

# 주짓수 대회 분석

US Grappling 2009~2019 대회 데이터를 활용하여

찰스 브리세트, 고려대학교 통계학과

11/26/2021

# 목차

- 1 탐구 동기와 목표
  - 1.1 주짓수 관련 단어
  - 1.2 US Grappling 데이터
- 2 시간이 지남에 따른 경기 지속 시간의 변화
  - 2.1 경기 지속 시간의 분포는 어떻게 생겼을까?
  - 2.2 경기 지속 시간의 분포는 로그-정규분포 일까?
  - 2.3 대회 간 경기 지속 시간의 평균은 다른가?
  - 2.4 Games-Howell 사후 검정 결과
- 3 서브미션에 대한 인기의 변화
  - 3.1 대회 간 서브미션의 비율은 다른가?
  - 3.3 Fisher의 정확 검정 결과
  - 3.4 Straight Ankle에 대한 사후 검정 결과
  - 3.5 Straight Ankle - 대회 별로 비율
- 4 서브미션과 기술 수준의 관계
  - 4.1 기술 수준이 서브미션에 영향을 미치는가?
  - 4.2 Fisher의 정확 검정 결과
  - 4.3 기술수준별 서브미션 비율
- 5 토의와 추후 연구 문제
- 6 부록 R 스크립트

# 1 탐구 동기와 목표

## | 동기

- 다른 스포츠에 비하여 주짓수에서는 데이터 분석을 기반한 접근이 부족함
- 데이터 분석을 주짓수에서 활용하면 더 개선된 코칭과 경기 준비를 할 수 있을 것으로 생각하였음

## | 분석의 목표

- 분석 1: 지난 11년간이 지만에 따라 경기 지속 시간에 변화가 있었는가?
- 분석 2: 지난 11년간 시간이 지남에 따라 인기 있는 서브미션 종류에 변화가 있었는가?
- 분석 3: 기술 수준이 서브미션에 영향을 미치는가?

## 1.1 주짓수 관련 단어

### | US Grappling이란?

- 2006년부터 미국 중대서양 지역에서 주짓수 대회를 개최하는 조직

### | 서브미션이란?

- 상대의 항복 사인을 받기 위한 행동
  - 예: 관절을 꺾기, 경동맥등을 조르기, 등

### | 주짓수 기술 수준

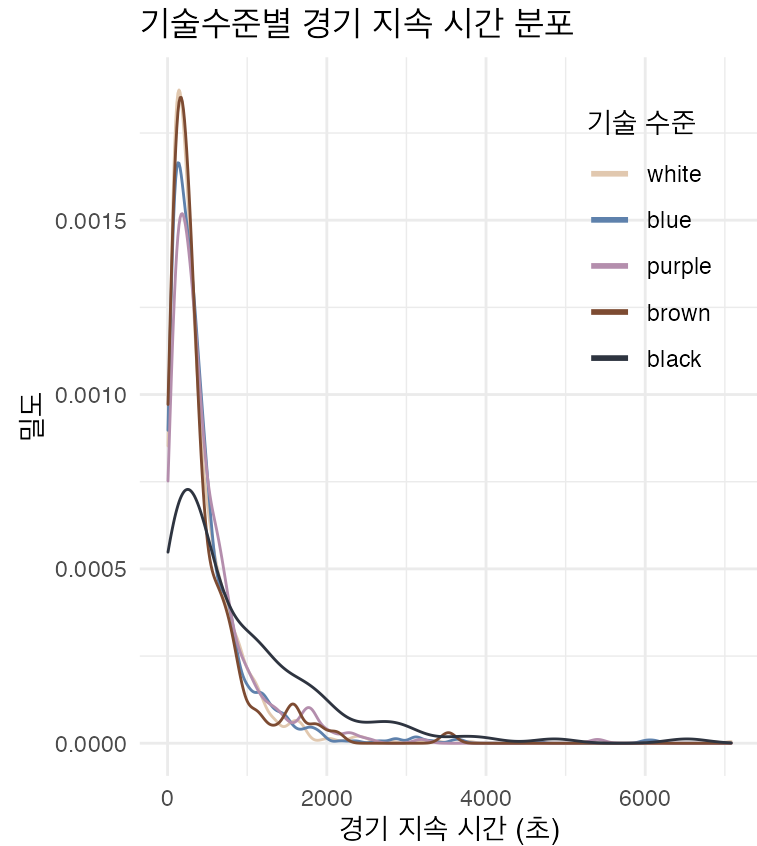
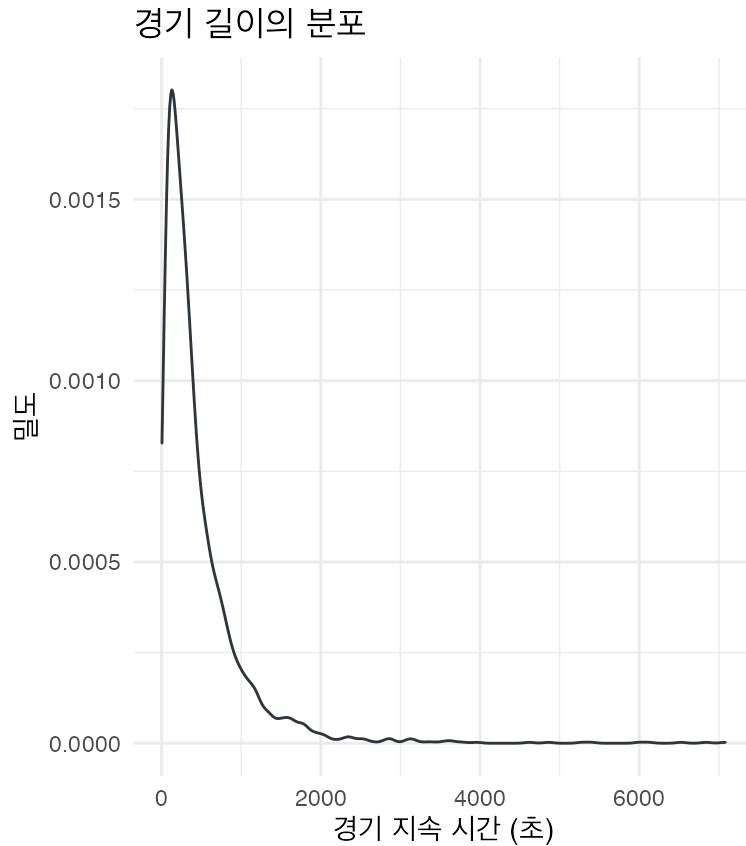
- 다른 무술과 동일하게 도복의 띠로 기술 수준을 나눔
  - 순서: 흰 띠, 파란 띠, 보라 띠, 갈 띠, 검은 띠
  - 흰 띠부터 검은 띠까지 걸리는 평균 기간은 12년

## 1.2 US Grappling 데이터

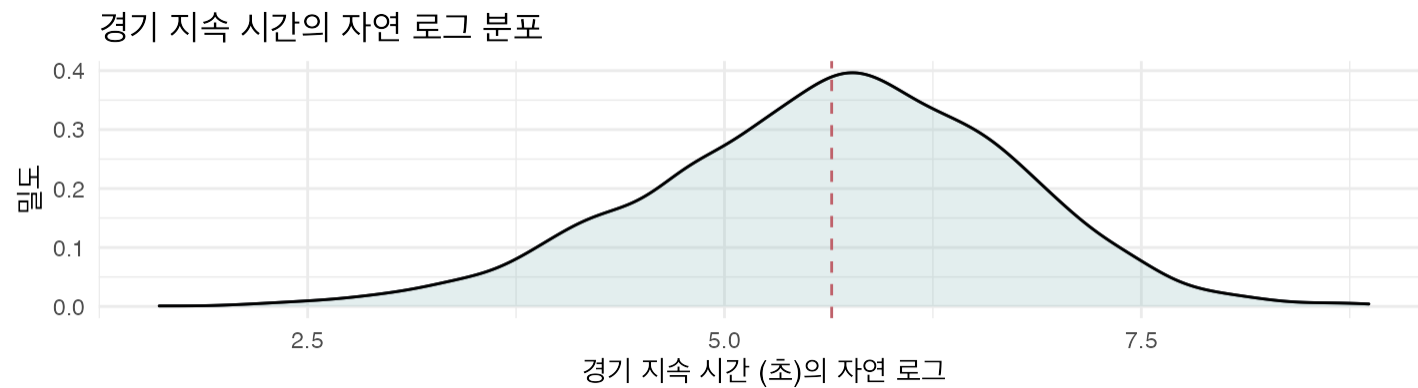
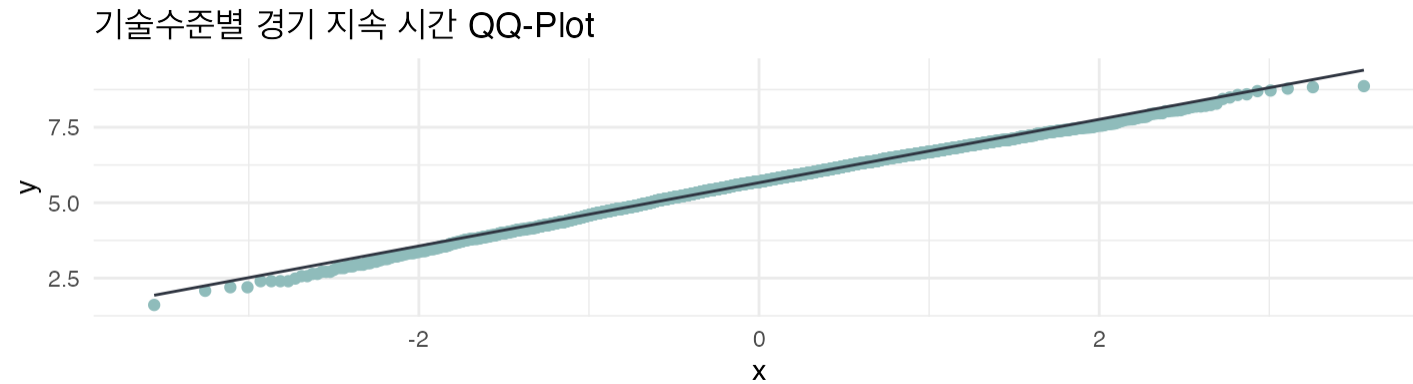
- 서브미션 온리 주짓수 대회에서 심판과 경기 보조원이 수집
- 2009~2019년까지 미국 동부에서 열린 총23회의 서브미션 온리 주짓수 대회
- 경기 지속 시간, 기술 수준, 서브미션 등 여러 변수에 대한 정보
- US Grappling에 재직 중인 Chrissy Linzy로부터 googledrive 링크 입수

## 2 시간이 지남에 따른 경기 지속 시간의 변화

### 2.1 경기 지속 시간의 분포는 어떻게 생겼을까?



## 2.2 경기 지속 시간의 분포는 로그-정규분포 일까?



## 2.3 대회 간 경기 지속 시간의 평균은 다른가?

### | 가설

- $H_0$ : 대회마다 평균 경기 지속 시간이 동일하다
- $H_1$ : 대회마다 평균 경기 지속 시간이 동일한 것은 아니다

### | 데이터 분석 절차

- 변수 간관계 분석 Welch's ANOVA
- Games-Howell 사후 검정



## 2.4 Games-Howell 사후 검정 결과

- 결과에 따라 귀무가설 기각하는 증거가 있다

belt	date1	date2	p.adj	p.adj.signif
black	2009-12-05	2017-07-29	0.040	.
black	2009-12-05	2018-08-18	0.044	.

### 3 서브미션에 대한 인기의 변화



## 대회 간 서브미션의 비율은 다른가?

### | 가설

- $H_0$ : 대회마다 사용된 서브미션의 비율은 동일하다
- $H_1$ : 대회마다 사용된 서브미션의 비율은 동일한 것은 아니다

### | 비율의 차에 대한 검정

- `prop.test()` 사용
- “Chi-squared approximation may be incorrect” 경고 발생
- Fisher의 정확 검정 (Fisher's Exact Test) 시도
- Pairwise Fisher 검정 시도

### 3.3 Fisher의 정확 검정 결과

- 결과에 따라 귀무가설은 straight ankle 경우만에 기각하는 증거가 있다

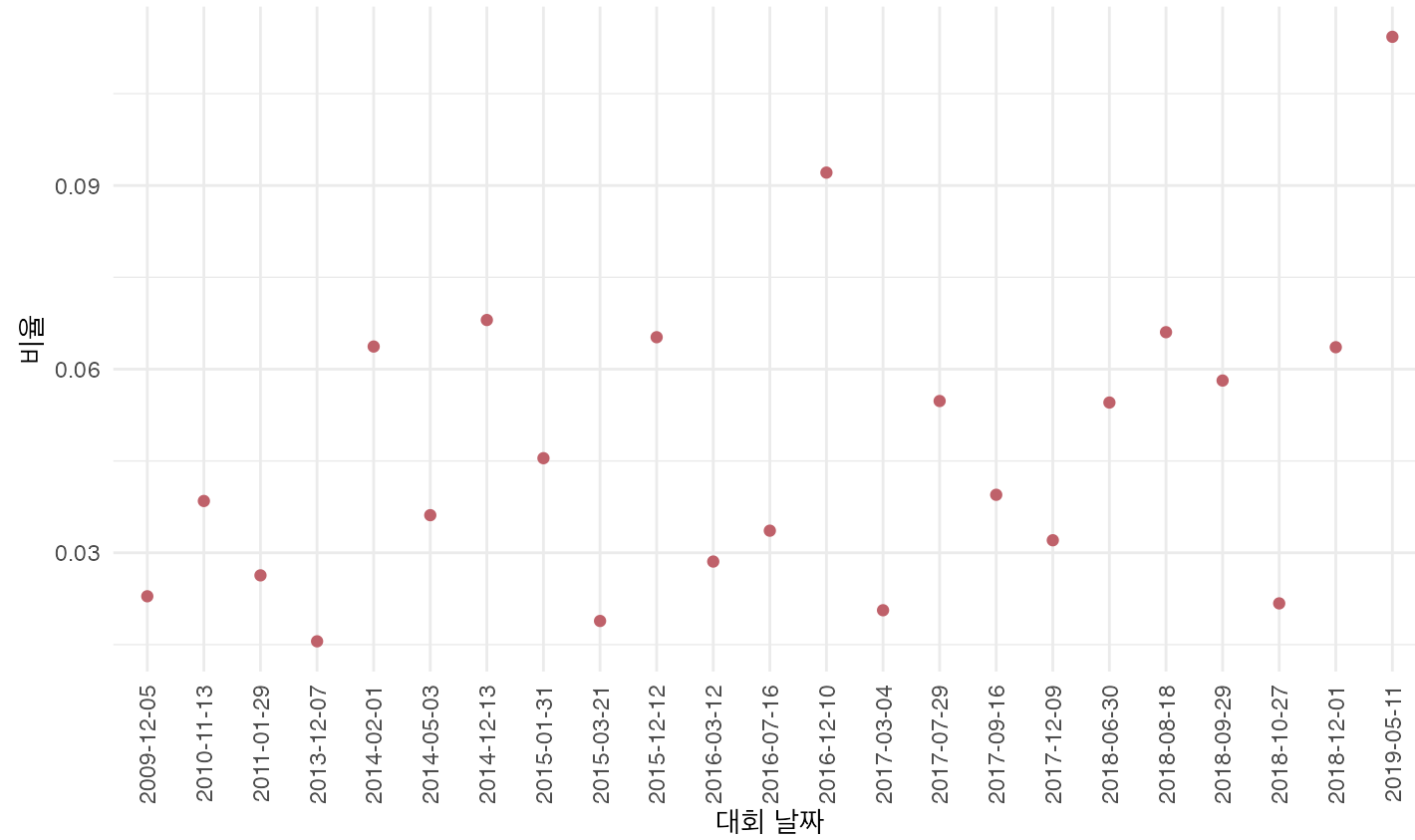
submission	p_value
straight ankle	0.0414793
armbar	0.0489755
bow and arrow	0.0959520
lapel choke	0.1239380
cross collar	0.1284358
triangle	0.1629185

### 3.4 Straight Ankle에 대한 사후 검정 결과

- Pairwise Fisher 검정 결과에 따라 귀무가설 기각할 증거가 부족하다

group1	group2	p.adj	p.adj.signif
row1	row21	0.0941	ns
row11	row21	0.4310	ns
row1	row2	1.0000	ns
row1	row3	1.0000	ns
row1	row4	1.0000	ns
row1	row5	1.0000	ns

### 3.5 Straight Ankle - 대회 별로 비율



## 4 서브미션과 기술 수준의 관계

arm triangle  
ezeziel  
bow and arrow  
key lock  
cross collar  
straight ankle  
mata leão

## 4.1 기술 수준이 서브미션에 영향을 미치는가?

### | 가설

- $H_0$ : 모든 기술 수준마다 사용한 서브미션의 비율은 동일하다
- $H_1$ : 모든 기술 수준마다 사용한 서브미션의 비율은 동일한 것은 아니다

### | 비율의 차에 대한 검정

- Fisher의 정확 검정 (Fisher's Exact Test) 시도

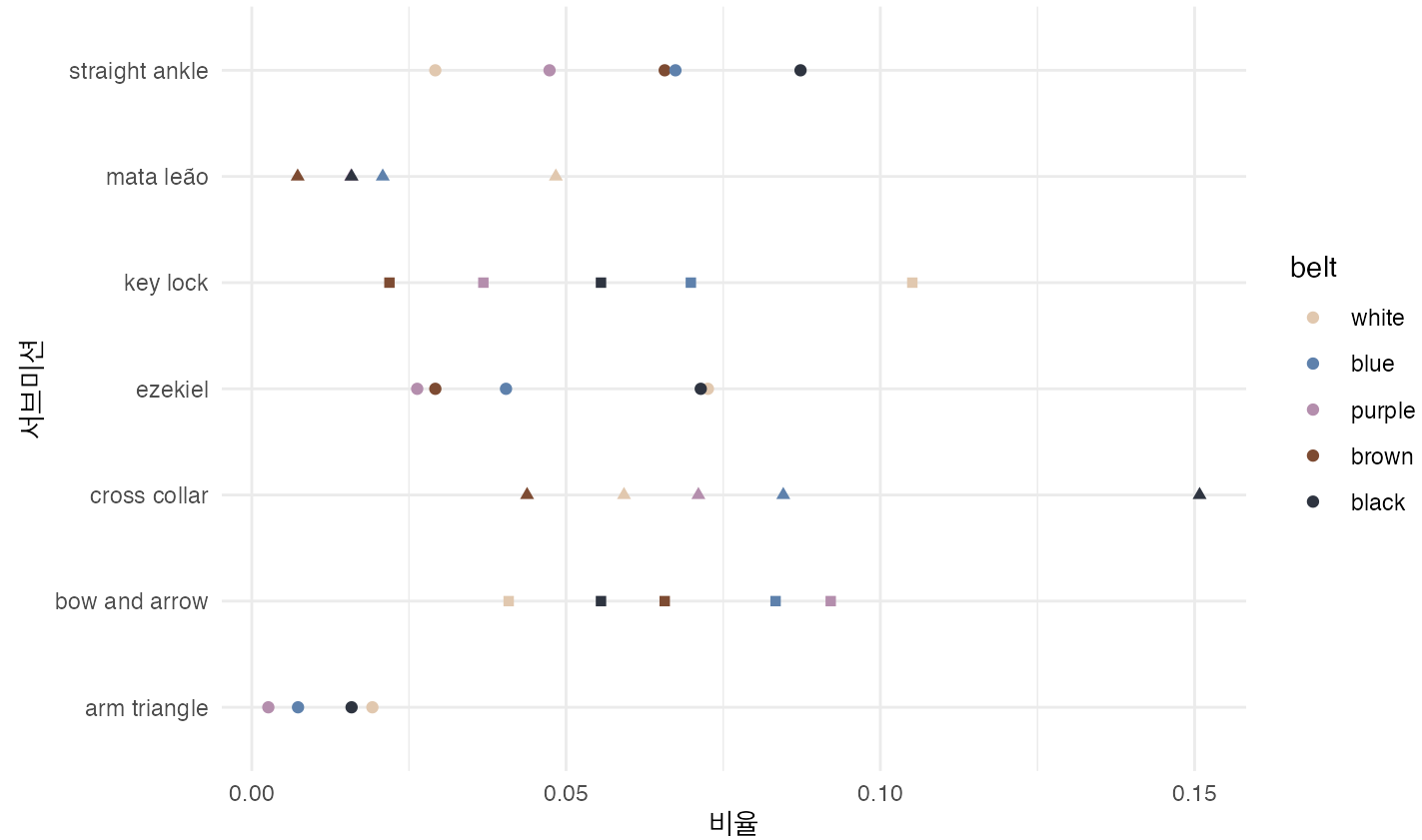


## 4.2 Fisher의 정확 검정 결과

- 결과에 따라 7개의 서브미션에서 귀무가설을 기각하는 증거가 있다

submission	p_value
bow and arrow	0.0002034
ezekiel	0.0005924
key lock	0.0000038
cross collar	0.0026753
straight ankle	0.0001757
arm triangle	0.0222817
mata leão	0.0004791

## 4.3 기술수준별 서브미션 비율



# 5 토의와 추후 연구 문제

## | 토의

- 차시마다 일관된 방식으로 데이터 수집이 진행되지 않아 분석에 영향을 미쳤을 수 있음
  - 예를 들어, 어떤 차수에는 “choke”라고 변수명이 설정되어 있으나, 다른 차수에는 “choke”의 종류가 세분화되어서 데이터가 수집됨
  - 따라서, “choke”의 종류에 속하는 “bow and arrow”, “cross collar”, “ezekiel”과 “mata leão”에 대한 분석을 수행할 때 (p.17), choke의 종류를 나누지 않고 데이터가 수집된 차수(42개 경기의 데이터)를 사용하지 못하였음
  - 이와 같은 데이터 수집의 한계가 결과에도 영향을 미쳤을 수 있음
- 경기 지속 시간에 더 적합한 분포가 있을 것인가?

## | 추후 연구 문제

- 도복(Gi) 주짓수 대 비도복(Nogi) 서브미션 비율
- 경기 지속 시간과 서브미션의 관계

## ## 6 부록 - R 스크립트

```
# libraries
library(tidyverse)
library(google sheets4)
library(tidymodels)
library(lubridate)
library(rstatix)
library(viridis)
library(nord)
library(kableExtra)
library(fmsb)
library(wordcloud)
library(ragg)

# seed for word cloud
seed_date <- ymd(20181217) # date I got my brown belt
set.seed(as.numeric(seed_date)) # for reproducibility

# categories
belt_lvls <- c("white", "blue", "purple", "brown", "black")
colors <- c("#E1C8AF", "#5E81AC", "#B48EAD", "#7D4B32", "#2E3440")

#import data from google sheets
usg_so_id <-
  as_sheets_id("https://docs.google
.com/spreadsheets/d/1L-atR2L9QWUmrLxGxurVByrqYTtZXzNKRKzbv9i3ZiQ/edit#gid=1253508644")
#vector of individual sheet names
s_names <- sheet_names(usg_so_id)

## read from googlesheets and do some basic tidying
get_dat <- function(s_name, id) {
  # get date from sheet name
  day <- as.Date(str_extract(s_name,
                             "[JFMASOND][:lower:]+\s[:digit:]{1,2}, [:digit:]{4}"), "%b %d, %Y")

  # read in the sheet
  sheet_dat <- id %>% read_sheet(sheet = s_name)
  # rename variables to make them easier to work with
```

```

sheet_dat <- sheet_dat %>%
  rename_with(~tolower(gsub("/", "_", .x, fixed = TRUE))) %>%
  rename_with(~gsub(" ", "_", .x, fixed = TRUE)) %>%
  # set na to 0 for time values
  mutate(hours = replace_na(hours, 0)) %>%
  mutate(minutes = replace_na(minutes, 0)) %>%
  mutate(seconds = replace_na(seconds, 0)) %>%
  # create in_seconds column
  mutate(in_seconds = hours*3600 + minutes*60 + seconds,
    .after = "seconds") %>%
  # conform observations
  mutate(across(!in_seconds, tolower)) %>%
  mutate(division = str_remove(division, "'s$")) %>%
  mutate(division = str_remove(division, "s$")) %>%
  mutate(submission = str_replace(submission, "&", "and")) %>%
  mutate(submission = str_replace(submission, "/", " ")) %>%
  # remove beginner, juvenile, and master (30+) divisions
  filter(division != "juvenile", belt_skill != "beginner",
    !str_starts(division, "30+"), !str_starts(belt_skill, "30+")) %>%
  # remove observations with 0 in in_seconds
  filter(!(in_seconds==0)) %>% # matches that were forfeited or time not recorded
  # change NA observations in gi_no_gi based on belt_skill
  mutate(gi_no_gi = if_else(belt_skill %in% belt_lvls, "gi", "no gi")) %>%
  # filter to gi only
  filter(gi_no_gi == "gi") %>%
  # change rename belt_skill and change into an ordered factor
  #rename(belt = belt_skill) %>%
  mutate(belt = factor(belt_skill, order=TRUE, levels=belt_lvls)) %>%
  # select only desired variables
  select(in_seconds, belt, submission) %>%
  # add match day column
  add_column(date = day)
}

# get data from each sheet except first two
usg_so_gi <- s_names[-c(1,2)] %>% map_df(~ get_dat(., usg_so_id))

```

```

# conform submission observations
dat <- usg_so_gi %>%
  mutate(submission = replace_na(submission, "unknown")) %>%
  mutate(submission = str_remove(submission, "(^mounted|mounted$)")) %>%
  mutate(submission = str_remove(submission, "^(reverse|inverted|modified)")) %>%
  mutate(submission = str_remove(submission, "((?<=arrow)|(?<=triangle))\\schoke$")) %>%
  mutate(submission = str_replace(submission, "^dq.*|.dq$", "dq")) %>%
  mutate(submission = str_replace(submission, ".*(arm\\s?lock|ar?mb?ar)$", "armbar")) %>%
  mutate(submission = str_replace(submission, ".*g(ui|iu)llotine$", "guillotine")) %>%
  mutate(submission = str_replace(submission, "^(\\?|not).*", "unknown")) %>%
  mutate(submission = str_replace(submission, "(head\\s(and|arm)|arm\\s(triangle).*", "arm triangle")) %>%
  mutate(submission = str_replace(submission, "ns choke", "north south choke")) %>%
  mutate(submission = str_replace(submission, "cho[jk]e", "choke")) %>%
  mutate(submission = str_replace(submission, ".*(ankle|foot)\\s?lock$", "straight ankle")) %>%
  mutate(submission = str_replace(submission, "^(cross|collar).*(collar|choke)$", "cross collar")) %>%
  mutate(submission = str_replace(submission, "rnc", "mata leão")) %>%
  mutate(submission = str_replace(submission, "^(darce|d'arce).*", "darce")) %>%
  mutate(submission = str_replace(submission, "om[oea]plata", "omoplata")) %>%
  mutate(submission = str_replace(submission, "bra[bv]o.*", "brabo choke")) %>%
  mutate(submission = str_replace(submission, ".*cutter.*", "paper cutter choke")) %>%
  mutate(submission = str_replace(submission, "e?z[ei].*", "ezekiel")) %>%
  mutate(submission = str_replace(submission, "he.*o[oc]k", "heel hook")) %>%
  mutate(submission = str_replace(submission, "^anaconda.*", "anaconda")) %>%
  mutate(submission = str_replace(submission, "americana|keylock", "key lock")) %>%
  mutate(submission = str_replace(submission, ".*arm$", "armbar")) %>%
  mutate(submission = str_replace(submission, "knee bar", "kneebar")) %>%
  mutate(submission = str_replace(submission, "toe hold", "toehold")) %>%
  mutate(submission = str_replace(submission, "shoulder lock", "key lock")) %>%
  mutate(submission = str_replace(submission, "^baseball.*choke$", "baseball choke")) %>%
  mutate(submission = str_replace(submission, ".*leaf$", "texas cloverleaf")) %>%
  mutate(submission = str_trim(submission, "both"))

# dat %>% summarise_all(~n_distinct())

# get the overall proportion of each submission
dat <- dat %>%
  add_count(submission) %>%

```

```
group_by(submission, n) %>%
nest() %>%
ungroup() %>%
mutate(P = n / sum(n), .after = n) %>%
unnest(data)
```

```
# match length dist plot 1
```

```
dat %>%
  ggplot(aes(x = in_seconds)) +
  geom_density(color = nord(palette="polarnight")[1]) +
  labs(title = "경기 길이의 분포", x = "경기 지속 시간 (초)", y = "밀도") +
  theme_minimal()
```

```
# match length dist plot 2
```

```
dat %>%
  ggplot(aes(x = in_seconds, color = belt)) +
  geom_density(show.legend=FALSE) +
  stat_density(geom="line", position="identity", size = 0) +
  guides(color = guide_legend(override.aes=list(size = 1))) +
  scale_color_manual(values = colors) +
  labs(title = "기술수준별 경기 지속 시간 분포",
       x = "경기 지속 시간 (초)",
       y = "밀도",
       color = "기술 수준") +
  theme_minimal() +
  theme(legend.position = c(.95, .95),
       legend.justification = c("right", "top"),
       legend.box.just = "right",
       legend.margin = margin(6, 6, 6, 6))
```

```
# match length log-normal
```

```
seconds_ln <- dat %>%
  select(in_seconds) %>%
  summarize(ln = log(in_seconds))
```

```
# match length log-normal qq plot
```

```
seconds_ln %>%
  ggplot(aes(sample=ln)) +
```

```

stat_qq(color = nord(palette="frost")[1]) +
stat_qq_line(col = nord(palette="polarnight")[1]) +
labs(title = "기술수준별 경기 지속 시간 QQ-Plot") +
theme_minimal()

```

```

# match length log-normal dist plot
seconds_ln %>%
  ggplot(aes(x=ln)) +
  geom_density(color = nord(palette="polarnight")[1]) +
  geom_density(fill = nord(palette="frost")[1], alpha = 0.25) +
  geom_vline(xintercept = mean(seconds_ln$ln),
             col = nord(palette = "aurora")[1],
             lty = "dashed") +
  labs(title = "경기 지속 시간의 자연 로그 분포",
       x = "경기 지속 시간 (초)의 자연 로그", y = "밀도") +
  theme_minimal()

```

```

# tidy data for answering q1
q1_dat <- dat %>%
  # calculate ln for match times
  mutate(ln = log(in_seconds)) %>%
  # only need belt_skill, date, and the natural log of the match length in seconds
  select(belt, date, ln) %>%
  # count how many matches occurred by belt_skill on each match day
  group_by(belt, date) %>%
  mutate(n = n()) %>%
  mutate(mean = mean(ln)) %>%
  ungroup() %>%
  # filter out those belt_skills which had less than X on a given match day
  filter(!n < 4) %>%
  # prepare for one way analysis of variance
  mutate(date = as.factor(date)) %>%
  group_by(belt) %>%
  nest()

```

```

# welch anova test
ans1 <- q1_dat %>%
  mutate(waov = map(.x = data,

```



```

        .f = ~oneway.test(.x$ln ~ .x$date,
                          var.equal = FALSE))) %>%
mutate(p_value = map_dbl(.x = waov, .f = ~.x$p.value)) %>%
filter(p_value < 0.05)

# games-howell post hoc test
ans1_ph <- ans1 %>%
  unnest(cols = c(data)) %>%
  games_howell_test(ln ~ date) %>%
  group_by(belt) %>%
  filter(p.adj < 0.05) %>%
  select(belt, group1, group2, p.adj, p.adj.signif) %>%
  rename(date1 = group1, date2 = group2)

# games-howell results table
ans1_ph %>%
  kbl() %>%
  kable_minimal(full_width = FALSE)

# submission word cloud
dat %>%
  count(submission, sort = TRUE) %>%
  filter(submission != "unknown", submission != "dq") %>%
  with(wordcloud(submission, n, random.order = FALSE, max.words = 20,
                 colors=nord(palette = "moose_pond")))

# overall top submissions
sub_NP <- dat %>%
  # get submission count
  count(submission) %>%
  # proportion of each submission
  mutate(p = n / sum(n)) %>%
  # remove any submission that has a proportion less than 0.01
  # or was unknown, dq, or choke
  filter(p >= 0.01, submission != "unknown", submission != "dq",
         submission != "choke", submission != "injury")

top_subs <- sub_NP$submission

```

```

# tidy data for answering q2
q2_dat <- dat %>%
  select(date, submission) %>%
  group_by(date) %>%
  mutate(bouts = n()) %>%
  ungroup() %>%
  filter(submission %in% top_subs) %>%
  group_by(date, submission) %>%
  mutate(n = n()) %>%
  mutate(p = n / bouts) %>%
  ungroup() %>%
  distinct() %>%
  mutate(date = as.factor(date)) %>%
  group_by(submission) %>%
  nest()

# q2 prop test
ans2_sa_p <- q2_dat %>%
  mutate(p_test = map(.x = data,
                      .f = ~prop.test(.x$n, .x$bouts))) %>%
  mutate(p_value = map_dbl(.x = p_test, .f = ~.x$p.value)) %>%
  filter(p_value < 0.05)

# prop.test gave chi-squared approximation may be incorrect warnings
# trying fisher's exact test
ans2 <- q2_dat %>%
  mutate(f_test = map(.x = data,
                      .f = ~fisher.test(cbind(.x$n, .x$bouts-.x$n),
                                           simulate.p.value=TRUE))) %>%
  mutate(p_value = map_dbl(.x = f_test, .f = ~.x$p.value))

# fisher's test results table
ans2 %>%
  select(submission, p_value) %>%
  arrange(p_value) %>%
  head() %>%
  kbl() %>%
  kable_minimal(full_width = FALSE)

```

```

# pairwise fisher test
ans2_sa <- ans2 %>%
  filter(submission == "straight ankle") %>%
  mutate(pwf_test = map(data,
    ~pairwise_fisher_test(cbind(.x$n, .x$bouts-.x$n))))

# pairwise fisher test results table
ans2_sa$pwf_test[[1]] %>%
  select(group1, group2, p.adj, p.adj.signif) %>%
  arrange(p.adj) %>%
  head() %>%
  kbl() %>%
  kable_minimal(full_width = FALSE)

# straight ankle prop by date plot
ans2_sa %>%
  unnest(data) %>%
  ggplot(aes(x = date)) +
  geom_point(aes(y = p), color = nord(palette="aurora")[1]) +
  theme_minimal() +
  labs(y = "비율", x = "대회 날짜") +
  theme(legend.position = "none",
    axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

# tidy data for answering q3
q3_dat <- dat %>%
  select(belt, submission, P) %>%
  group_by(belt) %>%
  mutate(bouts = n()) %>%
  ungroup() %>%
  filter(submission %in% top_subs) %>%
  group_by(belt, submission) %>%
  mutate(n = n()) %>%
  mutate(p = n / bouts) %>%
  ungroup() %>%
  distinct() %>%
  group_by(submission) %>%
  nest()

```

```

# q3 fisher's test
ans3 <- q3_dat %>%
  mutate(f_test = map(.x = data, .f = ~fisher.test(cbind(.x$n, .x$bouts-.x$n),
                                                    workspace=2e8))) %>%
  mutate(p_value = map_dbl(.x = f_test, .f = ~.x$p.value)) %>%
  filter(p_value < 0.05)

# submission word cloud based on fisher's test results
a3_subs <- ans3$submission

dat %>%
  filter(submission %in% a3_subs) %>%
  count(submission, sort = TRUE) %>%
  with(wordcloud(submission, n, scale = c(3, 0.25), random.order = FALSE,
            colors=nord(palette = "moose_pond")))

# fisher's test results table
ans3 %>%
  select(submission, p_value) %>%
  kbl() %>%
  kable_minimal(full_width = FALSE)

# proportion plot of submissions based on fisher's test results
shapes <- c("circle", "square", "triangle", "circle",
           "square", "triangle", "circle")

ans3 %>%
  unnest(data) %>%
  ggplot(aes(x = submission)) +
  geom_point(aes(y = p, color = belt, shape = submission)) +
  coord_flip() +
  theme_minimal() +
  scale_color_manual(values = colors) +
  scale_shape_manual(values = shapes, guide = FALSE) +
  labs(x = "서브미션", y = "비율") +
  theme()

```