

# Preliminary analysis

Samuel Pawel, Björn Siepe, František Bartoš

11 August 2023

## Visualizations

```
## libraries
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyr)
library(ggplot2)
library(colorspace)
library(ggpubr)
library(stringr)
library(forcats)
library(knitr)
library(kableExtra)

##
## Attaching package: 'kableExtra'
## The following object is masked from 'package:dplyr':
##
##   group_rows
# devtools::install_github("kupietz/kableExtra")

theme_set(theme_bw() +
  theme(legend.position = "top",
        panel.grid.minor = element_blank()))
pal <- "Harmonic" # change palette here
## colorspace::hcl_palettes("qualitative", plot = TRUE)

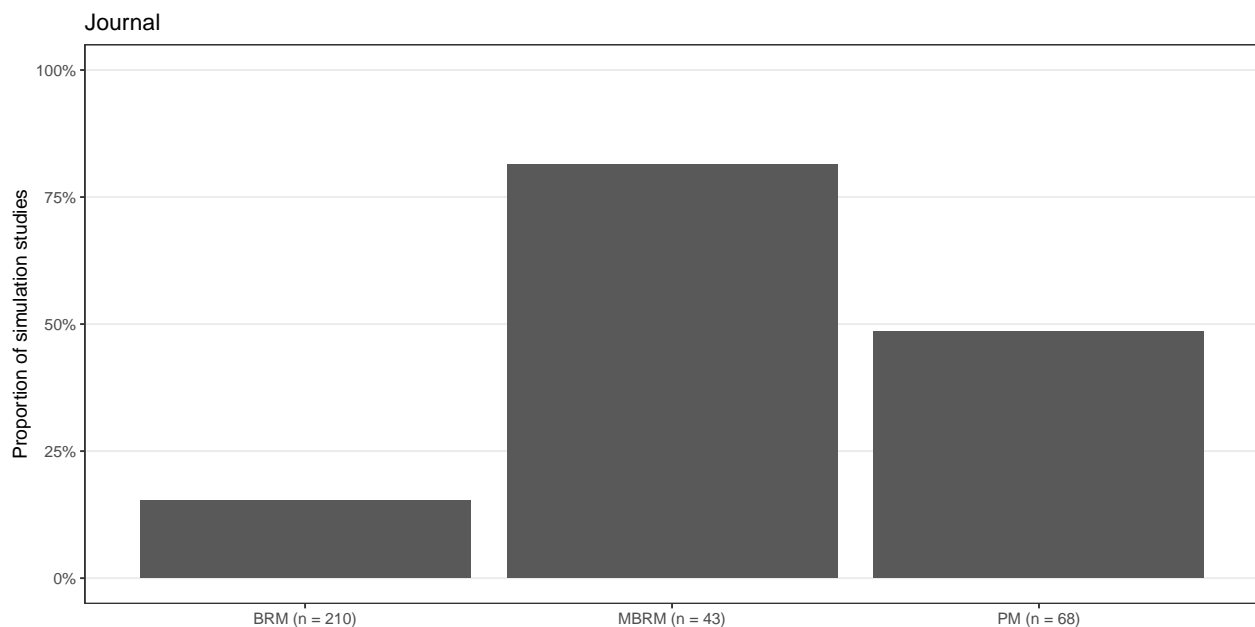
## data
sim_res_fac_full <- readRDS(file = "data/sim_res_fac.RDS")
sim_res_num_full <- readRDS(file = "data/sim_res_num.RDS")
```

```

# subset assessment only
sim_res_fac <- sim_res_fac_full %>%
  filter(simstudy_q1 == "yes",
         coding_type == "assessment")
sim_res_num <- sim_res_num_full %>%
  filter(simstudy_q1 == "yes",
         coding_type == "assessment")

## proportion of simulation studies by journal
sim_res_fac_full %>%
  filter(coding_type == "assessment" | is.na(coding_type)) %>%
  group_by(journal) %>%
  summarize(propSim = mean(simstudy_q1 == "yes"),
            n = n()) %>%
  mutate(journalLab = paste0(journal, " (n = ", n, ")")) %>%
  ggplot(aes(x = journalLab, y = propSim)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = scales::percent, limits = c(0, 1)) +
  labs(x = NULL, title = "Journal", y = "Proportion of simulation studies") +
  theme(panel.grid.major.x = element_blank())

```



```

sim_res_fac_full %>%
  filter(coding_type == "assessment" | is.na(coding_type),
         year == 2022) %>%
  summarize(propSim = mean(simstudy_q1 == "yes"),
            sim = sum(simstudy_q1 == "yes"),
            n = n())

```

```

## # A tibble: 1 x 3
##   propSim  sim    n
##   <dbl> <int> <int>
## 1  0.343    85   248

```

```

sim_res_fac_full %>%
  filter(coding_type == "assessment" | is.na(coding_type),

```

```

    year == 2022) %>%
group_by(journal) %>%
summarize(propSim = mean(simstudy_q1 == "yes"),
           sim = sum(simstudy_q1 == "yes"),
           n = n())

```

```

## # A tibble: 3 x 4
##   journal propSim   sim     n
##   <fct>     <dbl> <int> <int>
## 1 BRM       0.156    24   154
## 2 MBRM      0.814    35    43
## 3 PM        0.510    26    51

```

```

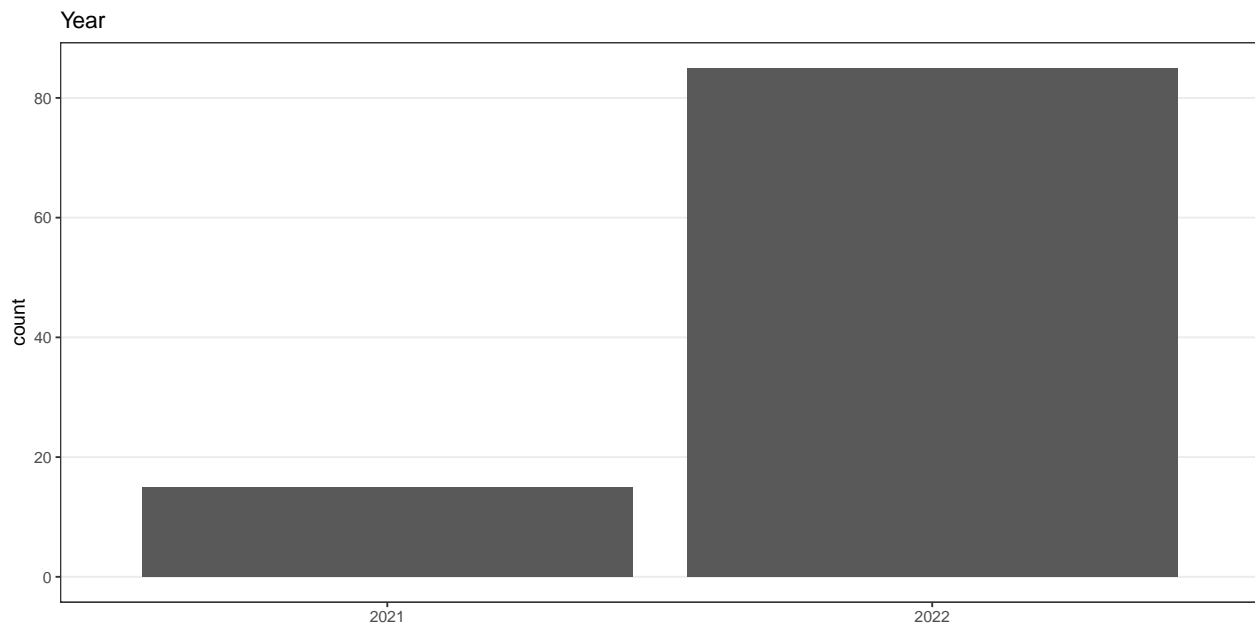
## Morris et al. (2019) find
## "264 articles of which 199 (75\%) included at least one simulation study"

```

```

## year
ggplot(data = sim_res_fac, aes(x = factor(year))) +
  geom_bar() +
  labs(x = NULL, title = "Year") +
  theme(panel.grid.major.x = element_blank())

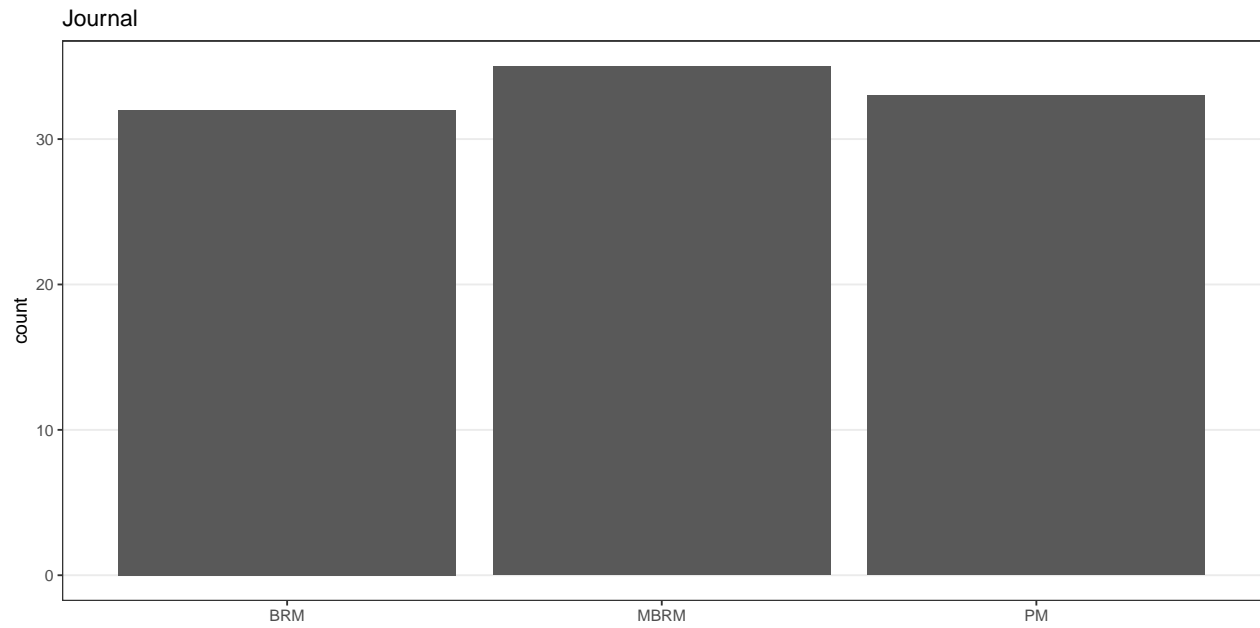
```



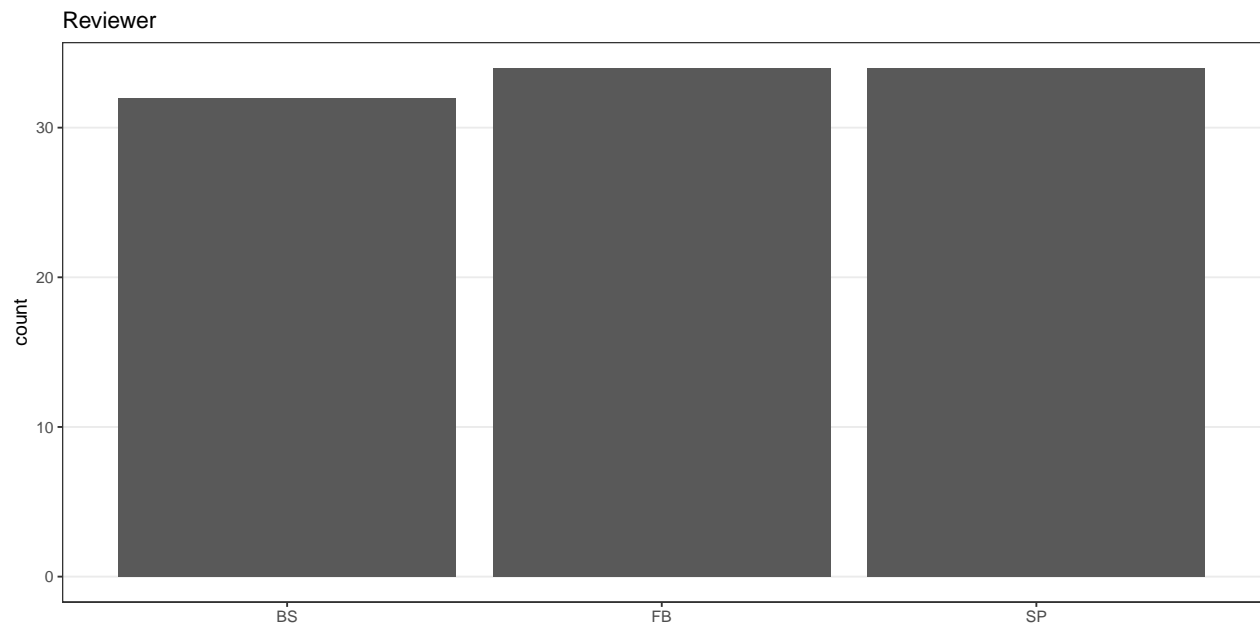
```

## journal
ggplot(data = sim_res_fac, aes(x = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Journal") +
  theme(panel.grid.major.x = element_blank())

```

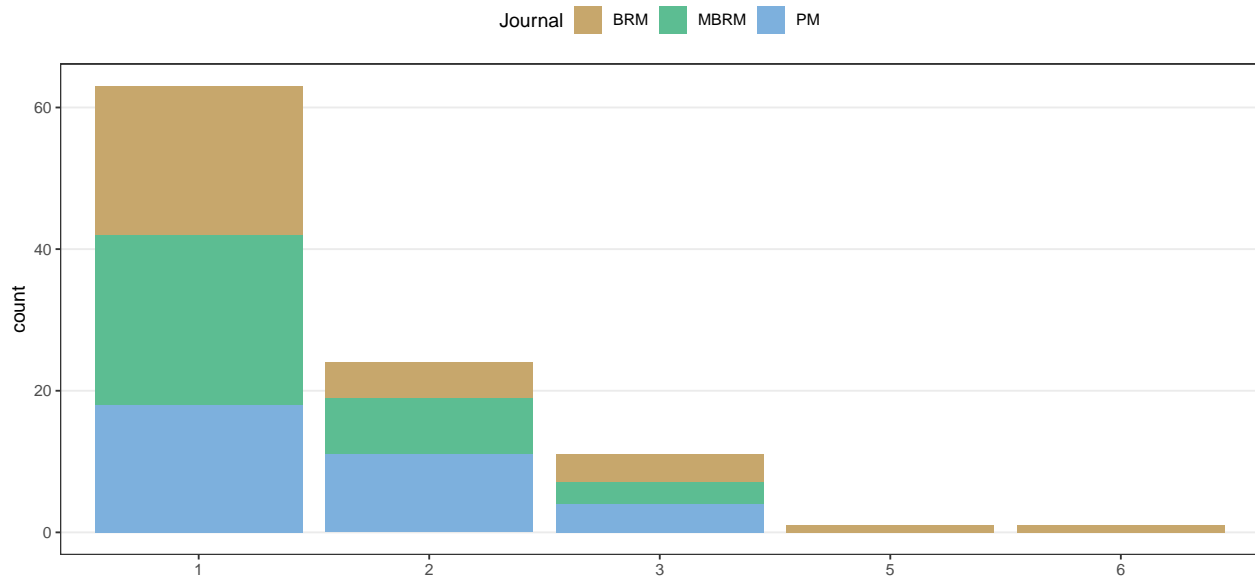


```
## reviewer
ggplot(data = sim_res_fac, aes(x = reviewer)) +
  geom_bar() +
  labs(x = NULL, title = "Reviewer") +
  theme(panel.grid.major.x = element_blank())
```



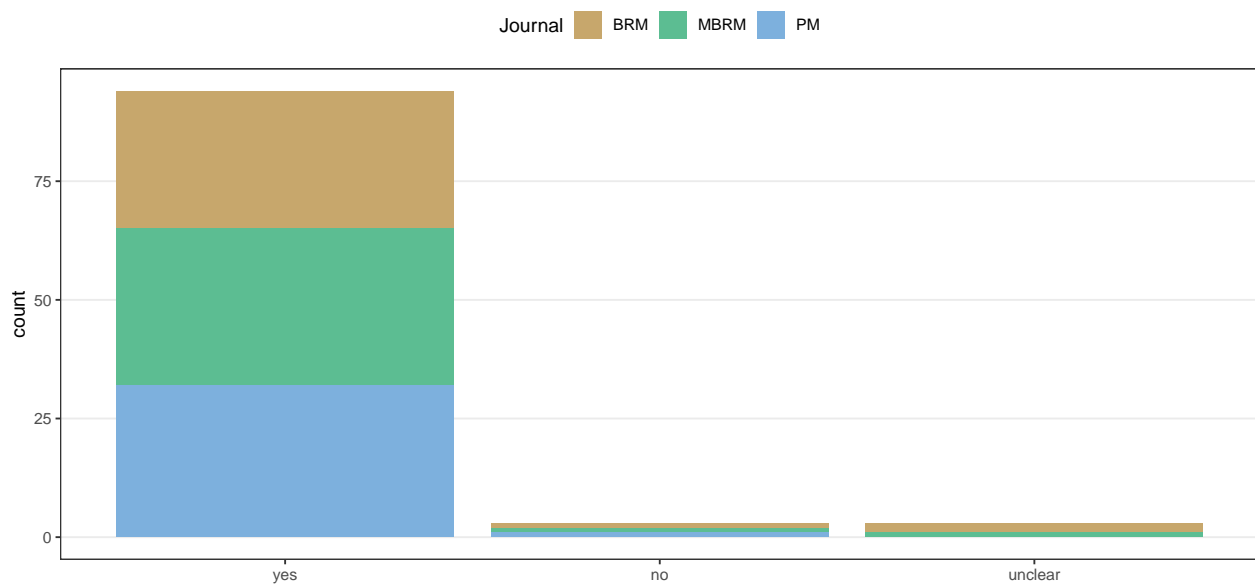
```
## Q2 number of simulation studies
q2 <- ggplot(data = sim_res_fac, aes(x = nsimstudies_q2, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Number of simulation studies in article", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q2
```

Number of simulation studies in article



```
## Q3 are the aims of the study defined
q3 <- sim_res_fac %>%
  mutate(aimsdefined_q3 = factor(aimsdefined_q3,
                                levels = c("yes", "no", "unclear"))) %>%
ggplot(aes(x = aimsdefined_q3, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Aims of the study defined?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q3
```

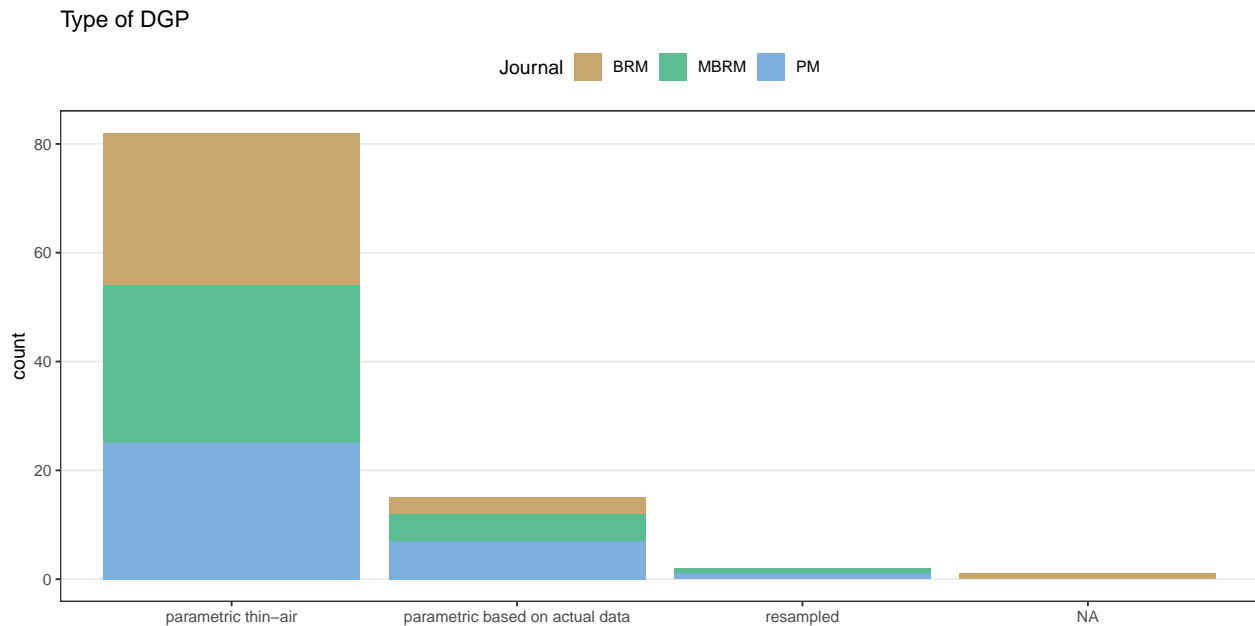
Aims of the study defined?



```
## Q4 type of DGP
q4 <- sim_res_fac %>%
  mutate(dgptype_q4 = as.factor(dgptype_q4)) %>%
```

```
mutate(dgptype_q4 = reorder(dgptype_q4, dgptype_q4, length, decreasing = TRUE)) %>%
  ggplot(aes(x = dgptype_q4, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Type of DGP", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
```

q4



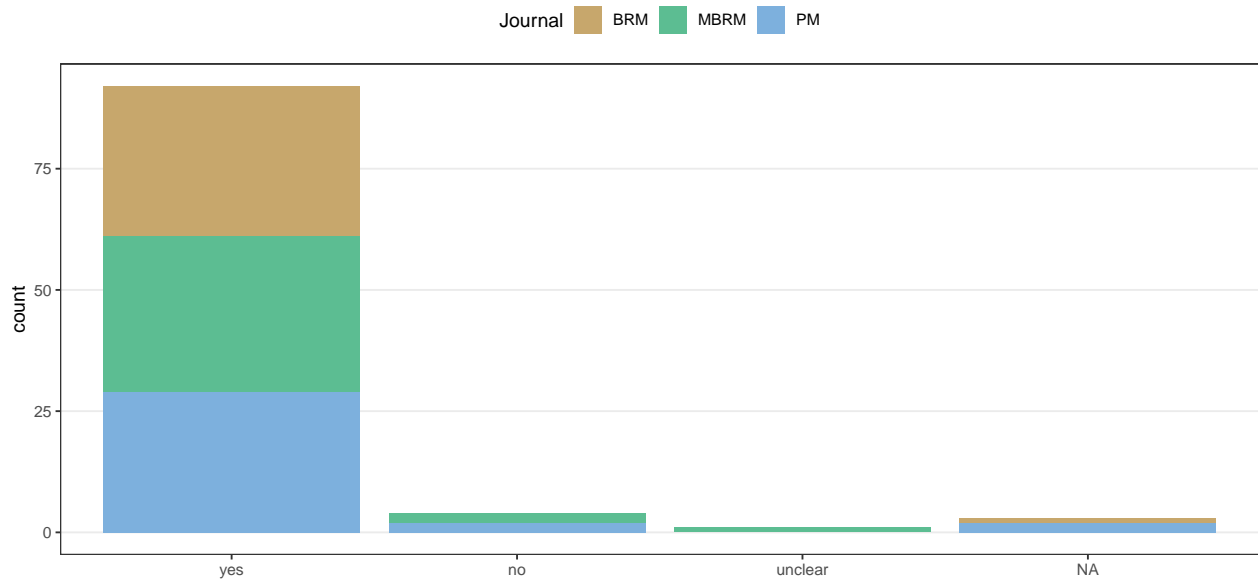
*## from Morris: "97 simulation studies used some form of parametric model to generate data while three used resampling methods. Of the 97 that simulated from a parametric model, 27 based parameter values on data, one based parameter values partly on data, and the remaining 69 on no data. Of these 97, 91 (94%) provided the parameters used." (p. 2079)*

*## Q5 DGP parameters provided?*

```
q5 <- sim_res_fac %>%
  mutate(dgpparameters_q5 = factor(dgpparameters_q5,
    levels = c("yes", "no", "unclear"))) %>%
  ggplot(aes(x = dgpparameters_q5, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Are DGP parameters provided?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
```

q5

Are DGP parameters provided?



## Q6 How many conditions?

```
summary(sim_res_num$nconds_q6)
```

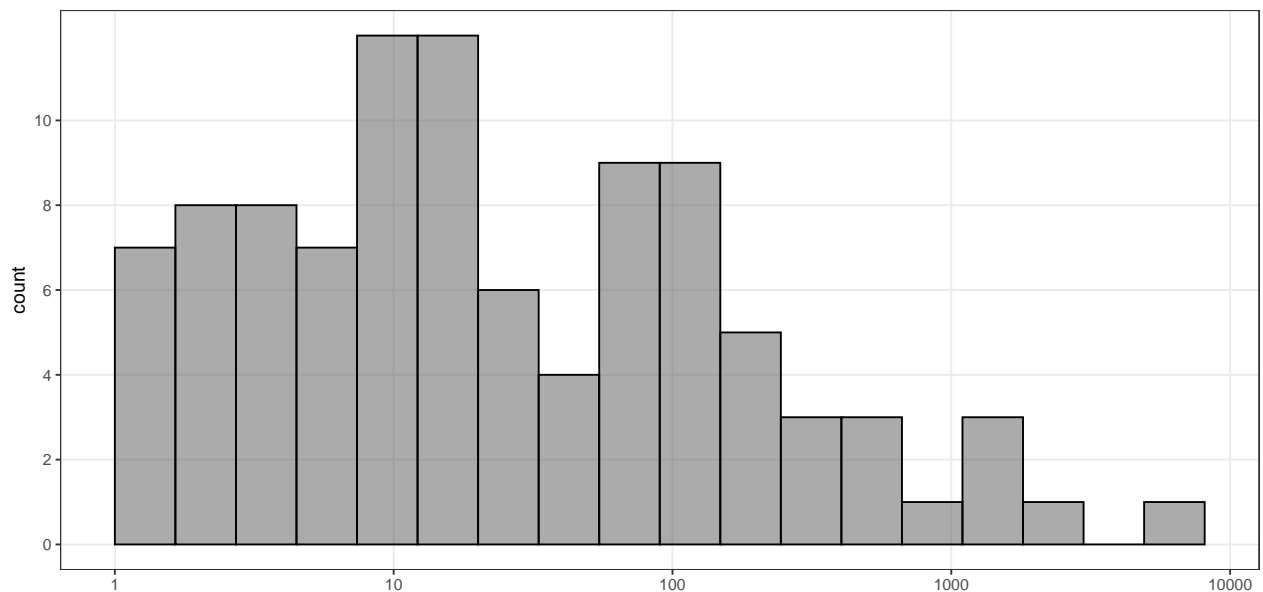
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      1.0     5.0     16.0   185.8   96.0   6000.0         1
```

```
breaks <- c(1, 10, 100, 1000, 10000)
```

```
q6 <- ggplot(data = sim_res_num, aes(x = log(nconds_q6))) +
  geom_histogram(breaks = seq(0, log(10000), 0.5), col = 1, alpha = 0.5) +
  scale_x_continuous(breaks = log(breaks), labels = breaks) +
  scale_y_continuous(breaks = seq(0, 10, 2)) +
  labs(x = NULL, title = "How many simulation conditions?", fill = "Journal")
q6
```

## Warning: Removed 1 rows containing non-finite values (`stat\_bin()`).

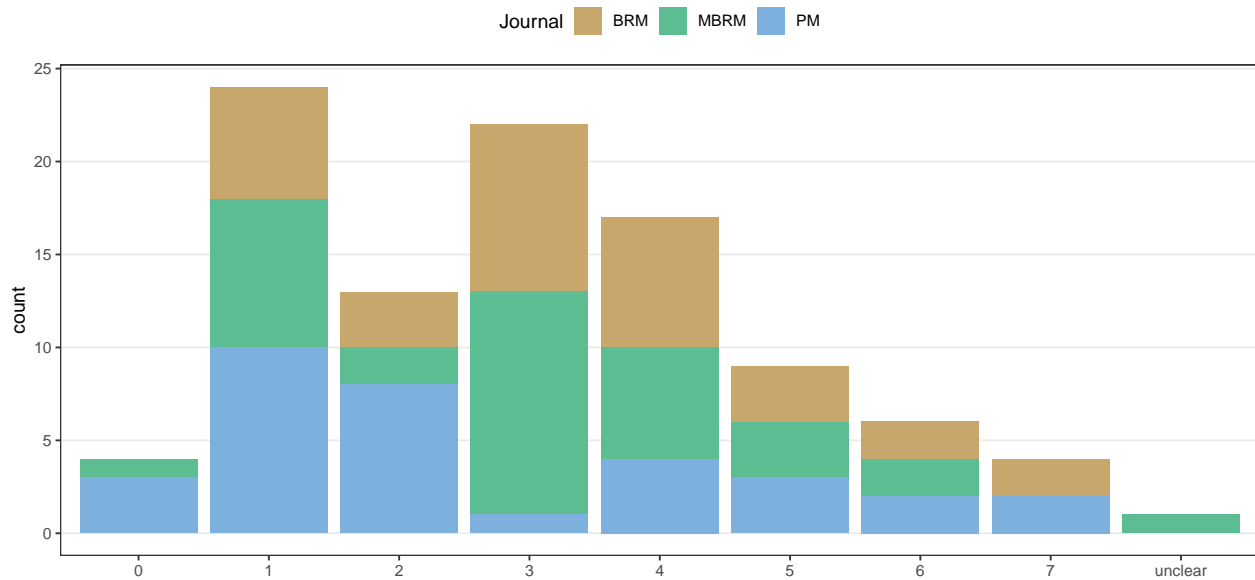
How many simulation conditions?



```
## Q7 How many factors?
q7a <- sim_res_num %>%
  mutate(factorsvaried_q7_fac = ifelse(is.na(factorsvaried_q7),
                                       "unclear", factorsvaried_q7)) %>%

  ggplot(aes(x = factorsvaried_q7_fac, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "How many factors varied?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q7a
```

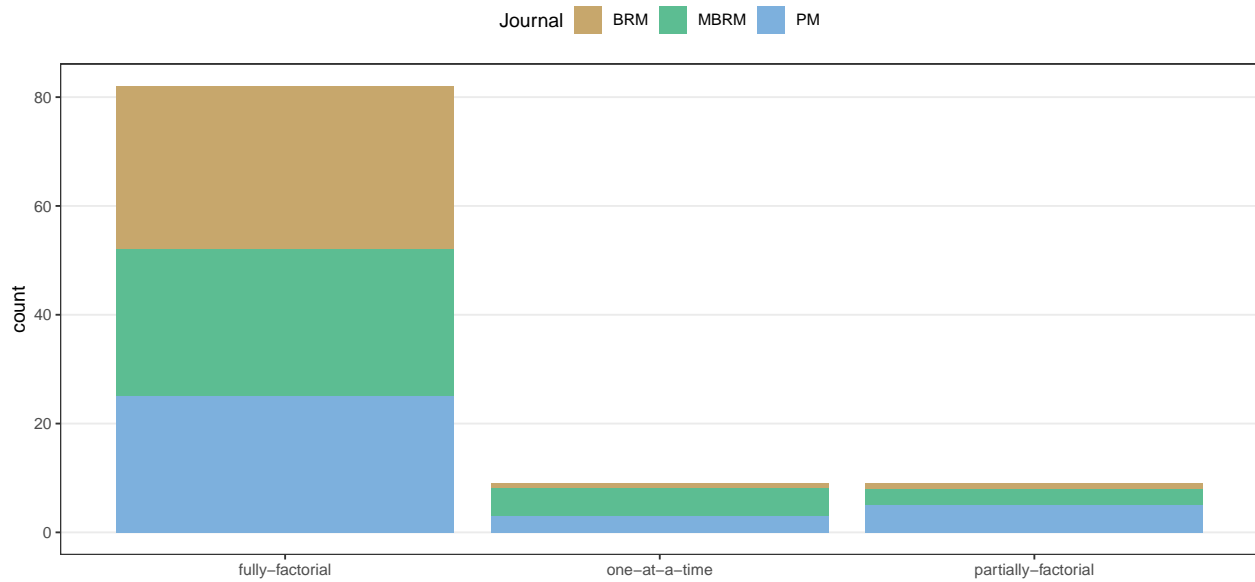
How many factors varied?



```
## Q7 Fully factorial?
q7b <- ggplot(data = sim_res_fac, aes(x = dgmfactorial_q7, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "How are factors varied?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q7b
```



How are factors varied?



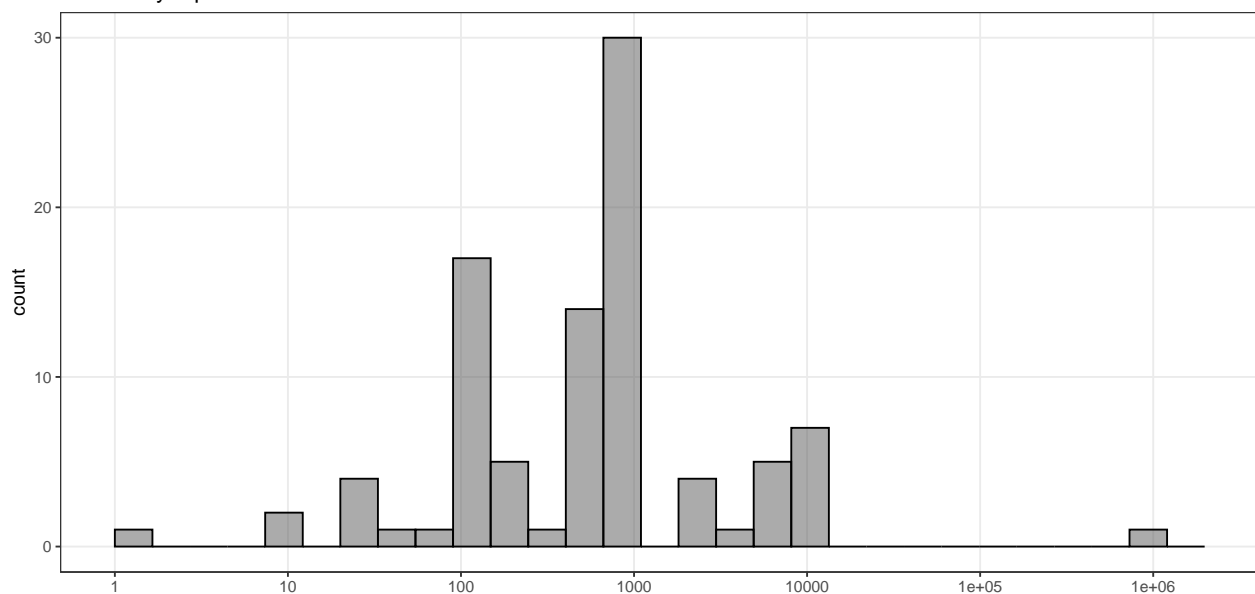
```
## Q8 How many repetitions?
summary(sim_res_num$nsim_q8)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##         1     100     900   12198    1000 1000000         6
```

```
breaks <- c(1, 10, 100, 1000, 10000, 100000, 1000000)
q8 <- ggplot(data = sim_res_num, aes(x = log(nsim_q8))) +
  geom_histogram(breaks = seq(0, log(2000000), 0.5), col = 1, alpha = 0.5) +
  labs(x = NULL, title = "How many repetitions?", fill = "Journal") +
  scale_x_continuous(breaks = log(breaks), labels = breaks)
q8
```

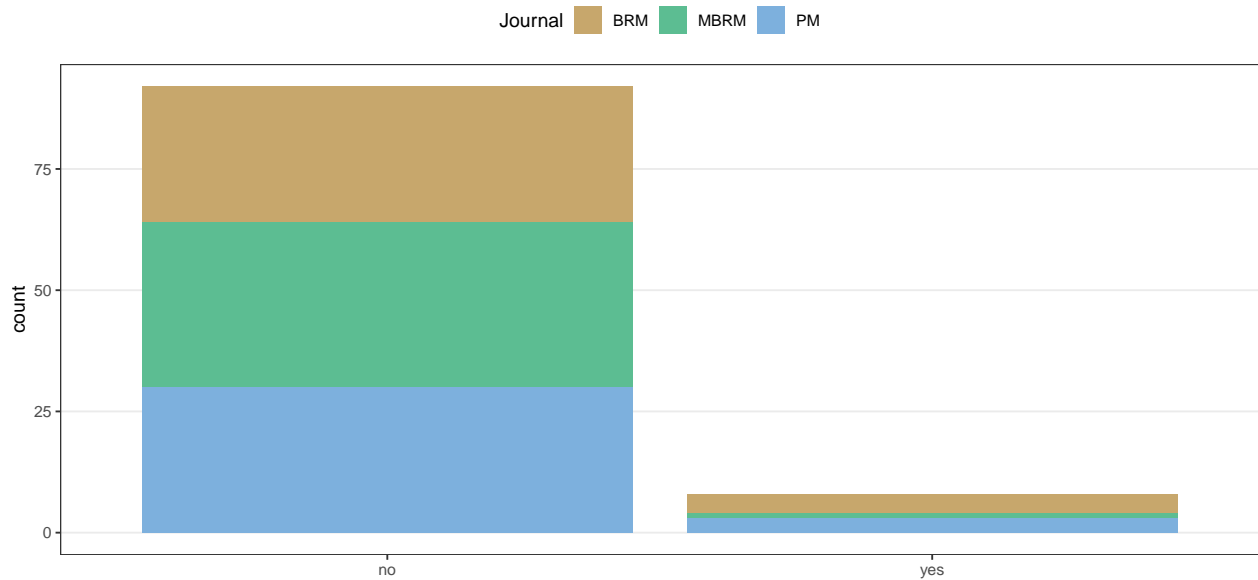
```
## Warning: Removed 6 rows containing non-finite values (`stat_bin()`).
```

How many repetitions?



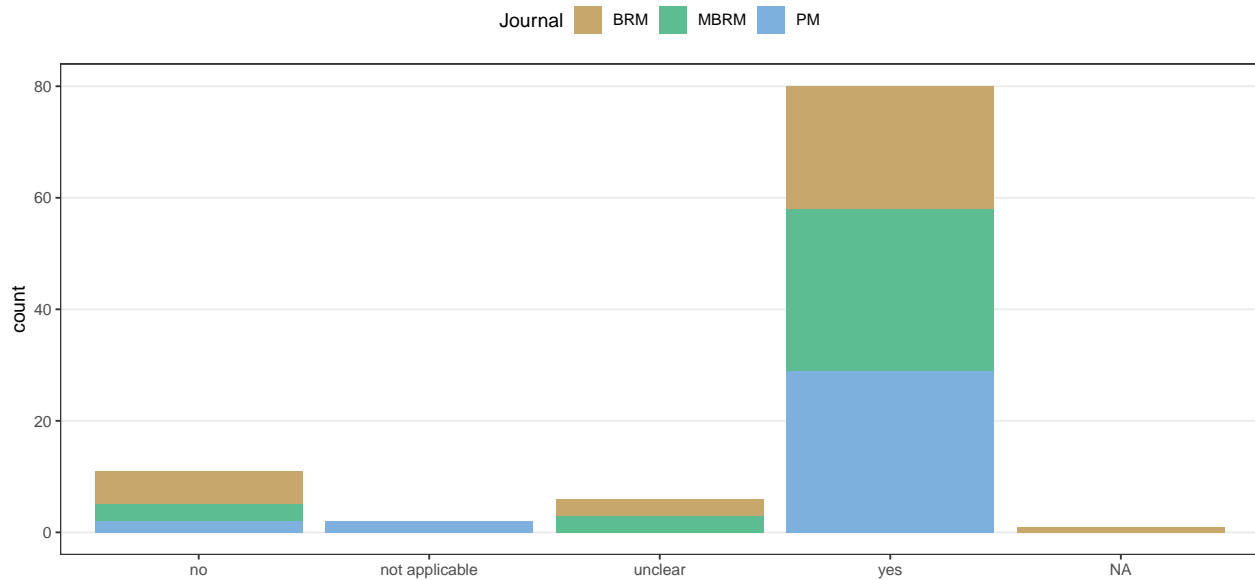
```
## Q9 Are the number of repetitions justified?
q9 <- ggplot(data = sim_res_fac, aes(x = nsimjustified_q9, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Are the number of repetitions justified?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q9
```

Are the number of repetitions justified?



```
## Q10 Is the estimand stated?
q10 <- ggplot(data = sim_res_fac, aes(x = estimandstated_q10, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Is the estimand stated?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q10
```

Is the estimand stated?



## Q11 How many estimands?

```
summary(sim_res_num$nestimands_q11)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      1.00   2.00   4.00  20.11  15.00  384.00      11
```

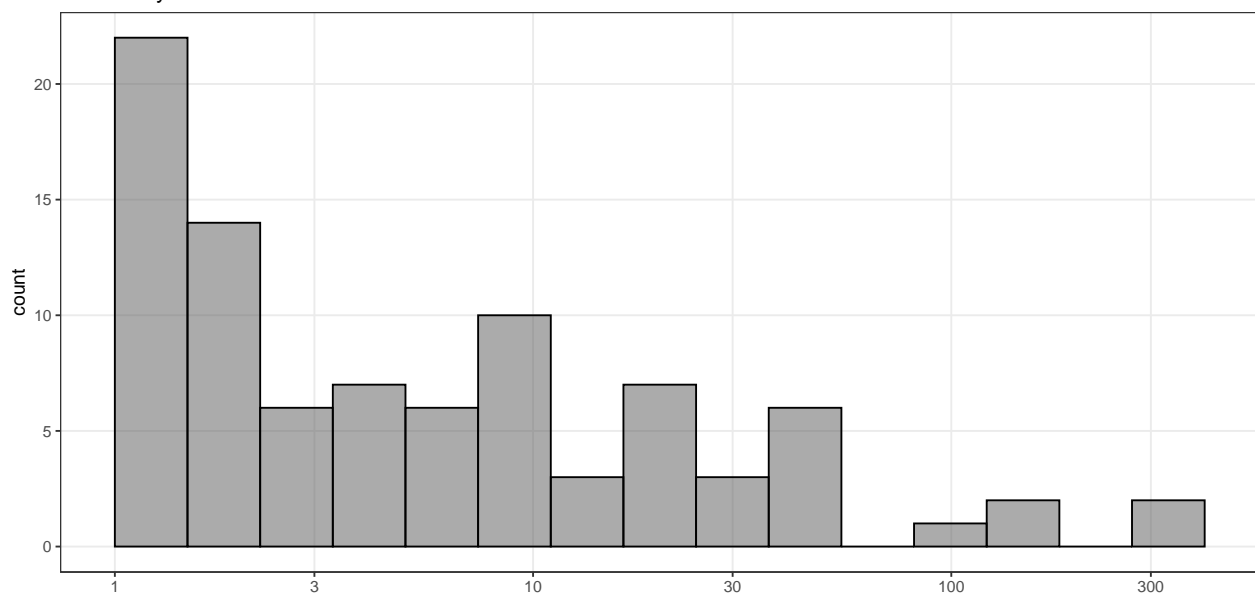
```
breaks <- c(1, 3, 10, 30, 100, 300)
```

```
q11 <- ggplot(data = sim_res_num, aes(x = log(nestimands_q11))) +
  geom_histogram(breaks = seq(0, log(500), 0.4), col = 1, alpha = 0.5) +
  scale_x_continuous(breaks = log(breaks), labels = breaks) +
  labs(x = NULL, title = "How many estimands?", fill = "Journal")
```

```
q11
```

## Warning: Removed 11 rows containing non-finite values (`stat\_bin()`).

How many estimands?

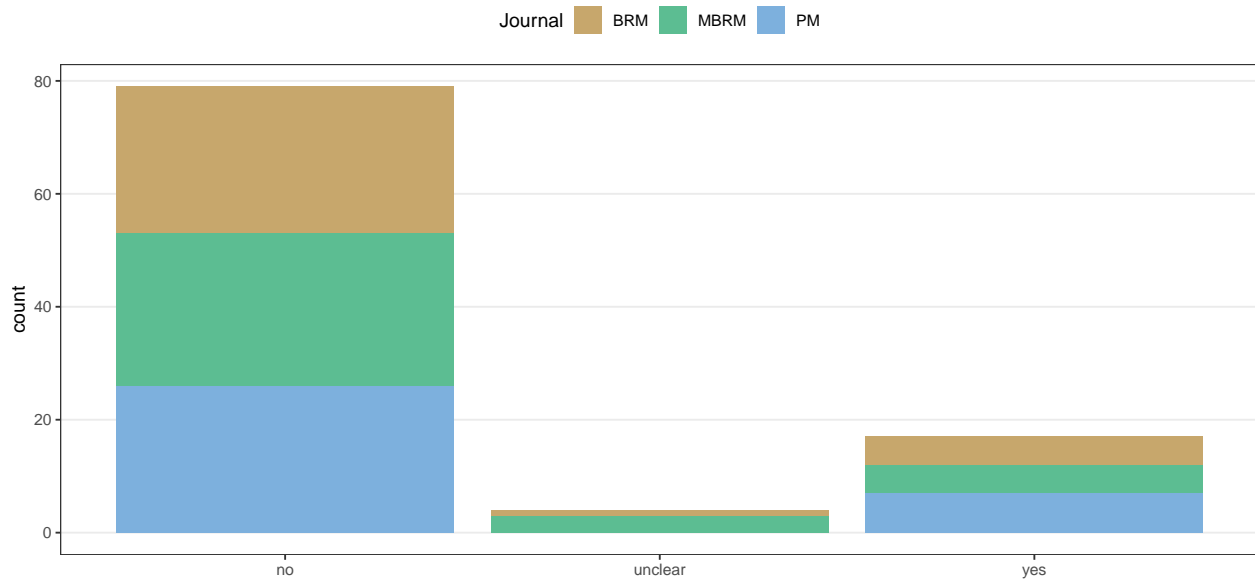


```
## Q12 Are estimands aggregated?
```

```
q12 <- ggplot(data = sim_res_fac, aes(x = estimandsagg_q12, fill = journal)) +  
  geom_bar() +  
  labs(x = NULL, title = "Are estimands aggregated?", fill = "Journal") +  
  scale_fill_discrete_qualitative(palette = pal) +  
  theme(panel.grid.major.x = element_blank())
```

q12

Are estimands aggregated?

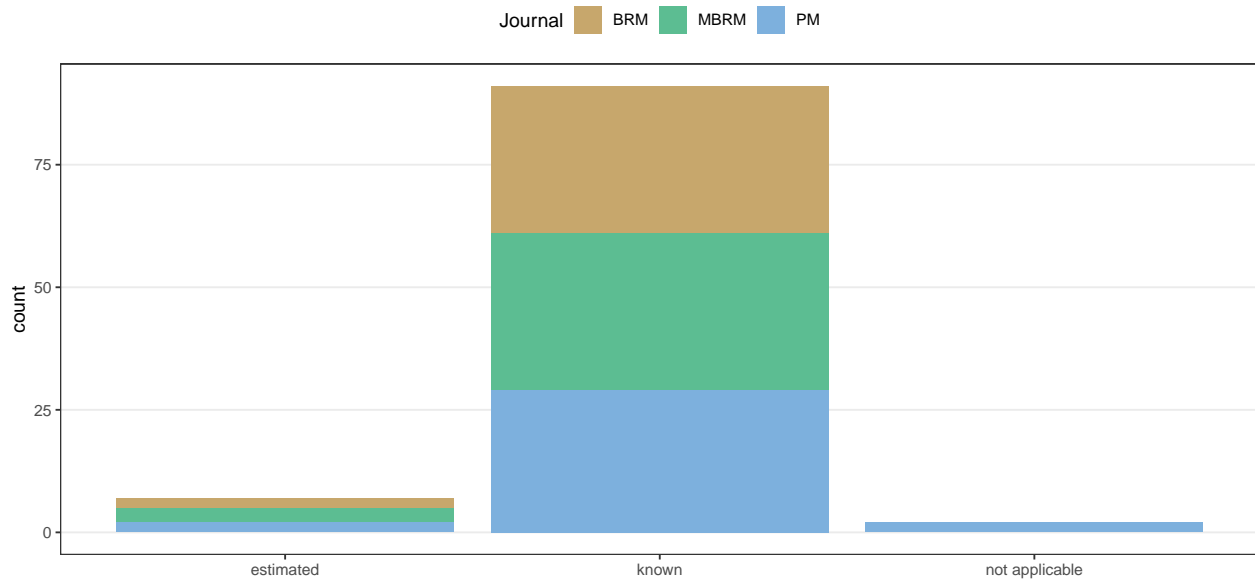


```
## Q13 How are the true parameters specified?
```

```
q13 <- ggplot(data = sim_res_fac, aes(x = truetheta_q13, fill = journal)) +  
  geom_bar() +  
  labs(x = NULL, title = "How are the true parameters specified?", fill = "Journal") +  
  scale_fill_discrete_qualitative(palette = pal) +  
  theme(panel.grid.major.x = element_blank())
```

q13

How are the true parameters specified?



*## Q14 How many methods are included?*

summary(sim\_res\_num\$nmethods\_q14)

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   2.00   3.00   5.63   5.00  192.00
```

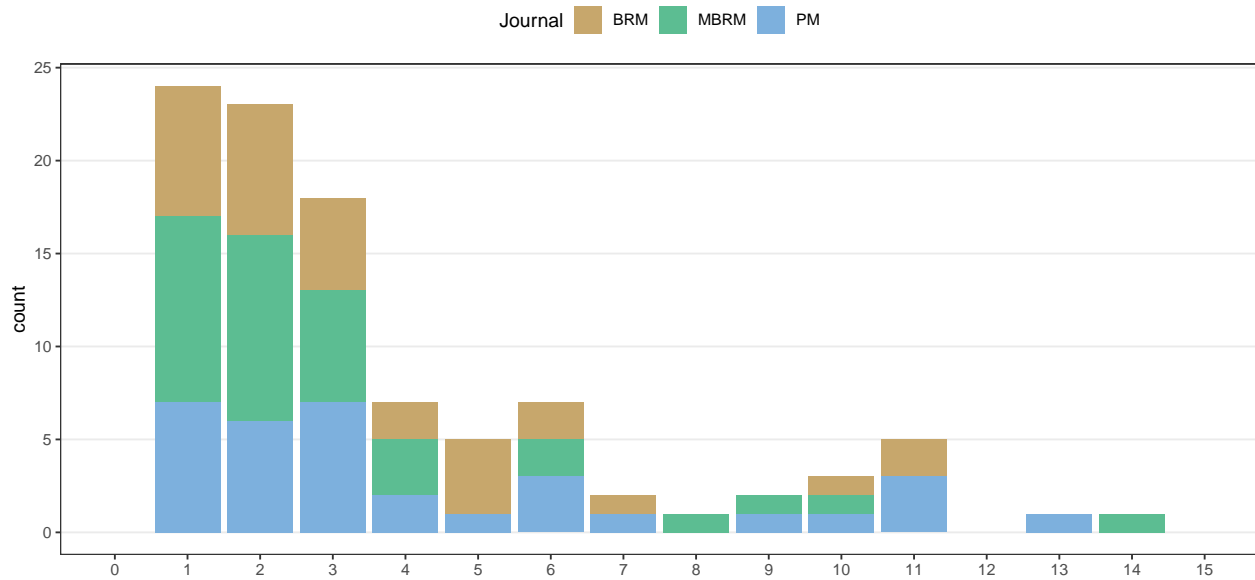
*# HACK there is one study with 192 methods, let's exclude it for a moment*

```
q14 <- ggplot(data = sim_res_num, aes(x = nmethods_q14, fill = journal)) +
  geom_bar() +
  scale_x_continuous(breaks = seq(0, 15), limits = c(0, 15)) +
  labs(x = NULL, title = "How many methods are included?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
```

q14

## Warning: Removed 1 rows containing non-finite values (`stat\_count()`).

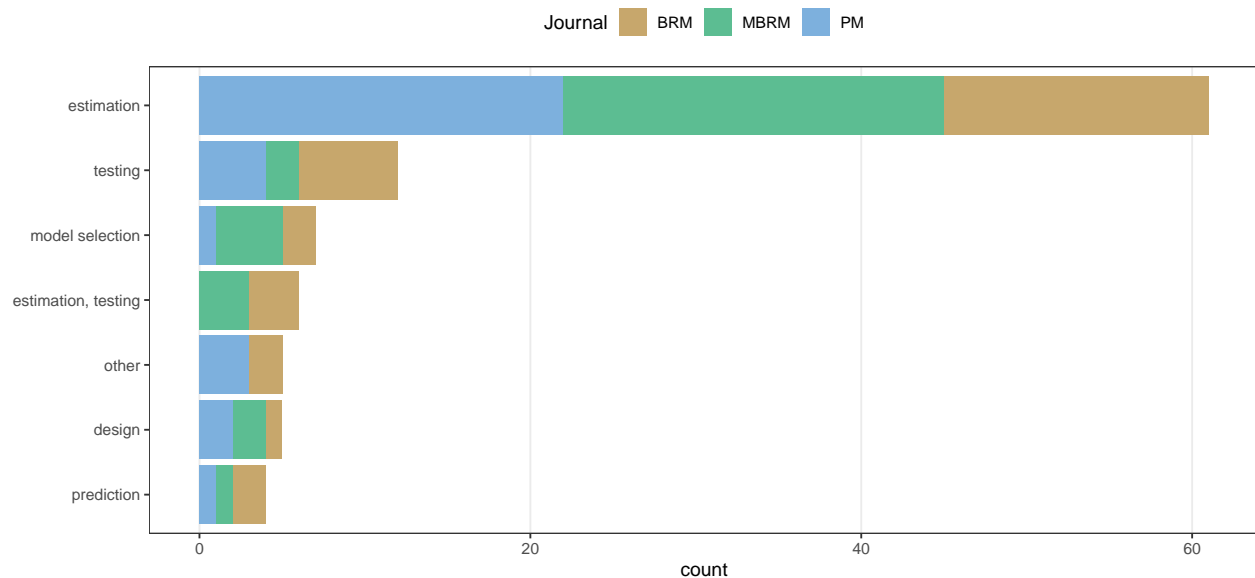
How many methods are included?



## Q15 What is the evaluation target of the simulation?

```
q15a <- sim_res_fac %>%
  mutate(target_q15 = as.factor(target_q15)) %>%
  mutate(target_q15 = reorder(target_q15, target_q15, length)) %>%
  ggplot(aes(x = target_q15, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "What is the evaluation target of the simulation?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.y = element_blank()) +
  coord_flip()
q15a
```

What is the evaluation target of the simulation?



## Q15 Which performance measures were used?

# Spread "Other" apart

# TODO add to the visualization after deciding on treatment of bias

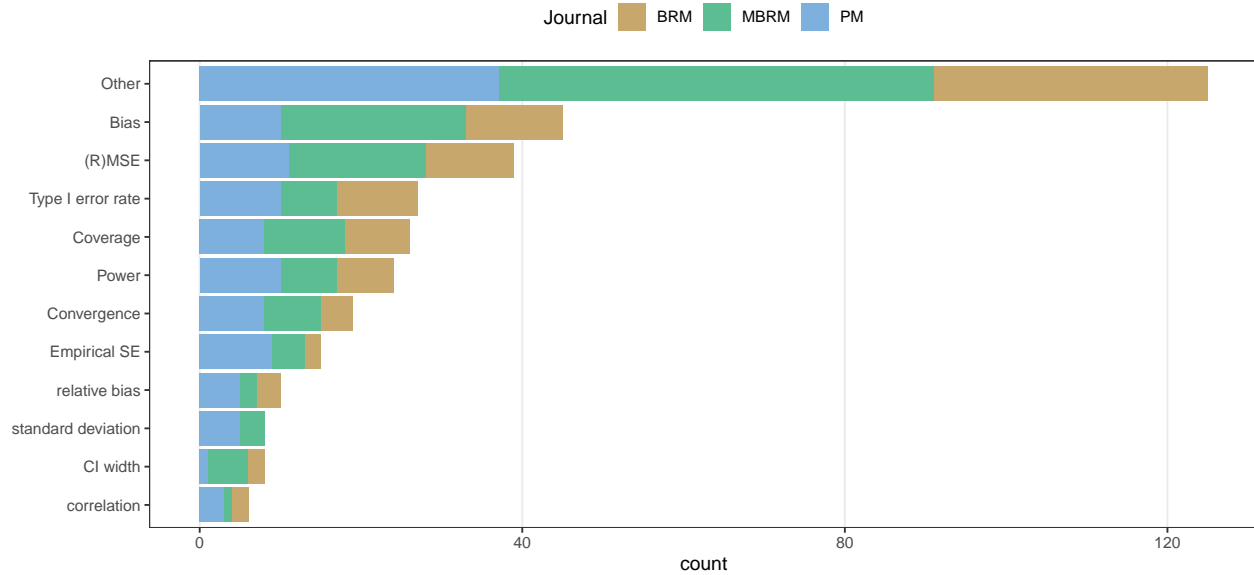
```

q15_other <- sim_res_fac %>%
  separate_wider_delim(pmother_q15,
    delim = ",",
    names_sep = "_",
    too_few = "align_start") %>%
  pivot_longer(cols = contains("pmother"),
    names_to = NULL,
    values_to = "pmother",
    values_drop_na = TRUE) %>%
  select(pmother, journal) %>%
  # remove whitespace
  mutate(pmother = str_trim(pmother)) %>%
  mutate(pmother = str_replace(pmother, ".*correlation.*", "correlation")) %>%
  mutate(pmother = str_replace(pmother, ".*standard deviation.*", "standard deviation")) %>%
  mutate(pmother = as.factor(pmother)) %>%
  mutate(pmother = forcats::fct_lump_n(pmother, 3)) %>%
  group_by(journal) %>%
  count(pmother) %>%
  rename(PM = pmother,
    count = n)

# Visualize
q15b <- sim_res_fac %>%
  group_by(journal) %>%
  summarise("Convergence" = sum(pmconvergence_q15 == "yes"),
    "Bias" = sum(pmbias_q15 == "yes"),
    "Empirical SE" = sum(pmempse_q15 == "yes"),
    "(R)MSE" = sum(pm_r_mse_q15 == "yes"),
    "Coverage" = sum(pmcover_q15 == "yes"),
    "Type I error rate" = sum(pmtypeierror_q15 == "yes"),
    "Power" = sum(pmpower_q15 == "yes"),
    "CI width" = sum(pmciwidth_q15 == "yes"),
    "Other" = sum(!is.na(pmother_q15))) %>%
  gather(key = "PM", value = "count", "Convergence", "Bias", "(R)MSE",
    "Empirical SE", "Coverage", "Type I error rate",
    "Power", "CI width", "Other") %>%
  bind_rows(q15_other) %>%
  mutate(PM = as.factor(PM)) %>%
  mutate(PM = reorder(PM, count, sum)) %>%
  ggplot(aes(x = PM, y = count, fill = journal)) +
  geom_bar(stat = "identity") +
  labs(x = NULL, title = "Performance measure", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.y = element_blank()) +
  coord_flip()
q15b

```

## Performance measure



*# Double check if absolute/relative bias ever occur with bias rated as "no"*

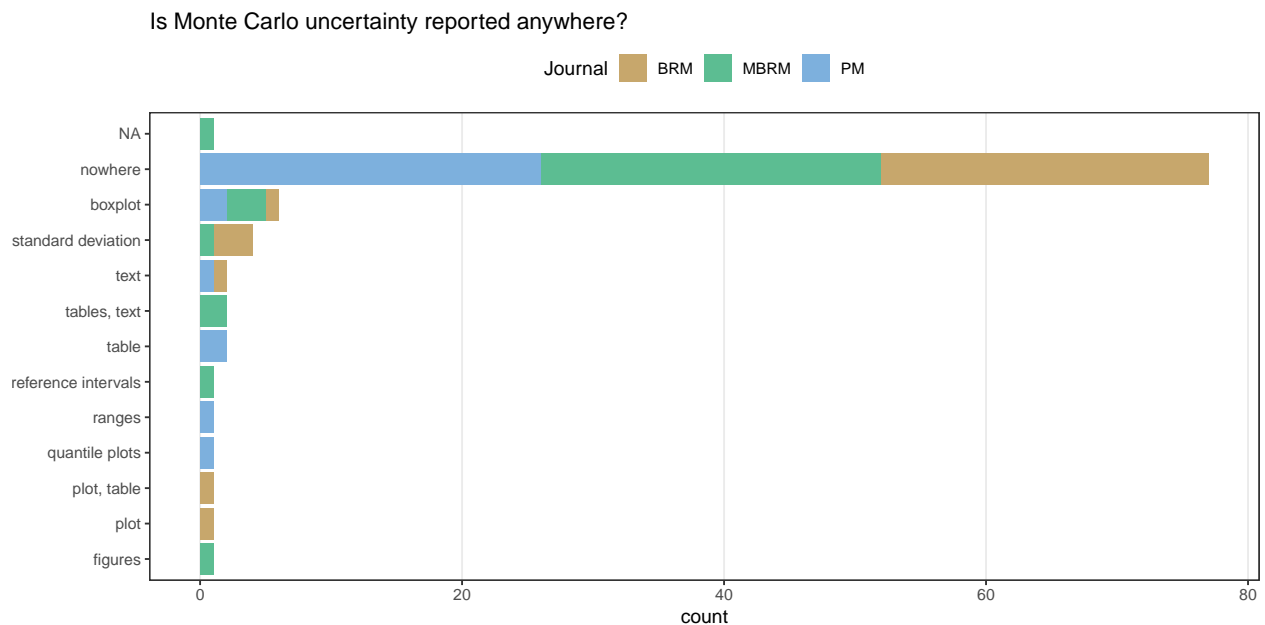
```
sim_res_fac %>%
  separate_wider_delim(pmother_q15,
                        delim = ",",
                        names_sep = "_",
                        too_few = "align_start") %>%
  pivot_longer(cols = contains("pmother"),
               names_to = NULL,
               values_to = "pmother",
               values_drop_na = TRUE) %>%
  mutate(pmother = str_trim(pmother)) %>%
  # mutate(pmother = str_replace(pmother, ".*bias.*", "bias")) %>%
  filter(grepl("bias", pmother)) %>%
  select(reviewer, pmbias_q15, pmother)
```

```
## # A tibble: 18 x 3
##   reviewer pmbias_q15 pmother
##   <fct>    <fct>      <chr>
## 1 FB      yes      absolute bias
## 2 FB      yes      relative bias
## 3 FB      yes      relative bias
## 4 FB      no       absolute bias
## 5 FB      yes      bias of standard errors
## 6 FB      no       relative bias
## 7 FB      yes      relative bias
## 8 FB      no       relative bias
## 9 FB      no       relative bias of standard errors
## 10 FB     no       relative bias
## 11 FB     no       relative bias
## 12 FB     no       relative bias
## 13 FB     no       relative bias of se
## 14 FB     no       relative bias
## 15 FB     no       absolute relative bias
## 16 FB     yes      relative bias of se
```



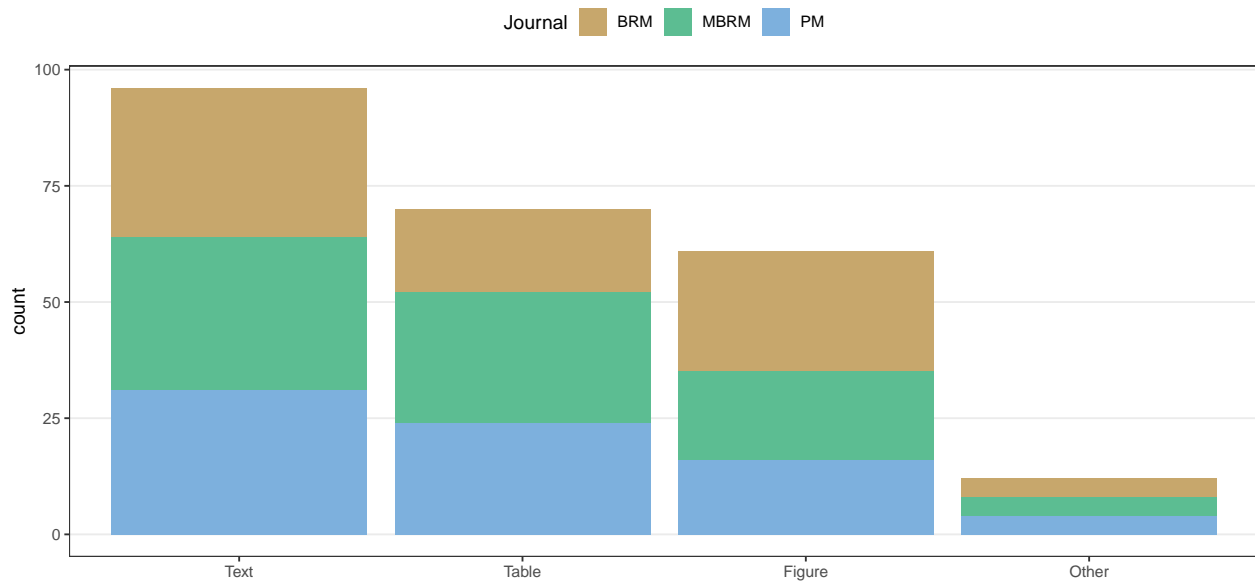
```
## 17 FB      no      relative bias
## 18 BS      yes     SD of SE bias (as uncertainty)

## Q16 Is Monte Carlo uncertainty reported anywhere?
q16 <- sim_res_fac %>%
  mutate(merrors_q16 = as.factor(merrors_q16)) %>%
  mutate(merrors_q16 = reorder(merrors_q16, merrors_q16, length)) %>%
  ggplot(aes(x = merrors_q16, fill = journal)) +
    geom_bar() +
    labs(x = NULL, title = "Is Monte Carlo uncertainty reported anywhere?", fill = "Journal") +
    scale_fill_discrete_qualitative(palette = pal) +
    theme(panel.grid.major.y = element_blank()) +
    coord_flip()
q16
```



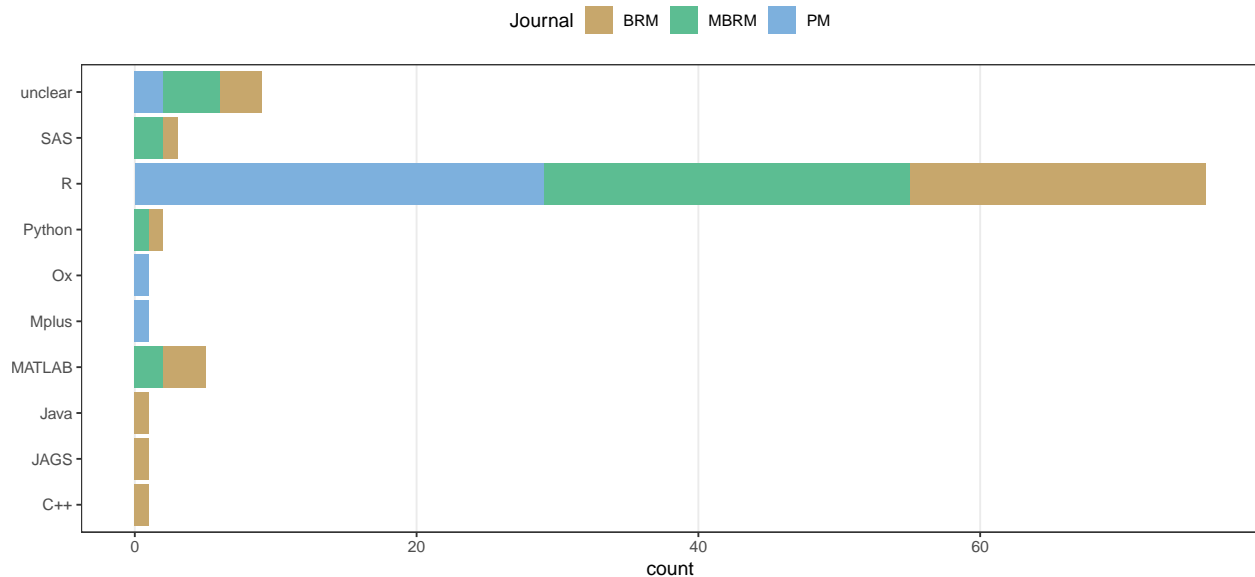
```
## Q17 In which way are the results reported?
q17 <- sim_res_fac %>%
  group_by(journal) %>%
  summarise("Figure" = sum(resultsfigure_q17 == "yes"),
            "Table" = sum(resultstable_q17 == "yes"),
            "Text" = sum(resultstext_q17 == "yes"),
            "Other" = sum(resultsother_q17 == "yes")) %>%
  gather(key = "Type", value = "count", "Figure", "Table", "Text", "Other") %>%
  mutate(Type = as.factor(Type)) %>%
  mutate(Type = reorder(Type, count, sum, decreasing = TRUE)) %>%
  ggplot(aes(x = Type, y = count, fill = journal)) +
    geom_bar(stat = "identity") +
    labs(x = NULL, title = "In which way are the results reported?", fill = "Journal") +
    scale_fill_discrete_qualitative(palette = pal) +
    theme(panel.grid.major.x = element_blank())
q17
```

In which way are the results reported?



```
## Q18 Which software was used to conduct the simulation?
q18a <- ggplot(data = sim_res_fac, aes(x = software_1_q18, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Which primary software was used?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.y = element_blank()) +
  coord_flip()
q18a
```

Which primary software was used?

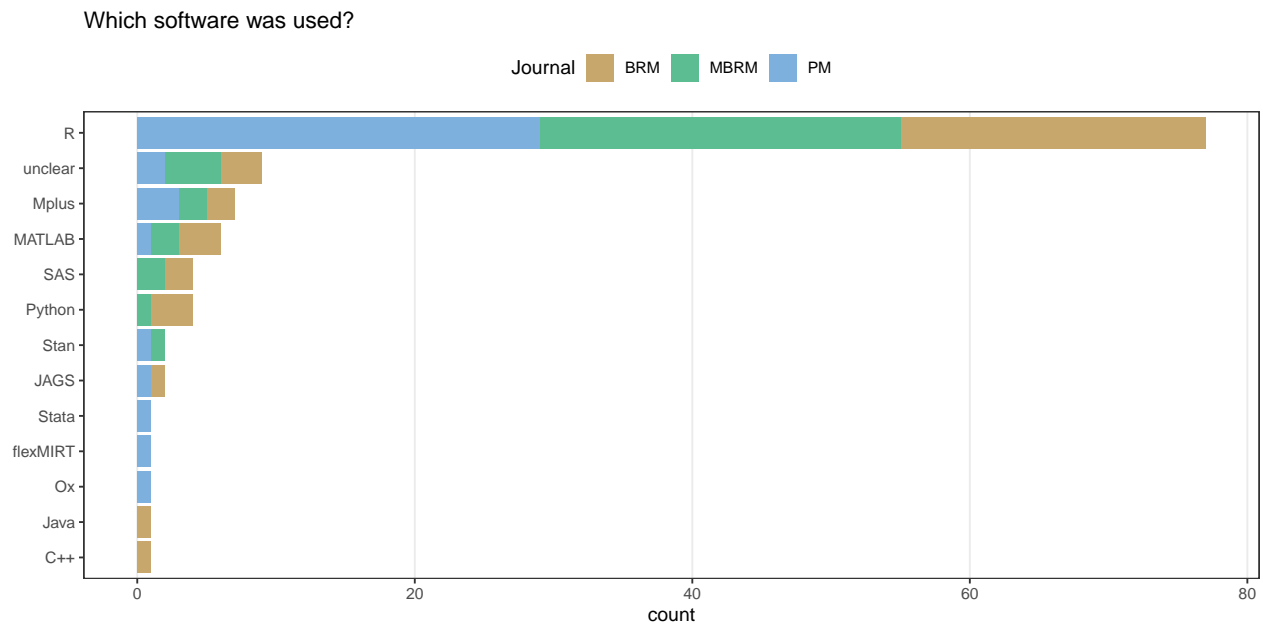


```
# add information from software_2_q18 and software_3_q18
q18b <- sim_res_fac %>%
  select(starts_with("software"), journal) %>%
  pivot_longer(cols = starts_with("software"),
               names_to = NULL,
```

```

      values_to = "software",
      values_drop_na = TRUE) %>%
mutate(software = as.factor(software)) %>%
mutate(software = reorder(software, length)) %>%
ggplot(aes(x = software, fill = journal)) +
geom_bar() +
labs(x = NULL, title = "Which software was used?", fill = "Journal") +
scale_fill_discrete_qualitative(palette = pal) +
theme(panel.grid.major.y = element_blank()) +
coord_flip()
q18b

```

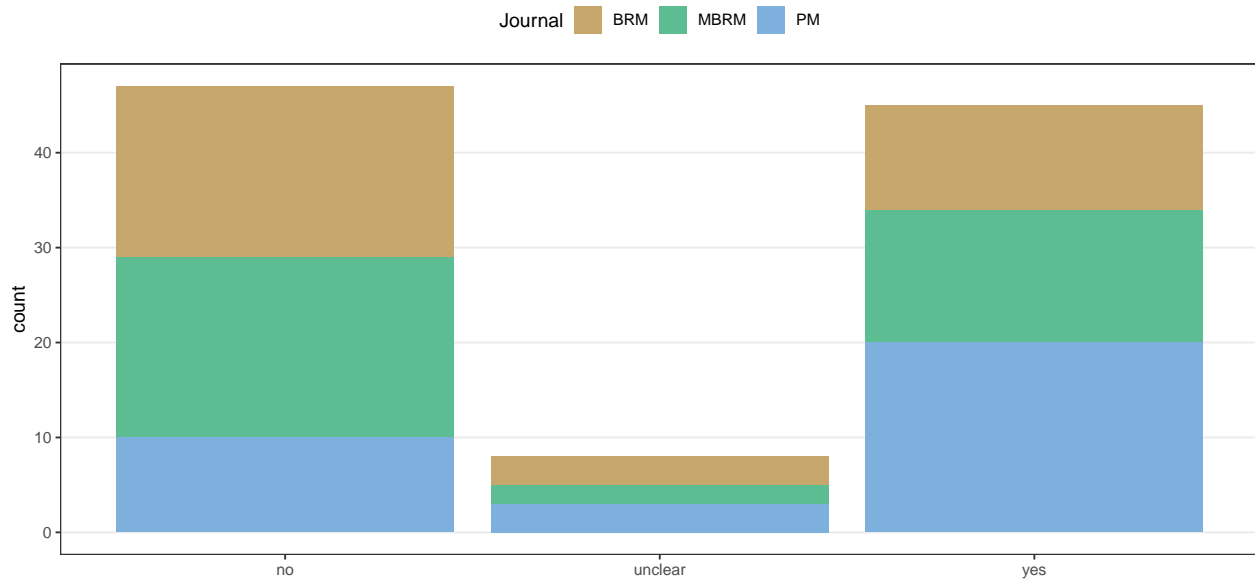


```

## Q19 Are there userwritten commands/packages/macros?
q19 <- ggplot(data = sim_res_fac, aes(x = userwritten_q19, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Are there userwritten commands/packages/macros?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q19

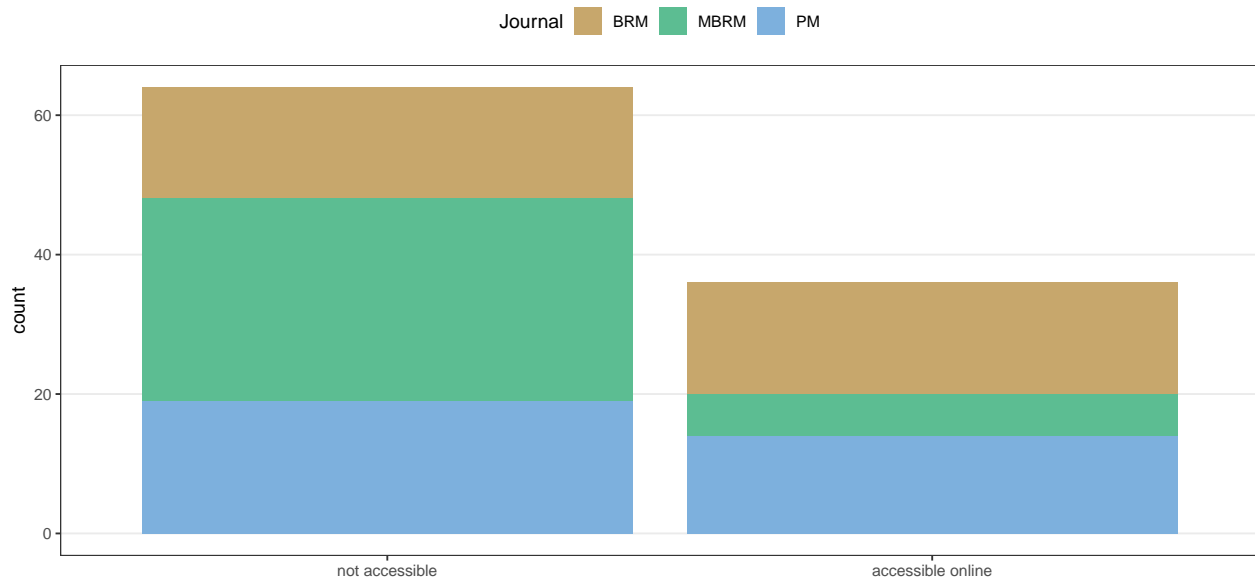
```

Are there userwritten commands/packages/macros?



```
## Q20 Is code provided?
# Reordered this such that negative is on the left to stay consistent
q20 <- sim_res_fac %>%
  mutate(codeprovided_q20 = as.factor(codeprovided_q20)) %>%
  mutate(codeprovided_q20 = reorder(codeprovided_q20, codeprovided_q20,
                                    length, decreasing = TRUE)) %>%
  ggplot(aes(x = codeprovided_q20, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Is code provided?", fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q20
```

Is code provided?

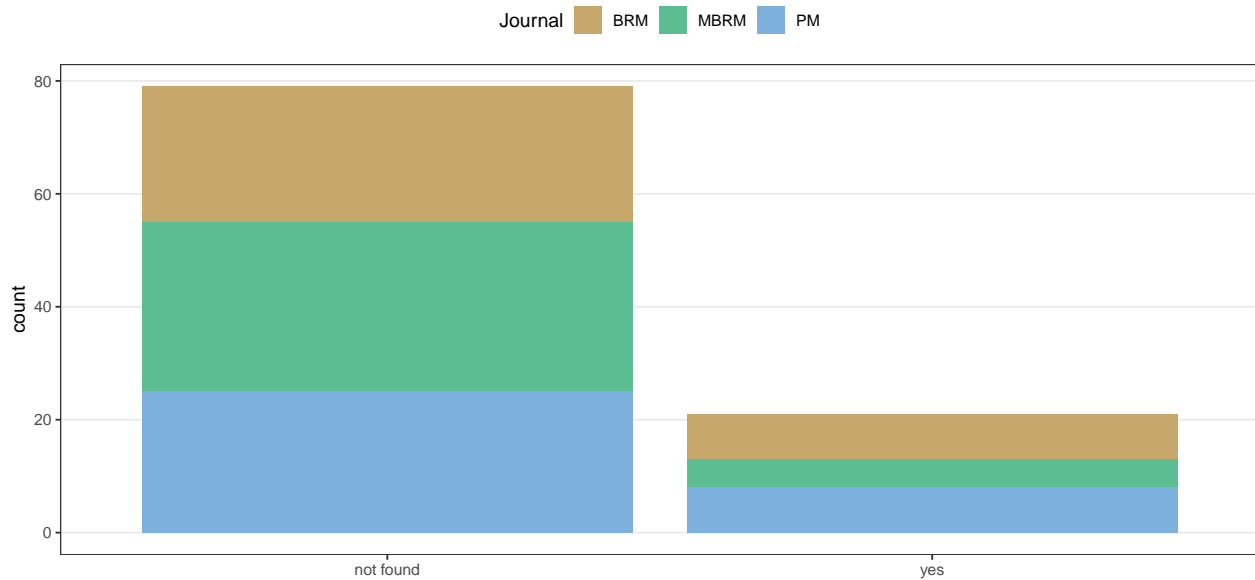


```
## Q21 If code is provided, is a seed provided?
```

```
q21 <- ggplot(data = sim_res_fac, aes(x = seedprovided_q21, fill = journal)) +  
  geom_bar() +  
  labs(x = NULL, title = "If code is provided, is a seed provided?", fill = "Journal") +  
  scale_fill_discrete_qualitative(palette = pal) +  
  theme(panel.grid.major.x = element_blank())
```

q21

If code is provided, is a seed provided?

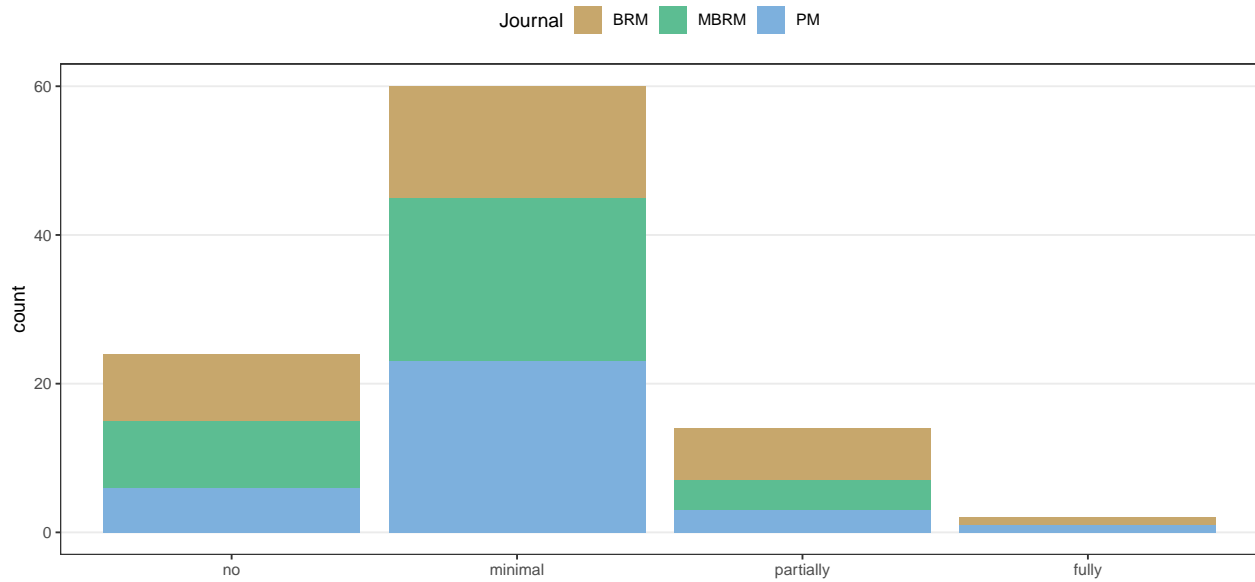


```
## Q22 Is information on the computational environment provided?
```

```
q22 <- sim_res_fac %>%  
  mutate(compenvironment_q22 = factor(compenvironment_q22,  
                                       levels = c("no", "minimal", "partially", "fully"))) %>%  
  ggplot(aes(x = compenvironment_q22, fill = journal)) +  
  geom_bar() +  
  labs(x = NULL, title = "Is information on the computational environment provided?",  
       fill = "Journal") +  
  scale_fill_discrete_qualitative(palette = pal) +  
  theme(panel.grid.major.x = element_blank())
```

q22

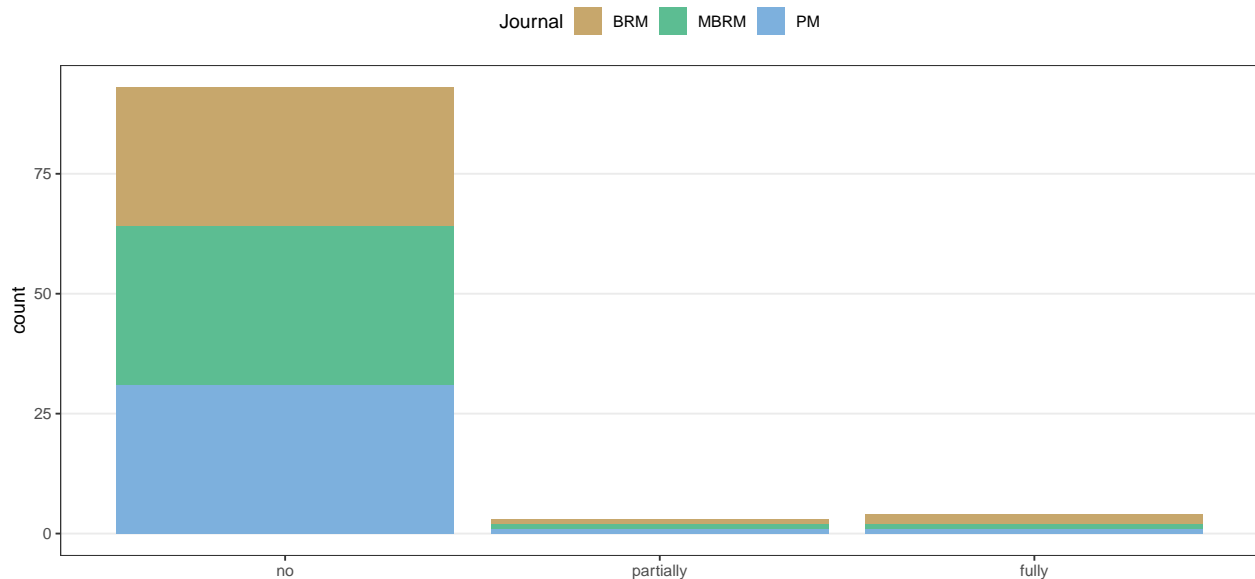
Is information on the computational environment provided?



## Q23 Is information on the operating system provided?

```
q23 <- sim_res_fac %>%
  mutate(compos_q23 = factor(compos_q23,
                             levels = c("no", "partially", "fully"))) %>%
  ggplot(aes(x = compos_q23, fill = journal)) +
  geom_bar() +
  labs(x = NULL, title = "Is information on the operating system provided?",
       fill = "Journal") +
  scale_fill_discrete_qualitative(palette = pal) +
  theme(panel.grid.major.x = element_blank())
q23
```

Is information on the operating system provided?



## Q24 How confident was reviewer in coding of the article?

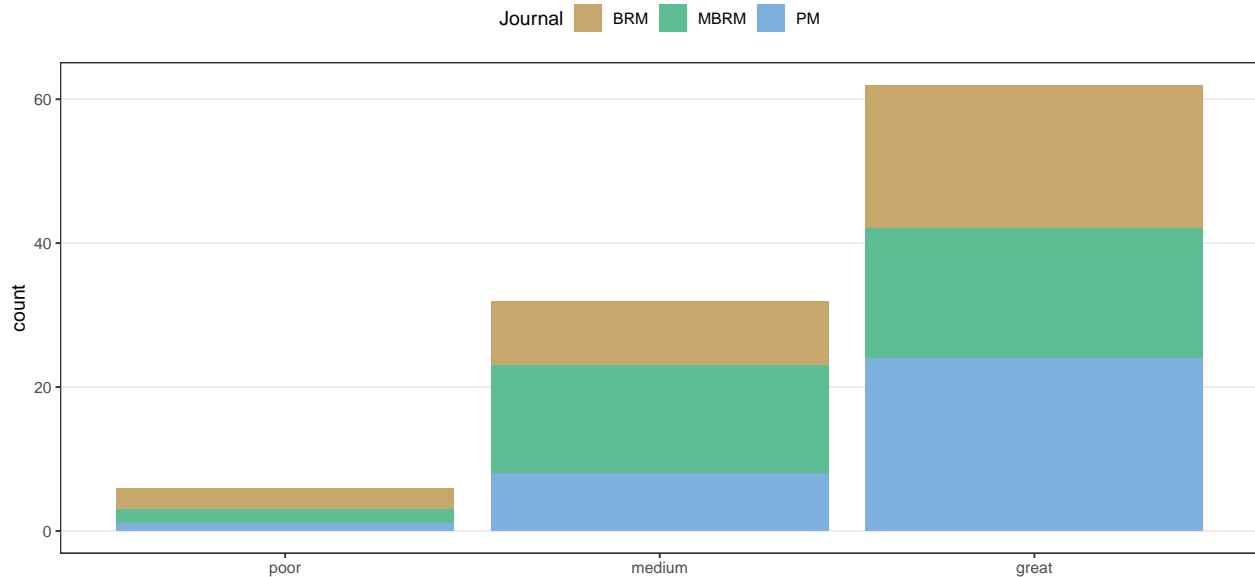
```
q24 <- sim_res_fac %>%
```

```

mutate(coding_confidence = factor(coding_confidence,
                                  levels = c("poor", "medium", "great"))) %>%
ggplot(aes(x = coding_confidence, fill = journal)) +
geom_bar() +
labs(x = NULL, title = "How confident was reviewer in coding of the article?",
     fill = "Journal") +
scale_fill_discrete_qualitative(palette = pal) +
theme(panel.grid.major.x = element_blank())
q24

```

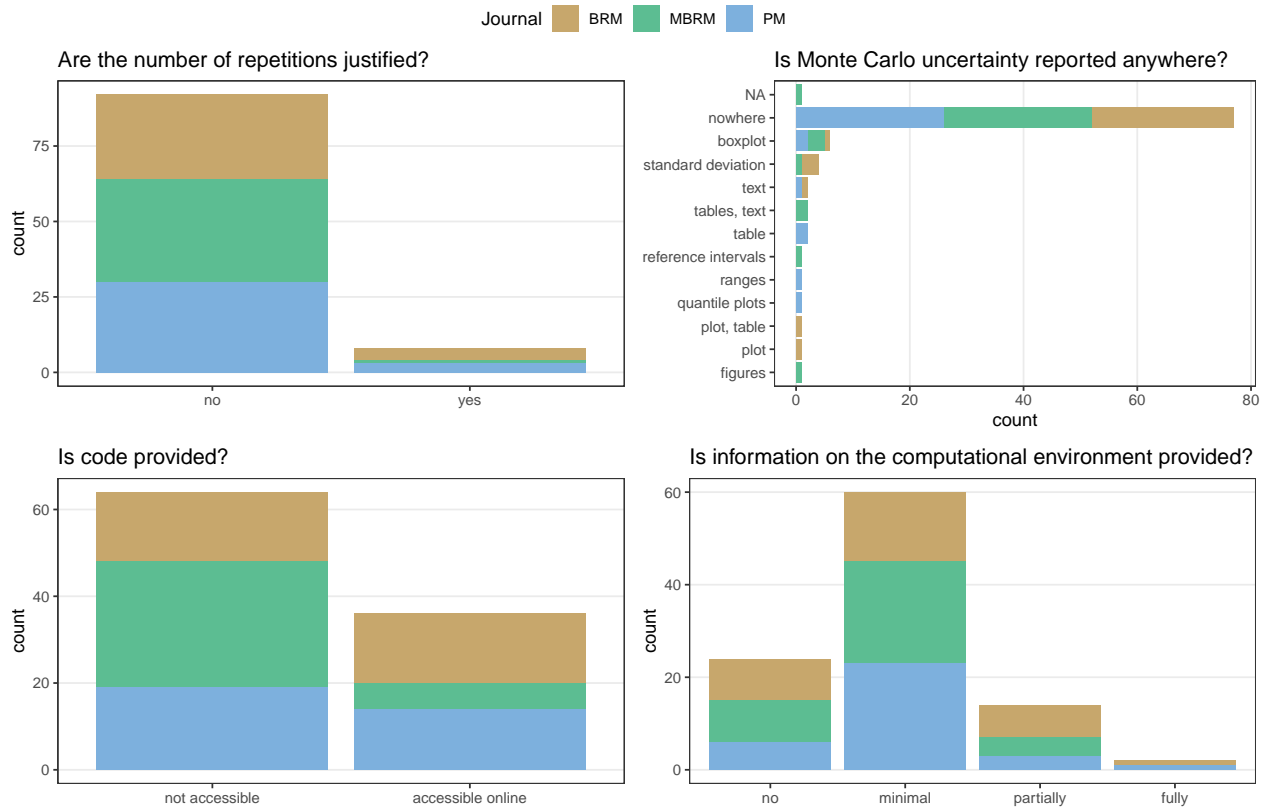
How confident was reviewer in coding of the article?



```

## composite plot 1 - problematic questions
ggpubr::ggarrange(q9, q16, q20, q22, ncol = 2, nrow = 2, common.legend = TRUE,
                  align = "h")

```



```
## composite plot 2 - descriptives
```

```
ggpubr::ggarrange(q2, q4, q6, q7a, q7b, q8, q11, q14, q15a, q15b, q17, q18b,
  ncol = 3, nrow = 4, common.legend = TRUE, align = "h")
```

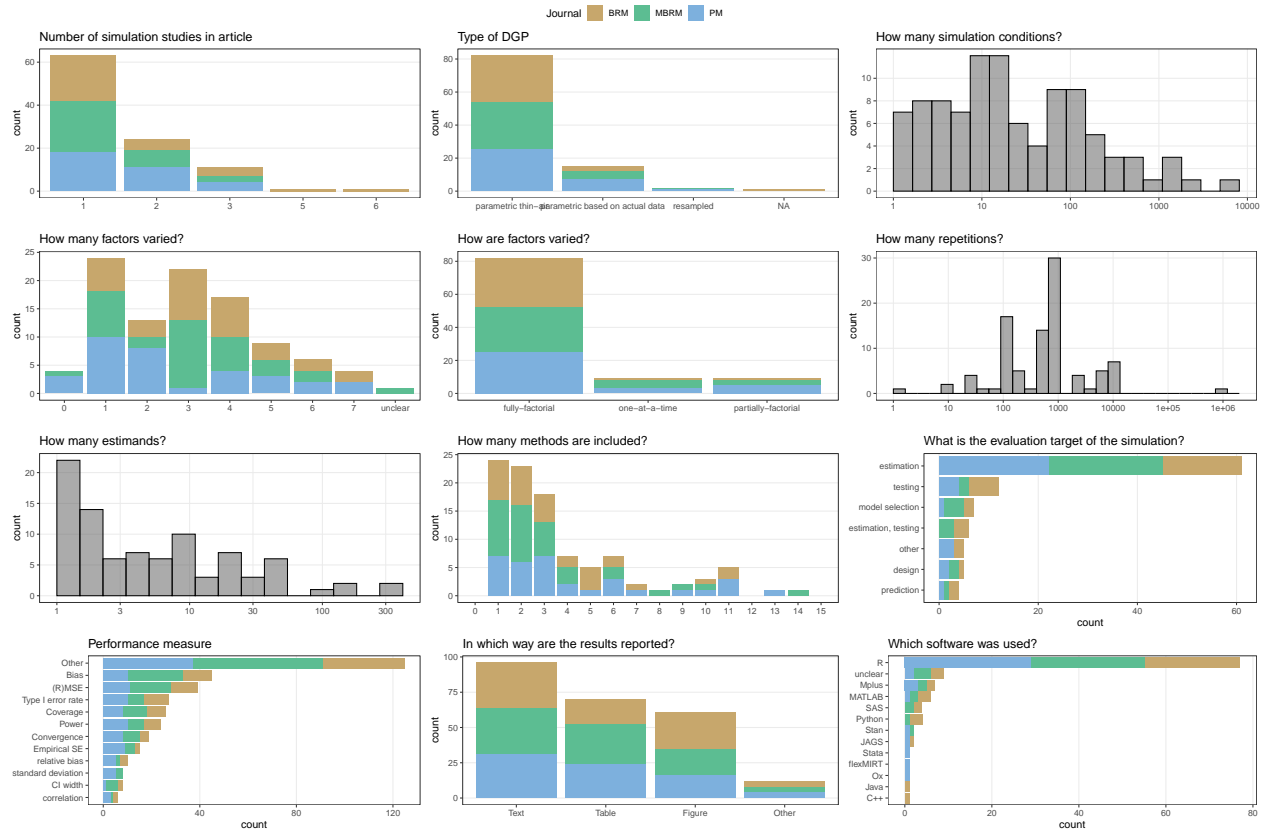
```
## Warning: Removed 1 rows containing non-finite values (`stat_bin()`).
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_bin()`).
```

```
## Warning: Removed 11 rows containing non-finite values (`stat_bin()`).
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_count()`).
```





## Descriptives

The following still needs some cleaning.

```
# detailed per-journal descriptives
journal_describe <- sim_res_fac %>%
  split(.$journal) %>%
  purrr::map(~Hmisc::describe(.x))

# In a tidy way for long summary table
summary_vars <- c(
  "reviewer", "simstudy_q1",
  "nsimstudies_q2", "whichsim", "aimsdefined_q3",
  "dgptype_q4",
  # "dgpparameters_q5",
  # "nconds_q6",
  "factorsvaried_q7", "dgmfactorial_q7",
  # "nsim_q8",
  "nsimjustified_q9",
  "estimandstated_q10",
  # "nestimands_q11",
  "estimandsagg_q12", "truetheta_q13",
  # "nmethods_q14",
  "target_q15",
  "pmconvergence_q15", "pmbias_q15", "pmempse_q15", "pm_r_mse_q15",
  "pmcover_q15", "pmtypeerror_q15", "pmpower_q15", "pmciwidth_q15",
```

```

"pmsclear_q15", "mcerrors_q16",
"resultsfigure_q17", "resultstable_q17", "resultstext_q17",
"resultsother_q17", "software_1_q18", "software_2_q18",
"software_3_q18", "software", "userwritten_q19",
"codeprovided_q20", "seedprovided_q21",
"compenvironment_q22", "compos_q23",
"coding_confidence"
)

sim_res_fac %>%
  as.data.frame() %>%
  group_by(journal) %>%
  pivot_longer(cols = starts_with("software"),
               names_to = NULL,
               values_to = "software",
               values_drop_na = TRUE) %>%
  mutate(software = as.factor(software)) %>%
  pivot_longer(cols = any_of(summary_vars),
               names_to = "col", values_to = "response") %>%
  select(journal, col, response) %>%
  group_by(journal, col) %>%
  count(response) %>%
  # filter(!is.na(response)) %>%
  # View()
  pivot_wider(id_cols = c(col, response),
              names_from = journal,
              values_from = n,
              values_fill = 0) %>%
  arrange(factor(col, levels = summary_vars)) %>%
  rowwise() %>%
  mutate(Sum = sum(BRM, MBRM, PM)) %>%
  knitr::kable("latex", longtable = TRUE, caption = "Grouped by Journal") %>%
  kableExtra::column_spec(1, bold = TRUE) %>%
  kableExtra::collapse_rows(columns = 1:2, valign = "top")

```

Table 1: Grouped by Journal

col	response	BRM	MBRM	PM	Sum
<b>reviewer</b>	BS	11	15	13	39
	FB	5	16	16	37
	SP	22	7	11	40
<b>simstudy_q1</b>	yes	38	38	40	116
<b>nsimstudies_q2</b>	1	24	26	21	71
	2	6	8	14	28
	3	6	4	5	15
	5	1	0	0	1
	6	1	0	0	1
<b>whichsim</b>	1	31	26	31	88
	NA	7	11	8	26
	2	0	1	0	1
	3	0	0	1	1
	no	1	1	1	3
	unclear	2	1	0	3

aimsdefined\_q3

	yes	35	36	39	110
dgptype_q4	parametric based on actual data	3	6	8	17
	parametric thin-air	34	31	31	96
	NA	1	0	0	1
	resampled	0	1	1	2
factorsvaried_q7	1	8	8	12	28
	2	3	2	9	14
	3	11	14	1	26
	5	3	4	4	11
	6	3	2	2	7
	4	8	6	6	20
	7	2	0	2	4
	unclear	0	1	0	1
	0	0	1	4	5
dgmfactorial_q7	fully-factorial	35	29	31	95
	one-at-a-time	1	5	4	10
	partially-factorial	2	4	5	11
nsimjustified_q9	no	32	36	37	105
	yes	6	2	3	11
estimandstated_q10	no	6	3	2	11
	unclear	4	5	0	9
	yes	27	30	36	93
	NA	1	0	0	1
	not applicable	0	0	2	2
estimandsagg_q12	no	31	29	33	93
	unclear	1	4	0	5
	yes	6	5	7	18
truetheta_q13	estimated	2	3	2	7
	known	36	35	36	107
	not applicable	0	0	2	2
target_q15	design	1	2	2	5
	estimation	19	26	28	73
	estimation, testing	4	3	0	7
	model selection	2	4	1	7
	other	3	0	4	7
	prediction	3	1	1	5
	testing	6	2	4	12
pmconvergence_q15	no	33	30	25	88
	yes	5	8	12	25
	unclear	0	0	3	3
pmbias_q15	no	22	13	26	61
	yes	16	25	14	55
pmempse_q15	no	36	33	29	98
	yes	2	5	11	18
pm_r_mse_q15	no	24	20	24	68
	yes	14	18	16	48
pmcover_q15	no	27	27	31	85
	yes	11	11	9	31
pmtypeierror_q15	no	26	31	28	85
	yes	12	7	12	31
pmpower_q15	no	29	31	28	88
	yes	9	7	12	28
	no	36	33	38	107

<b>pmciwidth_q15</b>	yes	2	5	2	9
<b>pmsclear_q15</b>	no	2	1	3	6
	unclear	3	0	2	5
	yes	33	37	32	102
	NA	0	0	3	3
<b>mcerrors_q16</b>	boxplot	1	4	2	7
	nowhere	29	28	33	90
	plot	1	0	0	1
	plot, table	1	0	0	1
	standard deviation	5	1	0	6
	text	1	0	1	2
	figures	0	1	0	1
	reference intervals	0	1	0	1
	tables, text	0	2	0	2
	NA	0	1	0	1
	quantile plots	0	0	1	1
	ranges	0	0	1	1
	table	0	0	2	2
<b>resultsfigure_q17</b>	no	7	17	23	47
	yes	31	21	17	69
<b>resultstable_q17</b>	no	16	8	9	33
	yes	22	30	31	83
<b>resultstext_q17</b>		38	36	38	112
<b>resultsother_q17</b>	no	0	2	2	4
		32	34	36	102
	yes	6	4	4	14
<b>software</b>	unclear	3	4	2	9
	C++	1	0	0	1
	JAGS	1	0	1	2
	Java	1	0	0	1
	MATLAB	3	2	1	6
	Mplus	2	2	3	7
	Python	3	1	0	4
	R	22	26	29	77
	SAS	2	2	0	4
	Stan	0	1	1	2
	Ox	0	0	1	1
	flexMIRT	0	0	1	1
	Stata	0	0	1	1
<b>userwritten_q19</b>	no	20	21	12	53
	unclear	4	2	4	10
	yes	14	15	24	53
<b>codeprovided_q20</b>	accessible online	19	6	18	43
	not accessible	19	32	22	73
<b>seedprovided_q21</b>	yes	9	5	11	25
	not found	29	33	29	91
<b>compenvironment_q22</b>	no	11	10	7	28
	fully	1	0	1	2
	minimal	16	24	27	67
	partially	10	4	5	19
<b>compos_q23</b>	no	35	36	37	108
	fully	2	1	1	4
	partially	1	1	2	4
	great	24	18	28	70

# coding\_confidence

	medium	10	18	11	39
	poor	4	2	1	7

*# Ungrouped and with proportions*

```
sim_res_fac %>%
  as.data.frame() %>%
  pivot_longer(cols = starts_with("software"),
               names_to = NULL,
               values_to = "software",
               values_drop_na = TRUE) %>%
  mutate(software = as.factor(software)) %>%
  pivot_longer(cols = any_of(summary_vars),
               names_to = "col", values_to = "response") %>%
  select(col, response) %>%
  group_by(col) %>%
  count(response) %>%
  arrange(factor(col, levels = summary_vars)) %>%
  knitr::kable("latex", longtable = TRUE, caption = "Ungrouped") %>%
  kableExtra::column_spec(1, bold = TRUE) %>%
  kableExtra::collapse_rows(columns = 1:2, valign = "top")
```

Table 2: Ungrouped

col	response	n
reviewer	BS	39
	FB	37
	SP	40
simstudy_q1	yes	116
nsimstudies_q2	1	71
	2	28
	3	15
	5	1
	6	1
whichsim	1	88
	2	1
	3	1
	NA	26
aimsdefined_q3	no	3
	unclear	3
	yes	110
dgptype_q4	parametric based on actual data	17
	parametric thin-air	96
	resampled	2
	NA	1
factorsvaried_q7	unclear	1
	1	28
	2	14
	3	26
	5	11
	6	7
	0	5
	4	20
	7	4

<b>dgmfactorial_q7</b>	fully-factorial	95
	one-at-a-time	10
	partially-factorial	11
<b>nsimjustified_q9</b>	no	105
	yes	11
<b>estimandstated_q10</b>	no	11
	unclear	9
	yes	93
	not applicable	2
	NA	1
<b>estimandsagg_q12</b>	no	93
	unclear	5
	yes	18
<b>truetheta_q13</b>	not applicable	2
	estimated	7
	known	107
<b>target_q15</b>	design	5
	estimation	73
	estimation, testing	7
	model selection	7
	other	7
	prediction	5
	testing	12
<b>pmconvergence_q15</b>	no	88
	unclear	3
	yes	25
<b>pmbias_q15</b>	no	61
	yes	55
<b>pmempse_q15</b>	no	98
	yes	18
<b>pm_r_mse_q15</b>	no	68
	yes	48
<b>pmcover_q15</b>	no	85
	yes	31
<b>pmtypeierror_q15</b>	no	85
	yes	31
<b>pmpower_q15</b>	no	88
	yes	28
<b>pmciwidth_q15</b>	no	107
	yes	9
<b>pmsclear_q15</b>	no	6
	unclear	5
	yes	102
	NA	3
	boxplot	7
	figures	1
	nowhere	90
	plot	1
	plot, table	1
	quantile plots	1
	ranges	1
	reference intervals	1
	standard deviation	6
	table	2

	tables, text	2
	text	2
	NA	1
<b>resultsfigure_q17</b>	no	47
	yes	69
<b>resultstable_q17</b>	no	33
	yes	83
<b>resultstext_q17</b>	no	4
	yes	112
<b>resultsother_q17</b>	no	102
	yes	14
<b>software</b>	unclear	9
	C++	1
	JAGS	2
	Java	1
	MATLAB	6
	Mplus	7
	Ox	1
	Python	4
	R	77
	SAS	4
	flexMIRT	1
	Stan	2
	Stata	1
<b>userwritten_q19</b>	no	53
	unclear	10
	yes	53
<b>codeprovided_q20</b>	accessible online	43
	not accessible	73
<b>seedprovided_q21</b>	yes	25
	not found	91
<b>compenvironment_q22</b>	no	28
	fully	2
	minimal	67
	partially	19
<b>compos_q23</b>	no	108
	fully	4
	partially	4
<b>coding_confidence</b>	great	70
	medium	39
	poor	7

```
sessionInfo()
```

```
## R version 4.3.1 (2023-06-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 12 (bookworm)
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p-r0.3.21.so; LAPACK version 3.11.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
```

```

## [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8   LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8      LC_NAME=C
## [9] LC_ADDRESS=C              LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Zurich
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] kableExtra_1.3.4.9000 knitr_1.43      forcats_1.0.0
## [4] stringr_1.5.0          ggpubr_0.6.0     colorspace_2.1-0
## [7] ggplot2_3.4.2          tidyr_1.3.0      dplyr_1.1.2
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.3      xfun_0.39        htmlwidgets_1.6.2 rstatix_0.7.2
## [5] vctr_0.6.3        tools_4.3.1      generics_0.1.3    tibble_3.2.1
## [9] fansi_1.0.4       highr_0.10       cluster_2.1.4     pkgconfig_2.0.3
## [13] data.table_1.14.8 checkmate_2.2.0   webshot_0.5.5     lifecycle_1.0.3
## [17] compiler_4.3.1    farver_2.1.1     munsell_0.5.0     carData_3.0-5
## [21] htmltools_0.5.5   yaml_2.3.7       htmlTable_2.4.1   Formula_1.2-5
## [25] pillar_1.9.0      car_3.1-2        Hmisc_5.1-0       rpart_4.1.19
## [29] abind_1.4-5       tidyselect_1.2.0 rvest_1.0.3       digest_0.6.33
## [33] stringi_1.7.12    purrr_1.0.1      labeling_0.4.2    cowplot_1.1.1
## [37] fastmap_1.1.1     grid_4.3.1       cli_3.6.1         magrittr_2.0.3
## [41] base64enc_0.1-3   utf8_1.2.3       broom_1.0.5       foreign_0.8-84
## [45] withr_2.5.0       scales_1.2.1     backports_1.4.1   rmarkdown_2.23
## [49] httr_1.4.6        nnet_7.3-18      gridExtra_2.3     ggsignif_0.6.4
## [53] evaluate_0.21     viridisLite_0.4.2 rlang_1.1.1       glue_1.6.2
## [57] xml2_1.3.5        svglite_2.1.1    rstudioapi_0.15.0 R6_2.5.1
## [61] systemfonts_1.0.4

```