics		Comparison - Monte Carlo Samples From Country-Specific Posteriors	
	2.5th Percentile	97.5th Percentile	2.5th Percentile	97.5th Percentile	% Samples Host Year > Previous Olympics	Mean Ratio of Host Year to Previous Olympics
Australia	0.08	0.12	0.08	0.13	45.5%	1.00
Brazil	0.02	0.06	0.04	0.11	5.0%	0.61
Canada	0.01	0.05	0.01	0.05	64.2%	1.46
China	0.14	0.20	0.13	0.21	54.8%	1.03
Finland	0.05	0.12	0.12	0.27	0.4%	0.48
Great Britain	0.09	0.15	0.11	0.20	11.6%	0.81
Greece	0.02	0.06	0.05	0.15	1.0%	0.44
Italy	0.09	0.17	0.12	0.27	8.7%	0.72
Japan	0.07	0.10	0.08	0.13	9.7%	0.81
Mexico	0.02	0.06	0.00	0.04	91.9%	19.02
South Korea	0.06	0.11	0.07	0.16	17.8%	0.80
Soviet Union	0.34	0.46	0.25	0.36	99.1%	1.32
Spain	0.03	0.08	0.00	0.04	99.0%	3.88
United States	0.21	0.26	0.19	0.25	83.9%	1.10
West Germany	0.07	0.13	0.06	0.13	50.7%	1.04

For the most part, this approach does not lead me to believe there is evidence that host country advantage differs significantly by country. There are many confounding factors in this approach. The 1980 Olympics, hosted by the Soviet Union, were boycotted by many countries, including the United States. Similarly, the 1984 Olympics were hosted by the United States and boycotted by the Soviet Union. Those boycotts allowed for home country competitors to have a greater chance of winning medals against diluted competition. The two other host countries with some evidence of an advantage were Spain and Mexico. Mexico City, which hosted the Summer Olympics in 1968, is over 7,000 feet above sea level. The high elevation was a unique setting for the Summer Olympics (not the Winter Olympics) and may have led to an advantage for competitors from Mexico who were accustomed to that. I don't have as good of a hypothesis for why Spain would exhibit a strong host country advantage, but there is some evidence of that based on this analytical approach.

Part V: Conclusions

The main findings are that there is not a lot of evidence, based on this analytical approach, of a host country advantage in rate of medals won per participant during the host year compared to the previous year. This holds true even when analyzing each host country separately.

One area for further analysis is whether the host country advantage may be larger in certain Olympic events than others. The host country's fans may be particularly loud and exert a greater impact on certain events – basketball, for example, has a lot of fans in attendance that might provide a slight home court advantage. Some of the research done into home field/court advantage in other sports has indicated that referees are influenced slightly by the home fans – Olympic sports in which referees take more actions during the event may be worth separating.

Another area for further investigation would be to incorporate Olympic qualifying and data from other competitions for each sport to gain a more accurate forecast of the number of medals won in each event by the host country. As mentioned in previous sections, the host country has a lower barrier to entry for the events. This means that some of the athletes who qualify for the Olympics during the host year may not have qualified if they were in a different host country (I don't want to diminish their reaching the Olympics too much – it's an amazing accomplishment and must be extra special in their home country). If we only look at athletes competing for the host country who were likely to qualify even without the host country automatic entries, then the comparison of medals won per participant is more of an apples-to-apples comparison to the previous Olympics.

Sources

https://fivethirtyeight.com/features/is-there-home-field-advantage-at-the-olympics/

Code (appendix)

```
packages <- c("tidyverse", "gt", "xlsx", "ggimage", "countrycode")


install.packages(setdiff(packages, rownames(installed.packages())))


library(tidyverse)

library(gt)

library(xlsx)

library(ggimage)

library(countrycode)


setwd("C:/Users/richr/OneDrive/Documents/NC State/ST 540")


olympics_data <- read.csv(file = "Medals.csv")

colnames(olympics_data) <- tolower(colnames(olympics_data))


# add flags unicode

#olympics_data$flag_unicode <- countrycode(olympics_data$host.country, "country.name", "unicode.symbol")


##########################################################################################################
###################

# summarizing data

##########################################################################################################
###################


olympics_data %>%

  gt(rowname_col = "host.country") %>%

  fmt_number(columns = contains("year"), decimals = 0, use_seps = FALSE) %>%

  fmt_number(columns = contains(c("during")), decimals = 0) %>%

  grand_summary_rows(columns = contains("during"), fns = list(Aggregate = "sum"), missing_text = "", formatter =
fmt_number, decimals = 0) %>%
```

```
  cols_move(columns = c(medals.won.during.host.year, participating.athletes.during.host.year), after =
participating.athletes.during.previous.olympics) %>%
  tab_spanner(label = "Previous Olympics", columns = contains("previous")) %>%
  tab_spanner(label = "Host Year", columns = contains("host.year")) %>%
  cols_label(year = "Year",
             medals.won.during.previous.olympics = "Medals Won",
             medals.won.during.host.year = "Medals Won",
             participating.athletes.during.previous.olympics = "Participating Athletes",
             participating.athletes.during.host.year = "Participating Athletes") %>%
  # add outer border to table (gray)
  opt_table_outline(style = "solid", width = px(3)) %>%
  # add left and right borders to body
  tab_style(
    style = cell_borders(sides = c("right"), color = "#D3D3D3", weight = px(3), style = "solid"),
    locations = cells_body(columns = c(year, participating.athletes.during.previous.olympics)))

##########################################################################################################
###################

# part 1

##########################################################################################################
###################


set.seed(2024)


lambda_values <- seq(0, 1, by = 0.001)


# total medals won and participants for host countries in host year

y1 <- sum(olympics_data$medals.won.during.host.year)

n1 <- sum(olympics_data$participating.athletes.during.host.year)


# assume a likelihood based on number of medals won during host year and number of participants

likelihood1 <- (y1 / n1)
```

```
# use uninformative Gamma prior of a = b = 0.1
prior_a <- 0.1
prior_b <- prior_a

a1 <- (prior_a + y1)
b1 <- (prior_b + n1)

# total medals won and participants for host countries in previous year
y0 <- sum(olympics_data$medals.won.during.previous.olympics)
n0 <- sum(olympics_data$participating.athletes.during.previous.olympics)

a0 <- (prior_a + y0)
b0 <- (prior_b + n0)

# poisson-gamma model posterior
posterior_df <- data.frame(lambda_values, dgamma(lambda_values, shape = a1, rate = b1),
                           dgamma(lambda_values, shape = a0, rate = b0))
colnames(posterior_df) <- c("lambda", "host_year", "previous_year")

# plot both posterior distributions
posterior_df %>%
  pivot_longer(cols = c("host_year", "previous_year")) %>%
  ggplot(aes(x = lambda, y = value)) +
  geom_line(aes(color = name), size = 2) +
  scale_x_continuous(limits = c(0, 0.2)) +
  scale_color_manual(labels = c("Host Year Olympics", "Previous Olympics"), values = c("red", "blue")) +
  labs(x = "Estimated Medals Won per Participant",
       y = "Posterior Density",
       title = "Host Country Advantage in the Olympics?",
       caption = "Note the x-axis was truncated to range from 0 to 0.2",
```

```
      color = "Olympics Year") +
  theme(axis.title = element_text(size = 24),
        plot.title = element_text(size = 24),
        axis.text = element_text(size = 18),
        plot.subtitle = element_text(size = 18),
        legend.title = element_text(size = 18),
        legend.text = element_text(size = 18))

# credible intervals for medal rates when hosting
posterior_1_2.5 <- qgamma(0.025, shape = a1, rate = b1)
posterior_1_97.5 <- qgamma(0.975, shape = a1, rate = b1)

# credible intervals for medal rates in Olympics prior to hosting
posterior_0_2.5 <- qgamma(0.025, shape = a0, rate = b0)
posterior_0_97.5 <- qgamma(0.975, shape = a0, rate = b0)

#########################################################################################################
###################
# part 2 - hypothesis test
#########################################################################################################
###################

# test probability that host posterior > 97.5th CI percentile in previous posterior?
1 - pgamma(q = posterior_0_97.5, shape = a1, rate = b1)

# test probability that host posterior < 2.5th CI percentile in previous posterior?
pgamma(q = posterior_0_2.5, shape = a1, rate = b1)

# test probability that previous posterior > 97.5th CI percentile in host posterior?
1 - pgamma(q = posterior_1_97.5, shape = a0, rate = b0)
```

```
# monte carlo sampling from posterior distributions of medal per participant rates
set.seed(2024)

n_samples <- 100000

host_mc <- rgamma(n = n_samples, shape = a1, rate = b1)
previous_mc <- rgamma(n = n_samples, shape = a0, rate = b0)

# take means from monte carlo samples
mean(host_mc > previous_mc)

# are the results sensitive to the prior?
prior_a_list <- c(0.1, 0.5, seq(1, 5, by = 1), 10)
prior_b_list <- prior_a_list

mc_prior_results <- c()

for(a in 1:length(prior_a_list)) {
  a1_temp <- (prior_a_list[a] + y1)
  b1_temp <- (prior_b_list[a] + n1)

  a0_temp <- (prior_a_list[a] + y0)
  b0_temp <- (prior_b_list[a] + n0)

  host_mc_temp <- rgamma(n = n_samples, shape = a1_temp, rate = b1_temp)
  previous_mc_temp <- rgamma(n = n_samples, shape = a0_temp, rate = b0_temp)

  # take means from monte carlo samples
  mc_mean_temp <- mean(host_mc_temp > previous_mc_temp)

  # add to data frame storing test results
```

```
  mc_prior_results <- rbind(mc_prior_results, c(prior_a_list[a], prior_b_list[a], mc_mean_temp))
}

mc_prior_results <- as.data.frame(mc_prior_results)
colnames(mc_prior_results) <- c("prior_a", "prior_b", "mc_mean")
mc_prior_results <- mc_prior_results %>% mutate(prior_dist_character = paste0("Gamma(", prior_a, ", ", prior_b, ")"))

# table of priors and hypothesis test results
mc_prior_results %>%
  select(-c("prior_a", "prior_b")) %>%
  gt(rowname_col = "prior_dist_character")

##########################################################################################################################
# part 3 - prediction
##########################################################################################################################

olympics_data_comp <- olympics_data %>%
  mutate(host_year_previous_year_participants_ratio = participating.athletes.during.host.year /
participating.athletes.during.previous.olympics) %>%
  mutate(host_year_previous_year_participants_difference = (participating.athletes.during.host.year -
participating.athletes.during.previous.olympics)) %>%
  as.data.frame()

olympics_data_comp %>% ggplot(aes(x = host.country, y = participating.athletes.during.host.year, label =
participating.athletes.during.host.year)) +
  geom_point(stat = "identity", size = 10, fill = "blue") +
  geom_segment(aes(y = participating.athletes.during.previous.olympics, x = host.country, yend =
participating.athletes.during.host.year, xend = host.country),
               color = "blue") +
  geom_text(color = "white", size = 4) +
  labs(x = "Host Country",
```

```
      y = "Participants Difference",
      title = "Host Year Participants Compared to Previous Olympics Participants") +
  theme(axis.title = element_text(size = 24),
        plot.title = element_text(size = 24),
        axis.text = element_text(size = 18),
        plot.subtitle = element_text(size = 18)) +
  coord_flip()

#http://www.olympedia.org/countries/FRA
#france_olympic_data <- read.xlsx(file = "olympedia_france_data.xlsx", sheetIndex = 1)
# select only summer olympics data
#france_olympic_data <- france_olympic_data[grepl("Summer", france_olympic_data$Edition),]

# average host year to previous year participants ratio of countries similarly well-represented at Olympics as France
mean_host_year_previous_olympics_participant_ratio <-
mean(olympics_data_comp$host_year_previous_year_participants_ratio[which(olympics_data_comp$participating.athletes.during.previous.olympics >= 200)])

france_est_participants_2024 <- 398 * mean_host_year_previous_olympics_participant_ratio

# simulate using random medals per participant rate generated from host year posterior distribution
set.seed(2024)

n_samples_france_mc <- 100000

france_mc <- rgamma(n = n_samples_france_mc, shape = a1, rate = b1)

# create a data frame from random samples
france_forecast_df <- as.data.frame(france_mc)
colnames(france_forecast_df) <- c("estimated_medals_per_participant_rate")
france_forecast_df$forecasted_medals_won <- france_forecast_df$estimated_medals_per_participant_rate *
france_est_participants_2024
```

```
# plot distribution of forecasted number of medals won by France in 2024
france_forecast_df %>%
  ggplot(aes(x = forecasted_medals_won)) +
  #geom_bar() +
  #scale_x_binned() +
  geom_histogram(binwidth = 2) +
  geom_vline(xintercept = mean(france_forecast_df$forecasted_medals_won), color = "red", linetype = "dashed", size = 2) +
  labs(x = "Forecasted Medals Won",
       y = "Number of Simulations",
       title = "Forecasted Number of Medals Won by France in 2024 Summer Olympics") +
  theme(axis.title = element_text(size = 24),
        plot.title = element_text(size = 24),
        axis.text = element_text(size = 18),
        plot.subtitle = element_text(size = 18))

# quantiles of estimated medals won for France in 2024
quantile(france_forecast_df$forecasted_medals_won, c(0.025, 0.975))

##########################################################################################################################
# part 4 - country-specific analysis
##########################################################################################################################

# combine data for countries that have hosted twice - Australia, Japan, and USA
olympics_data_countries_combined <- olympics_data %>%
  #filter(year < 2020) %>%
  group_by(host.country) %>%
  select(-c(year)) %>%
  summarise_all(sum) %>%
```

```
  as.data.frame()

# use uninformative Gamma prior of a = b = 0.1
prior_a <- 0.1
prior_b <- prior_a

# data frame for countries posterior summaries
country_posterior_summary <- c()

# number of monte carlo samples
set.seed(2024)
n_samples <- 100000

# estimate prior and posterior distributions for each country
for(i in 1:nrow(olympics_data_countries_combined)) {
  host_country_temp <- olympics_data_countries_combined$host.country[i]

  # total medals won and participants for host countries in host year
  y1_temp <- olympics_data_countries_combined$medals.won.during.host.year[i]
  n1_temp <- olympics_data_countries_combined$participating.athletes.during.host.year[i]

  a1_temp <- (prior_a + y1_temp)
  b1_temp <- (prior_b + n1_temp)

  # total medals won and participants for host countries in previous year
  y0_temp <- olympics_data_countries_combined$medals.won.during.previous.olympics[i]
  n0_temp <- olympics_data_countries_combined$participating.athletes.during.previous.olympics[i]

  a0_temp <- (prior_a + y0_temp)
  b0_temp <- (prior_b + n0_temp)
```

```
# poisson-gamma model posterior
posterior_df_temp <- data.frame(lambda_values, dgamma(lambda_values, shape = a1_temp, rate = b1_temp),
                                dgamma(lambda_values, shape = a0_temp, rate = b0_temp))
colnames(posterior_df_temp) <- c("lambda", "host_year", "previous_year")

# plot both posterior distributions
# p_temp <- posterior_df_temp %>%
#   pivot_longer(cols = c("host_year", "previous_year")) %>%
#   ggplot(aes(x = lambda, y = value)) +
#   geom_line(aes(color = name), size = 2) +
#   scale_x_continuous(limits = c(0, 0.6)) +
#   scale_color_manual(values = c("red", "blue")) +
#   labs(x = "Estimated Medals Won per Participant",
#        y = "Posterior Density",
#        title = paste0("Host Country Advantage in the Olympics - ", host_country_temp),
#        subtitle = "Plot by Rich Ramsey") +
#   theme(axis.title = element_text(size = 24),
#         plot.title = element_text(size = 24),
#         axis.text = element_text(size = 18),
#         plot.subtitle = element_text(size = 18))
#
# plot(p_temp)

# credible intervals for medal rates when hosting
posterior_1_2.5_temp <- qgamma(0.025, shape = a1_temp, rate = b1_temp)
posterior_1_97.5_temp <- qgamma(0.975, shape = a1_temp, rate = b1_temp)

# credible intervals for medal rates in Olympics prior to hosting
posterior_0_2.5_temp <- qgamma(0.025, shape = a0_temp, rate = b0_temp)
posterior_0_97.5_temp <- qgamma(0.975, shape = a0_temp, rate = b0_temp)
```

```
  # hypothesis test for whether host year posterior estimate is less than previous year posterior estimate
  host_year_p_gt_previous_year_temp <- (1 - pgamma(q = posterior_0_97.5_temp, shape = a1_temp, rate = b1_temp))
  previous_year_p_gt_host_year_temp <- (1 - pgamma(q = posterior_1_97.5_temp, shape = a0_temp, rate = b0_temp))

  # monte carlo samples from host year and previous year posterior distributions
  host_mc_temp <- rgamma(n = n_samples, shape = a1_temp, rate = b1_temp)
  previous_mc_temp <- rgamma(n = n_samples, shape = a0_temp, rate = b0_temp)

  # take means from monte carlo samples
  mc_pct_host_gt_previous_temp <- mean(host_mc_temp > previous_mc_temp)
  mc_host_previous_estimated_ratio_temp <- mean(host_mc_temp / previous_mc_temp)

  country_posterior_summary <- rbind(country_posterior_summary,
                                     c(host_country_temp, posterior_1_2.5_temp, posterior_1_97.5_temp,
                                       posterior_0_2.5_temp, posterior_0_97.5_temp,
                                       host_year_p_gt_previous_year_temp, previous_year_p_gt_host_year_temp,
                                       mc_pct_host_gt_previous_temp, mc_host_previous_estimated_ratio_temp))
}

# put country posterior summary into data frame and set column names
country_posterior_summary <- as.data.frame(country_posterior_summary)
country_posterior_summary <- country_posterior_summary %>% mutate_at(c(2:ncol(country_posterior_summary)),
as.numeric)
colnames(country_posterior_summary) <- c("country", "host_year_posterior_ci_2.5", "host_year_posterior_ci_97.5",
                                         "previous_year_posterior_ci_2.5", "previous_year_posterior_ci_97.5",
                                         "prob_host_year_gt_previous_year", "prob_previous_year_gt_host_year",
                                         "mc_pct_host_year_gt_previous_year", "mc_host_year_previous_year_estimated_ratio")

# display summary of country posterior distributions as gt object
country_posterior_summary %>%
  select(-c("prob_host_year_gt_previous_year", "prob_previous_year_gt_host_year")) %>%
```

```
  gt(rowname_col = "country") %>%

  tab_spanner(label = "Host Year Olympics", columns = contains("host_year_posterior")) %>%

  tab_spanner(label = "Previous Olympics", columns = contains("previous_year_posterior")) %>%

  tab_spanner(label = "Comparison - Monte Carlo Samples From Country-Specific Posteriors", columns = contains("mc_")) %>%

  fmt_percent(columns = contains("pct"), decimals = 1) %>%

  fmt_number(columns = !contains("pct"), decimals = 2) %>%

  cols_label(host_year_posterior_ci_2.5 = "2.5th Percentile",

             host_year_posterior_ci_97.5 = "97.5th Percentile",

             previous_year_posterior_ci_2.5 = "2.5th Percentile",

             previous_year_posterior_ci_97.5 = "97.5th Percentile",

             mc_pct_host_year_gt_previous_year = "% Samples Host Year > Previous Olympics",

             mc_host_year_previous_year_estimated_ratio = "Mean Ratio of Host Year to Previous Olympics") %>%

  #add outer border to table (gray)

  opt_table_outline(style = "solid", width = px(3)) %>%

  # # add left and right borders to body

  tab_style(

    style = cell_borders(sides = c("right"), color = "#D3D3D3", weight = px(3), style = "solid"),

    locations = cells_body(columns = c(host_year_posterior_ci_97.5, previous_year_posterior_ci_97.5))) %>%

  # highlight host countries where at least 90% of samples from host year posterior were greater than samples from previous Olympics posterior

  tab_style(

    style = list(cell_fill(color = "yellow")),

    locations = cells_body(

      columns = c(mc_host_year_previous_year_estimated_ratio, mc_pct_host_year_gt_previous_year),

      rows = mc_pct_host_year_gt_previous_year > 0.9

    )

  )


#################################################################################################################

# part 5 - conclusions
```

##
###################