Load necessary packages

```r
library(tidyverse)
library(DescTools)
library(ryouready)
library(reshape2)
library(stringr)
library(ggthemes)
theme_set(theme_fivethirtyeight())
```

Load file and do some preprocessing

```r
file = 'C:/Users/Cole/Documents/DATA/PLIC_DATA/Experts/Experts_v08042020_09202020.csv'

info <- data.frame(lapply(read.csv(file, nrows = 1),
                          FUN = function(x) gsub("^.*- ", "", x)))

headers = read.csv(file, header = F, nrows = 1, as.is = T)
headers <- data.frame(lapply(headers, function(x) sub(' ', '.', x)))
df = read.csv(file, header = F)
df <- df %>%
  tail(., nrow(df) - 3) %>%
  `colnames<-`(headers) %>%
  filter(Finished == 1) %>%
  filter((Qt1_Page.Submit > 30) | (Qt2_Page.Submit > 30) | (Qt3_Page.Submit > 30) |
           (Qt4_Page.Submit > 30)) %>%
  filter(!(Q6a %in% c(1, 4)))

# check which entries to manually delete
df[df$Q6a %in% c(8, 9), c('Q6a', 'Q6a_8_TEXT')]
```

```
##     Q6a              Q6a_8_TEXT
## 4     8               HS Teacher
## 11    8 Teaching Professor Line
## 16    8                sophomore
```

```r
df <- df[!(row.names(df) %in% c(4, 16)),]

df <- recode2(df, vars = colnames(df)[colnames(df) %like any% c('%Q139%',
                                                                '%Q140%',
                                                                '%Q141%',
                                                                '%Q157%')],
              recodes = "6 = 1; 7 = 2; 9 = 3; 10 = 4")
df <- recode2(df, vars = c('Q1a'), recodes = "1 = 1; 2 = 2; 5 = 3; 3 = 4")
df <- recode2(df, vars = c('Q133'), recodes = "1 = 1; 2 = 2; 5 = 3; 6 = 4")
df <- recode2(df, vars = c('Q1a', 'Q2a', 'Q3a'), recodes = "1 = 'Support';
              2 = 'Refute'; 3 = 'Neither'; 4 = 'More evidence'")
df <- recode2(df, vars = c('Q133', 'Q144', 'Q148'),
              recodes = "1 = 'Data collection'; 2 = 'Amount of data';
```

```
                         3 = 'Analysis'; 4 = 'Others'")
df <- recode2(df, vars = c('Q3c'),
              recodes = "1 = 'fixed at zero'; 2 = 'not fixed at zero';
                         3 = 'other'")
df <- recode2(df, vars = c('Q4a'),
              recodes = "1 = 'Group 1'; 2 = 'Group 2'; 3 = 'Both';
                         4 = 'Neither'")
df <- recode2(df, vars = colnames(df)[colnames(df) %like any% c('%Q152%',
                                                                '%Q153%',
                                                                '%Q154%')],
              recodes = "1 = '1'; 2 = '2'; 3 = 'B'; 4 = 'N'")
```
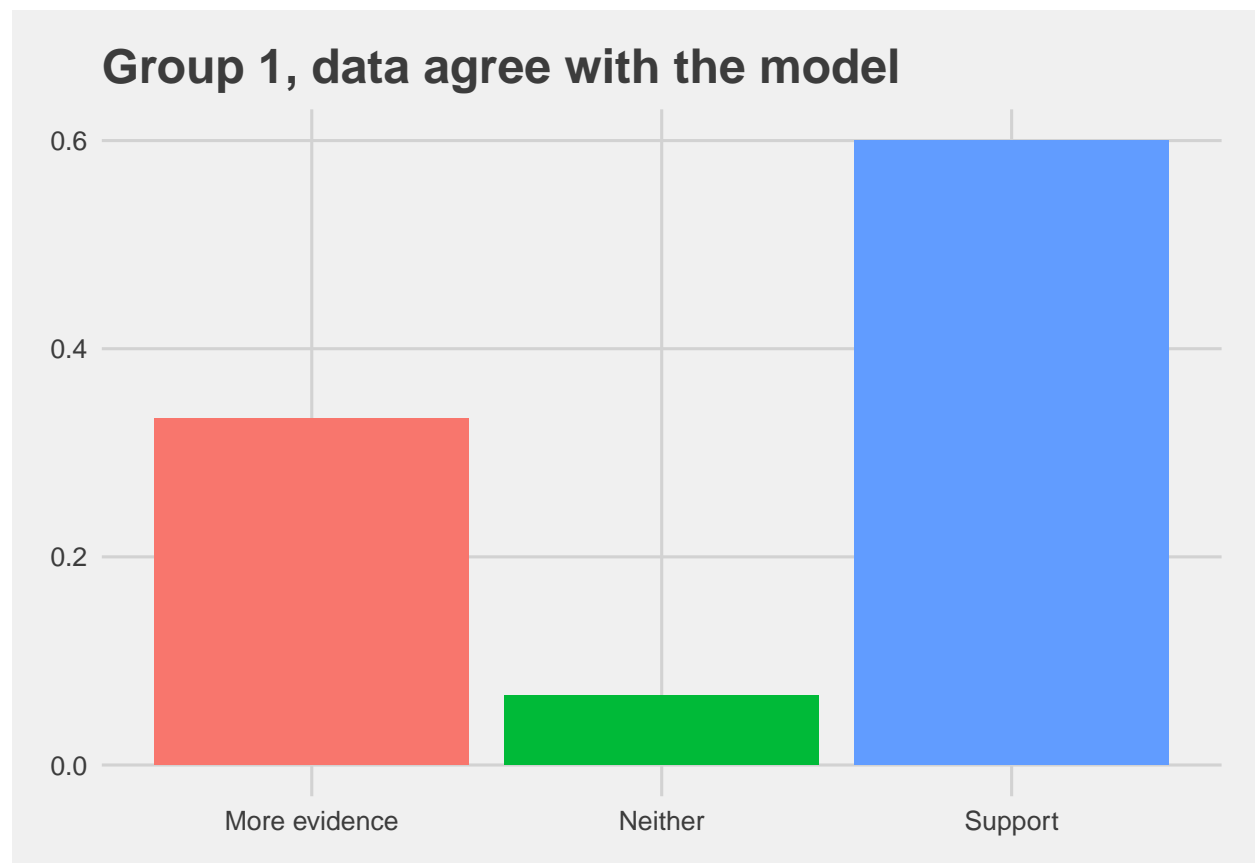
Analysis

Page 1

```
# What do you think Group 1 should say about whether their data agree with the
# model?

ggplot(df, aes(x = factor(Q1a), fill = factor(Q1a))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 1, data agree with the model') +
  theme(legend.position = "none")
```

```r
# Which statements below support your reasoning for whether Group 1's data agree
# with the model?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q1b%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                                t.info...cols.temp..
## Q1b_2                                     The two spring constant values are not exactly the same
## Q1b_5     The two spring constant values do not agree within experimental uncertainties
## Q1b_8               The percent difference between the spring constant values is large
## Q1b_35                          The average periods for the two masses are different
## Q1b_28    The size of the uncertainty (or variability between measurements) is large
## Q1b_19                                                         Other (Please describe)
## Q1b_37                                 The two spring constant values are similar
## Q1b_33        The two spring constant values agree within experimental uncertainties
## Q1b_34            The percent difference between the spring constant values is small
## Q1b_16                          The average periods for the two masses are similar
## Q1b_36    The size of the uncertainty (or variability between measurements) is small
```

```r
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity')  +
  ggtitle('Group 1, reasoning') +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```
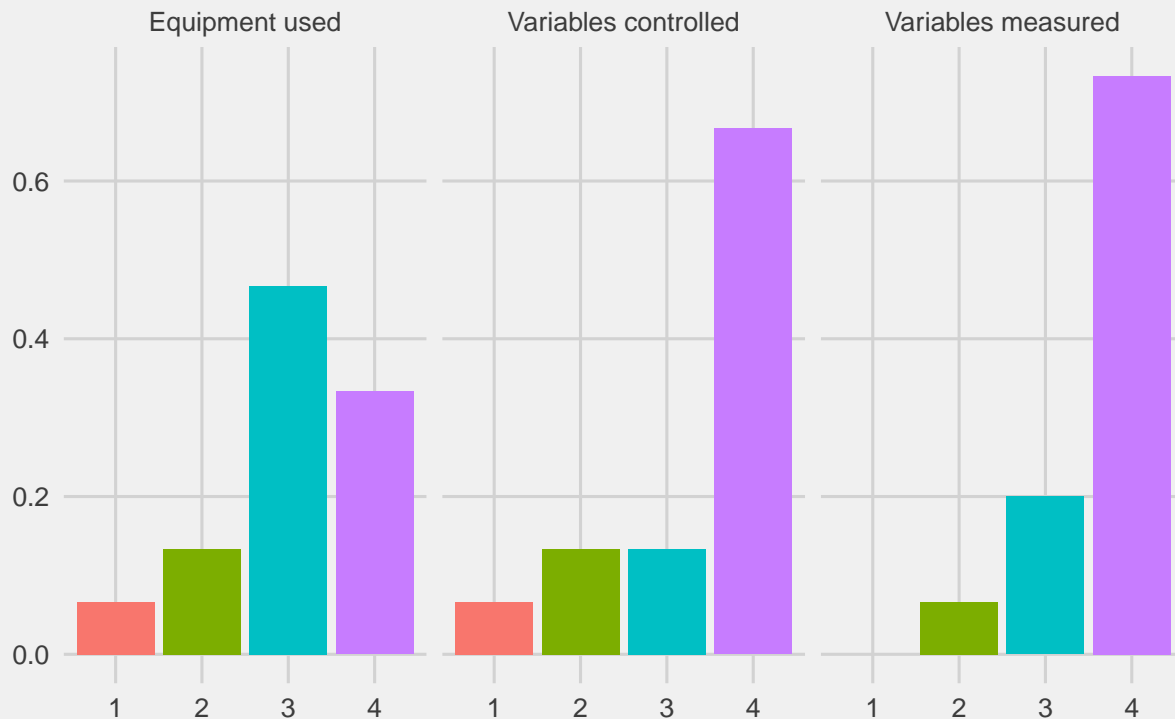
Group 1, reasoning
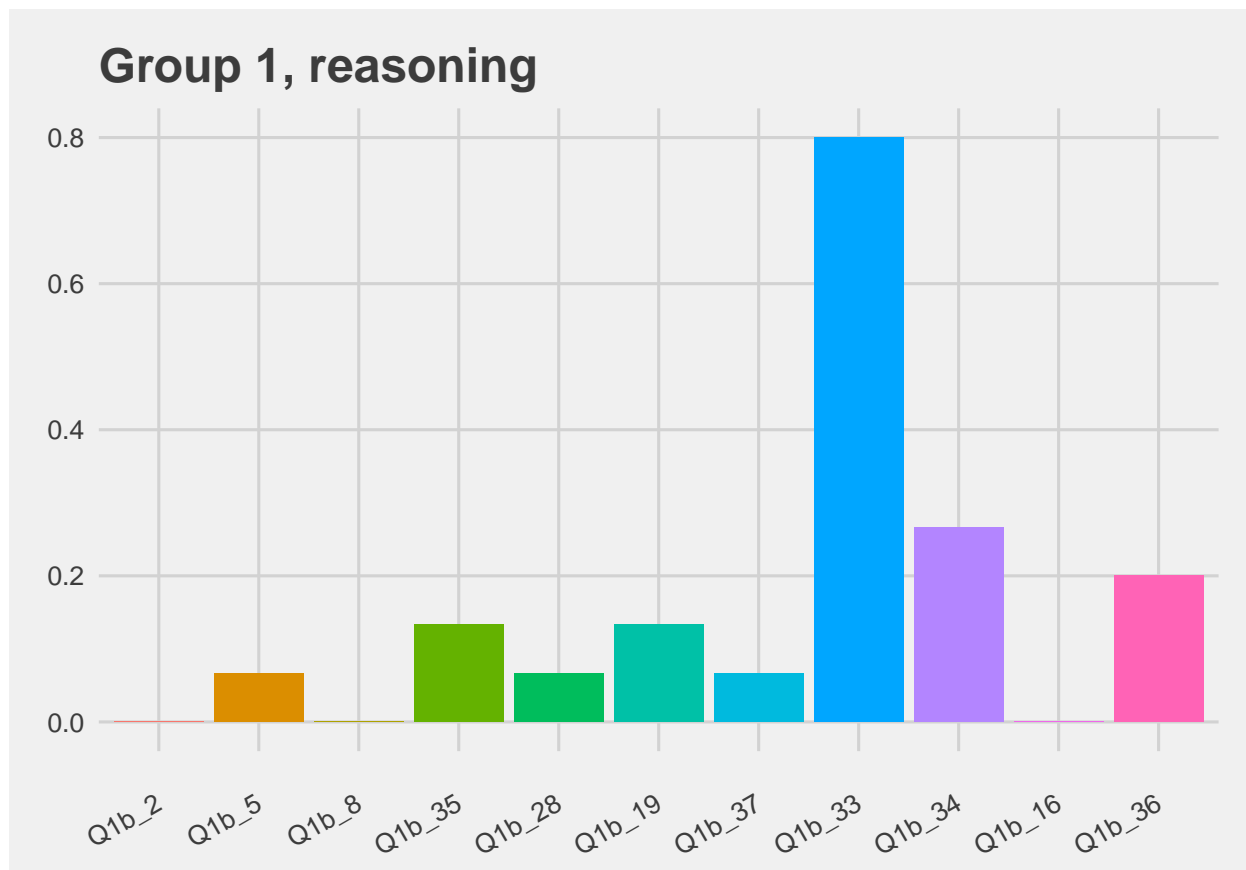

```r
cols.temp <- colnames(df)[(colnames(df) %like% '%Q139%') &
                            !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q139_1' ~ 'Equipment used',
    variable == 'Q139_2' ~ 'Variables measured',
    variable == 'Q139_3' ~ 'Variables controlled',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Group 1, data collection methods') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```

Group 1, data collection methods

Equipment used | Variables controlled | Variables measured
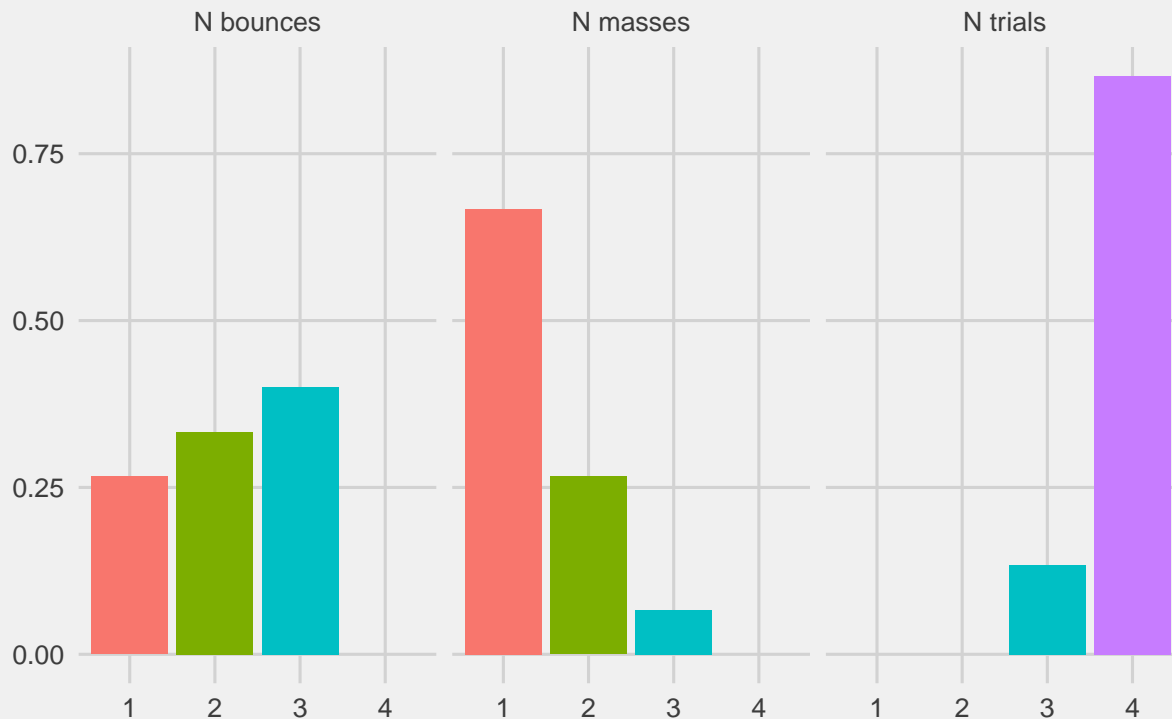
```
cols.temp <- colnames(df)[(colnames(df) %like% '%Q140%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q140_1' ~ 'N trials',
    variable == 'Q140_2' ~ 'N masses',
    variable == 'Q140_3' ~ 'N bounces',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Group 1, amount of data') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```

Group 1, amount of data

```
cols.temp <- colnames(df)[(colnames(df) %like% '%Q141%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q141_1' ~ 'Explanations',
    variable == 'Q141_2' ~ 'Analysis',
    variable == 'Q141_3' ~ 'Similar k values',
    variable == 'Q141_4' ~ 'Uncertainty in data and k',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Group 1, analysis and outcomes') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```
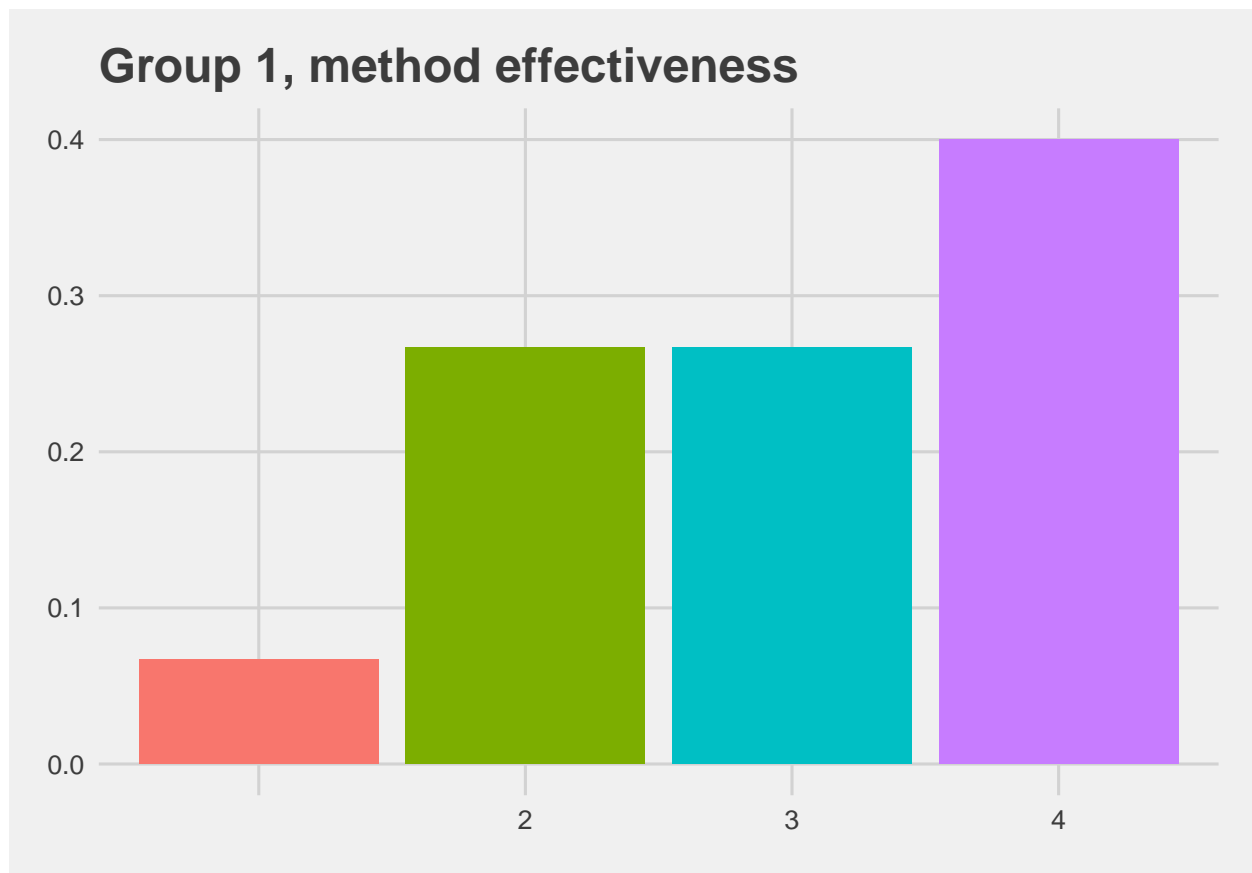
Group 1, analysis and outcomes
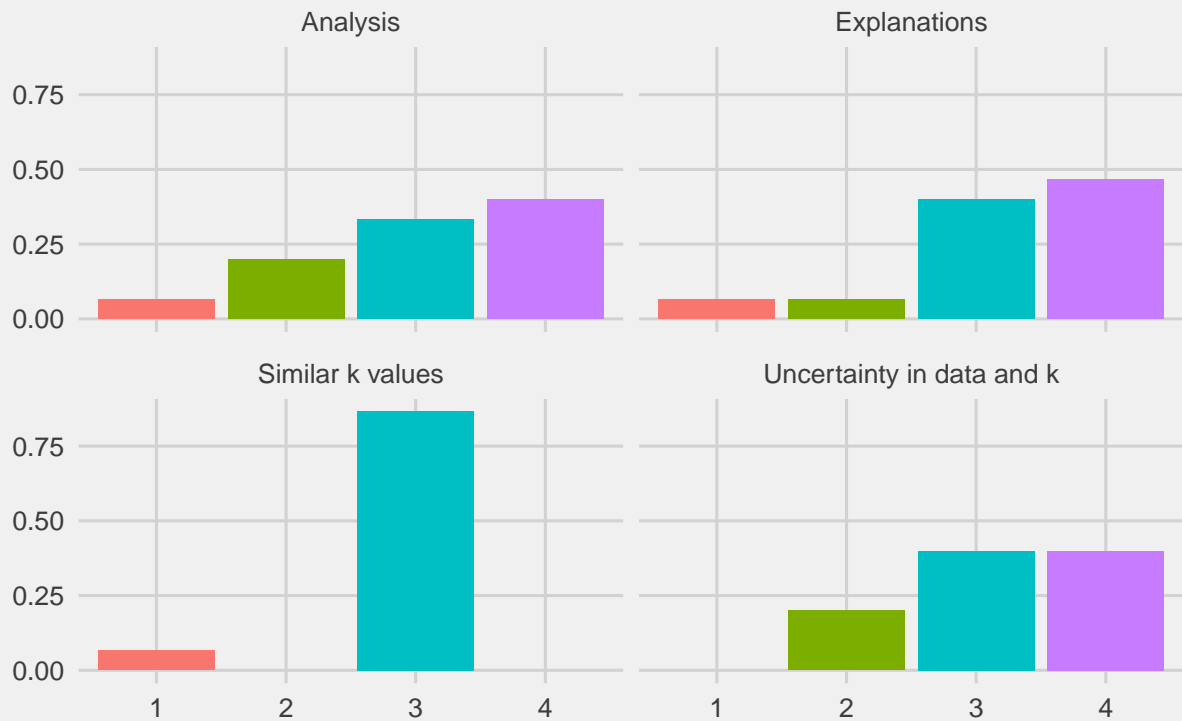

```r
# How effective do you think Group 1's method was for testing whether the data
# obey the model?

ggplot(df, aes(x = factor(Q1c), fill = factor(Q1c))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 1, method effectiveness') +
  theme(legend.position = "none")
```

Group 1, method effectiveness


```
# What feature was most important to you in evaluating the effectiveness of Group
# 1's methods (either as strengths or weaknesses)?

ggplot(df, aes(x = factor(Q133), fill = factor(Q133))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 1, feature of method') +
  theme(legend.position = "none")
```

Group 1, feature of method
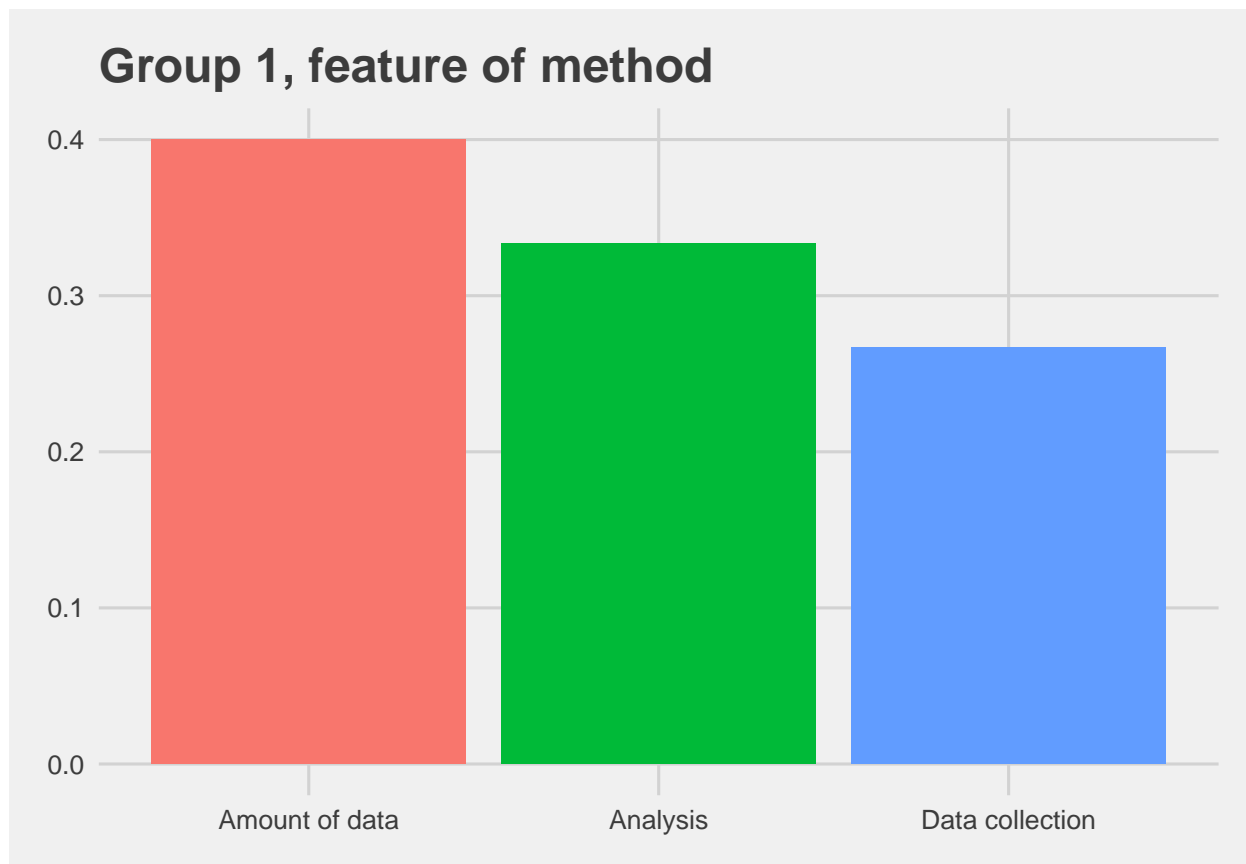
```r
# What do you think Group 1 should do next?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q134%') &
                            !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                            t.info...cols.temp..
## Q134_1                                              Test or control other variables
## Q134_2  Reduce uncertainty (e.g., more trials for the same masses, more bounces per trial, etc.)
## Q134_10                                                      Account for human error
## Q134_3                             Repeat the experiment with more and different masses
## Q134_9                                   Repeat the experiment with better equipment
## Q134_4       Use a different analysis (e.g., graph the results, incorporate systematic effects)
## Q134_6                                       Compare their k-values to the expected value
## Q134_8                                       Design a new experiment to test the results
## Q134_5                                              Check their work and write it up
## Q134_7                                                                          Other
```

```r
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity')  +
```

```
ggtitle('Group 1, what to do next') +
theme(legend.position = "none",
      axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```
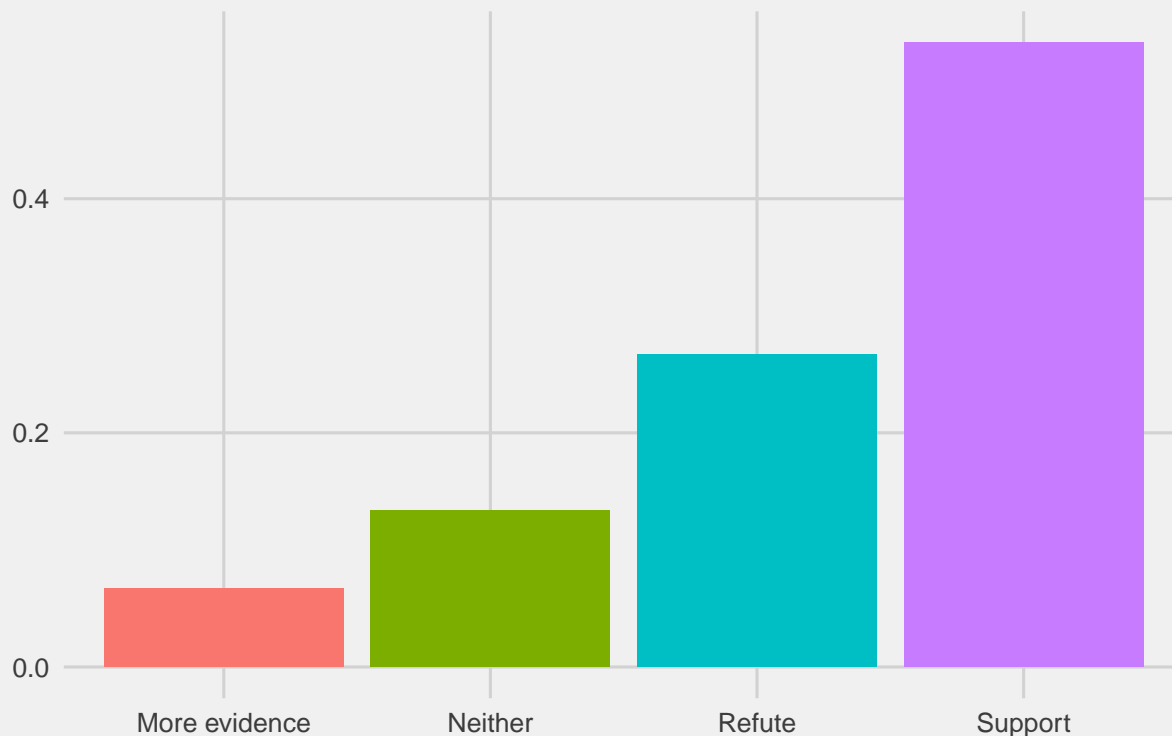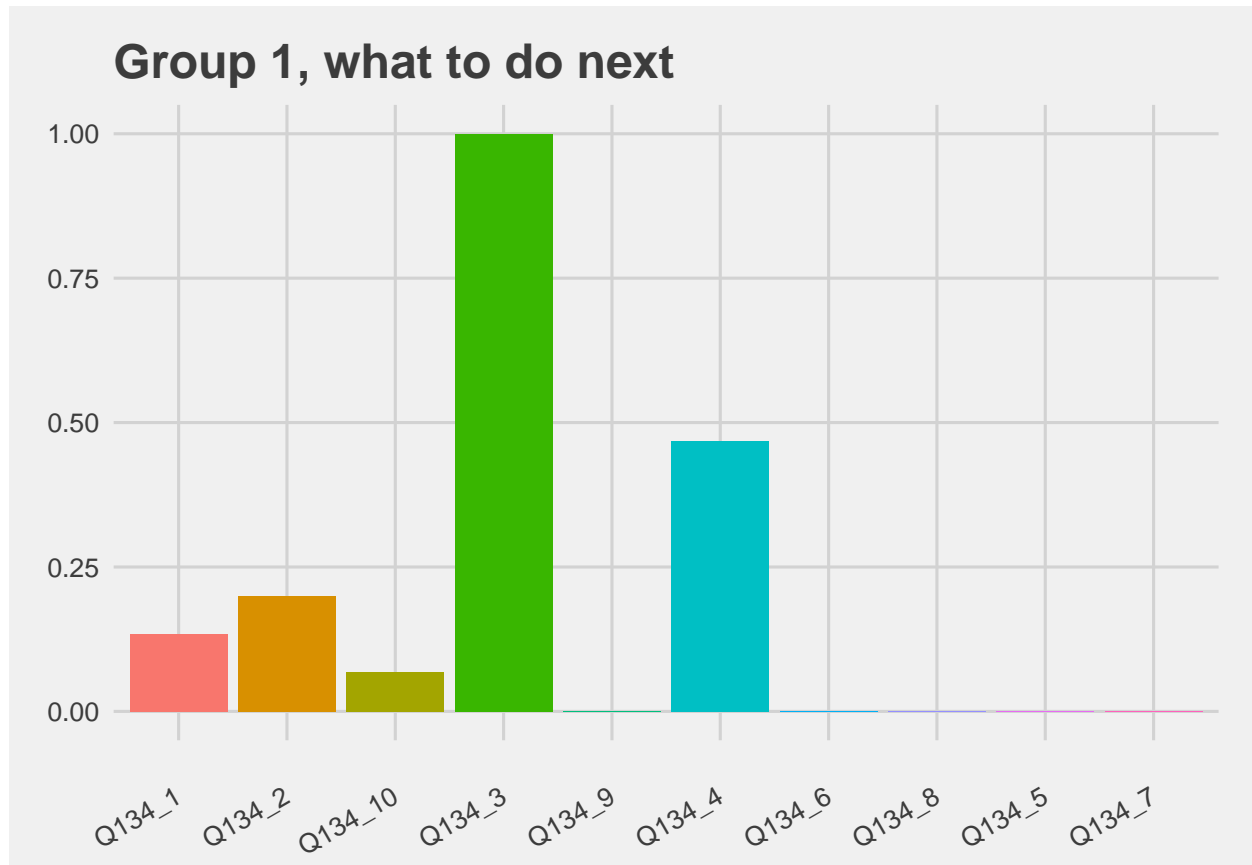
Group 1, what to do next

```
# What do you think Group 2 should say about whether their data agree with the
# model?

ggplot(df, aes(x = factor(Q2a), fill = factor(Q2a))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 2, data agree with the model') +
  theme(legend.position = "none")
```

Group 2, data agree with the model


```r
# Which statements below support your reasoning for whether Group 2's data agree
# with the model?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q2b%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                              t.info...cols.temp..
## Q2b_8                                            The points are close to the line
## Q2b_11          The points are close to the line compared to the uncertainties
## Q2b_21               There are equal numbers of points above and below the line
## Q2b_2               The points are randomly scattered above and below the line
## Q2b_9                                                  There are very few outliers
## Q2b_36                                    The uncertainty in the points is small
## Q2b_6       Many of the points cross the line (when considering their uncertainties)
## Q2b_1                                          The data are approximately linear
## Q2b_38                                                    Other (Please describe)
## Q2b_39                                           The points are far from the line
## Q2b_40          The points are far from the line compared to the uncertainties
## Q2b_41               There are not equal numbers of points above and below the line
```

```
## Q2b_42 There is a pattern in how the points are scattered above and below the line
## Q2b_43                                                   There are too many outliers
## Q2b_44                                      The uncertainty in the points is large
## Q2b_45          Too few points cross the line (when considering their uncertainties)
## Q2b_46                                              The data are curved (not linear)
```

```r
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity')  +
  ggtitle('Group 2, reasoning') +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```
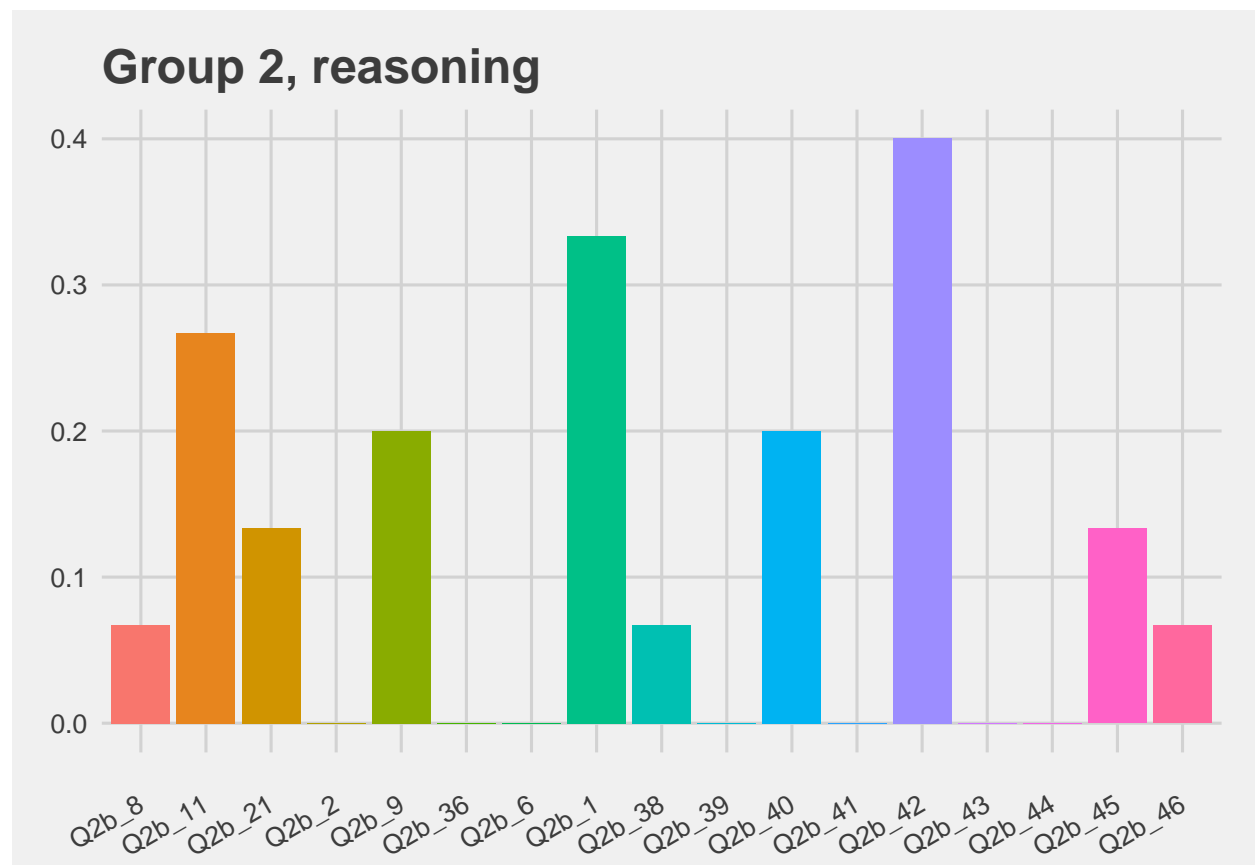


```r
cols.temp <- colnames(df)[(colnames(df) %like% '%Q2d%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q2d_1' ~ 'Equipment used',
    variable == 'Q2d_2' ~ 'Variables measured',
    variable == 'Q2d_3' ~ 'Variables controlled',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')
```

```
ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Group 2, data collection methods') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```


Group 2, data collection methods
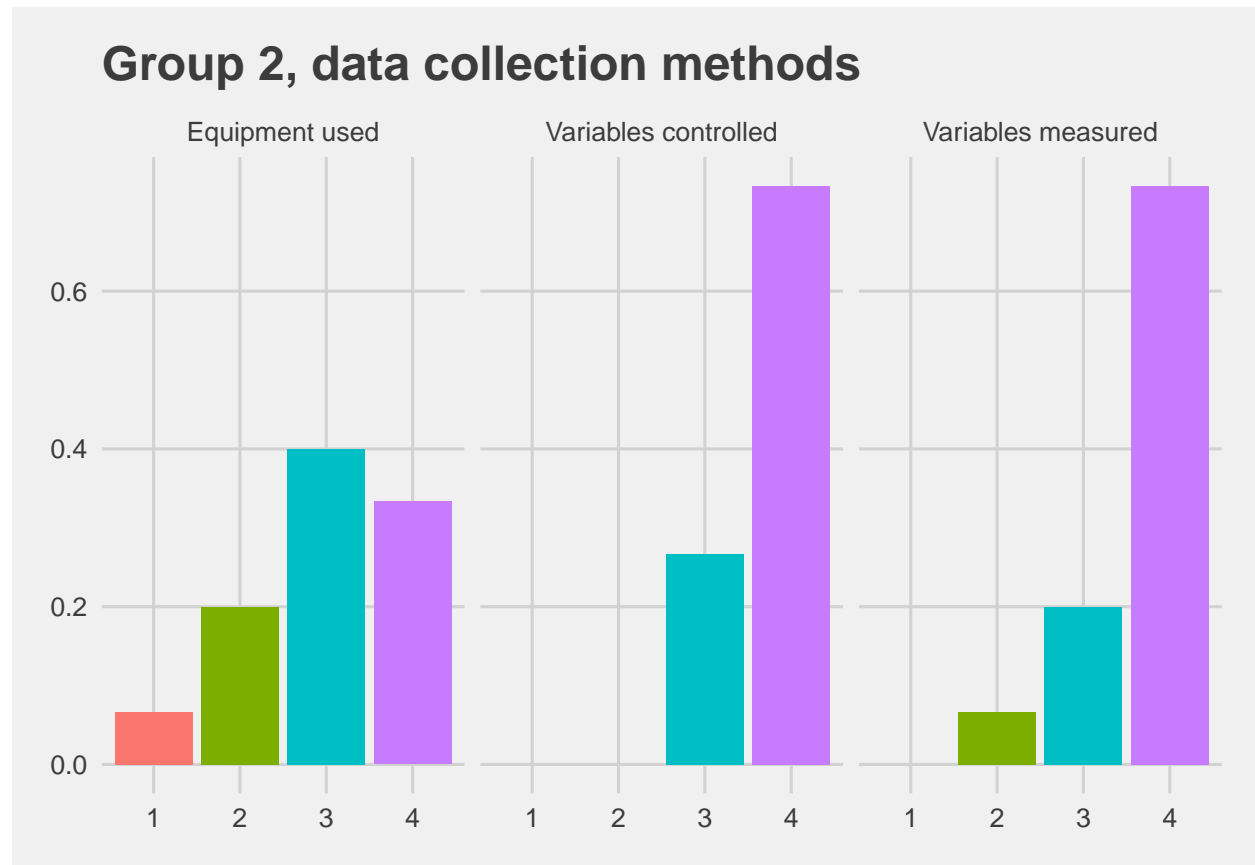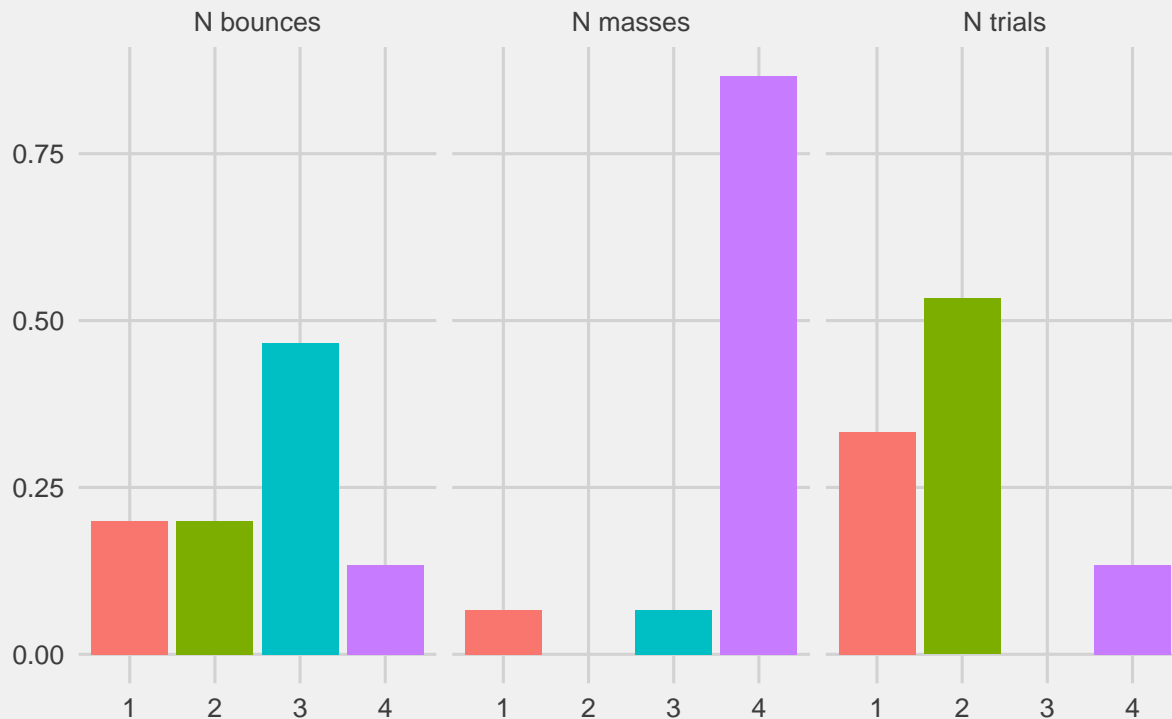
```
cols.temp <- colnames(df)[(colnames(df) %like% '%Q142%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q142_1' ~ 'N trials',
    variable == 'Q142_2' ~ 'N masses',
    variable == 'Q142_3' ~ 'N bounces',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Group 2, amount of data') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```

Group 2, amount of data

| N bounces | N masses | N trials |


```
cols.temp <- colnames(df)[(colnames(df) %like% '%Q157%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q157_1' ~ 'Explanations',
    variable == 'Q157_2' ~ 'Analysis',
    variable == 'Q157_3' ~ 'Similar k values',
    variable == 'Q157_4' ~ 'Uncertainty in data and k',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Group 2, analysis and outcomes') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```
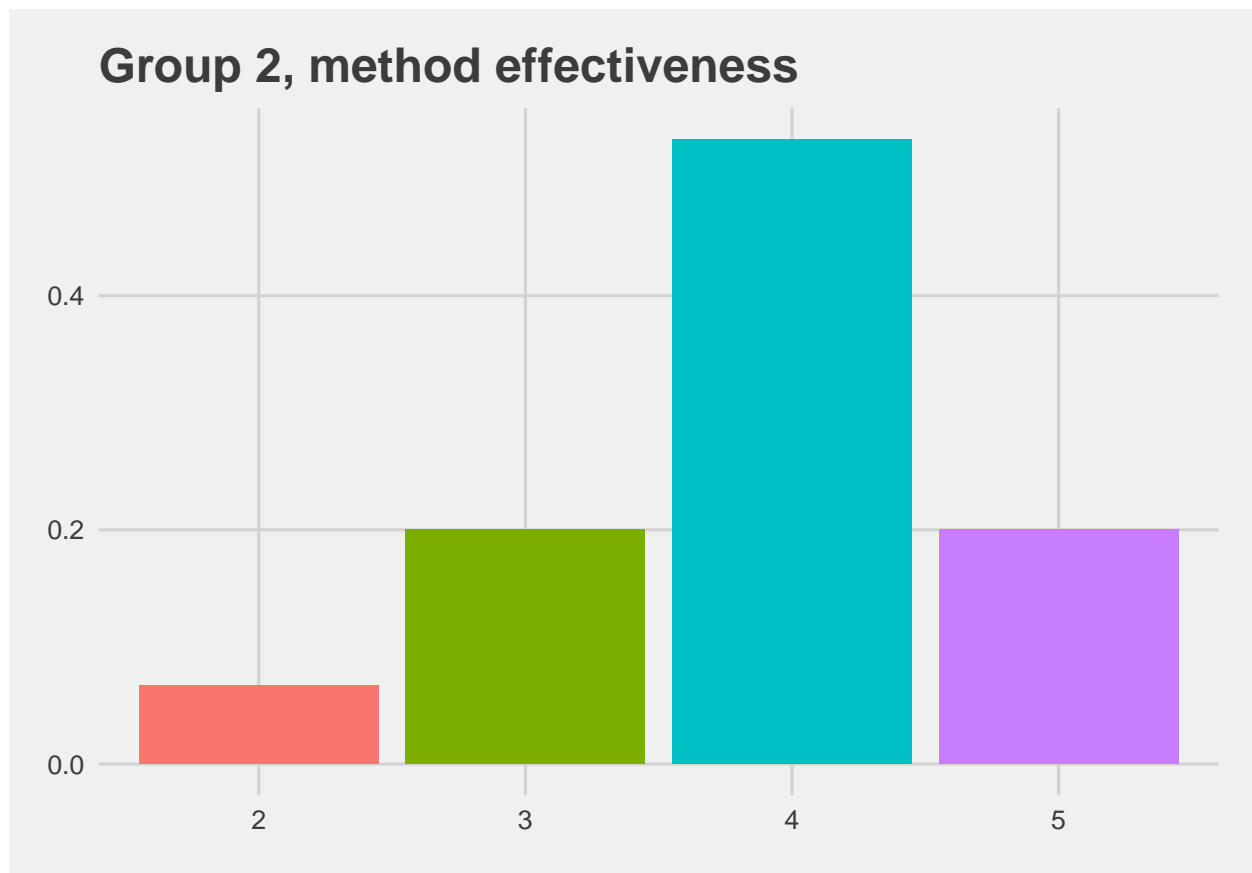
Group 2, analysis and outcomes
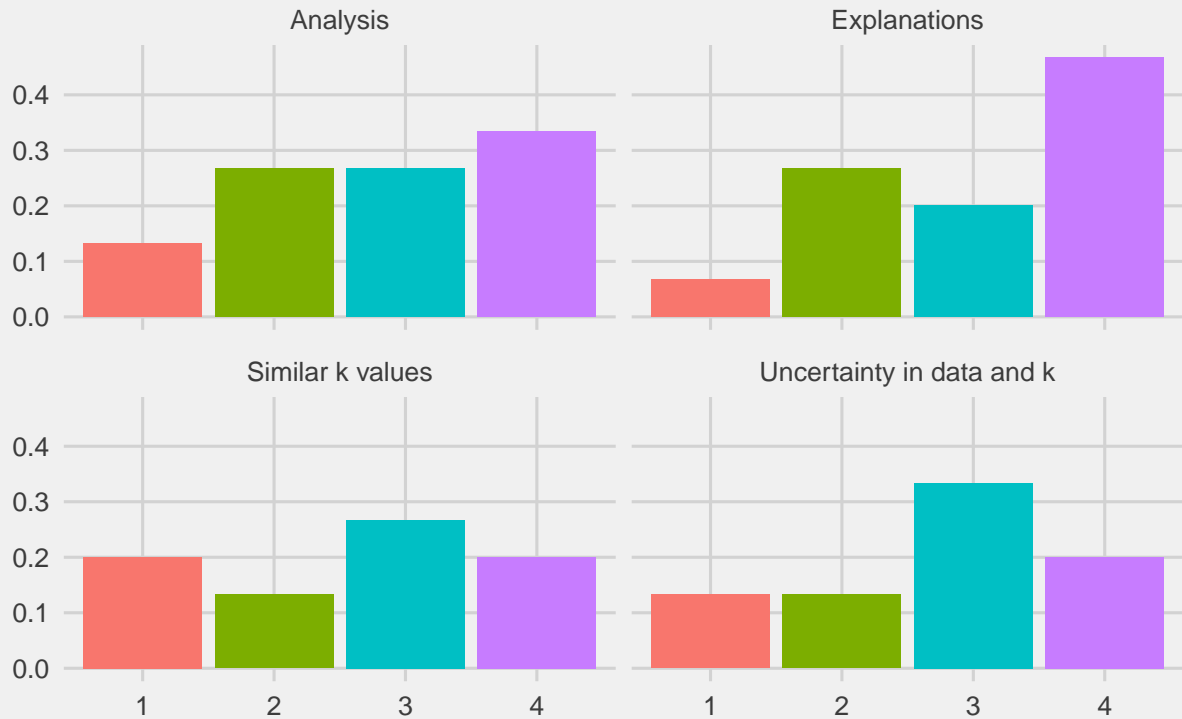

```
# How effective do you think Group 2's method was for testing whether the data
# obey the model?

ggplot(df, aes(x = factor(Q2c), fill = factor(Q2c))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 2, method effectiveness') +
  theme(legend.position = "none")
```
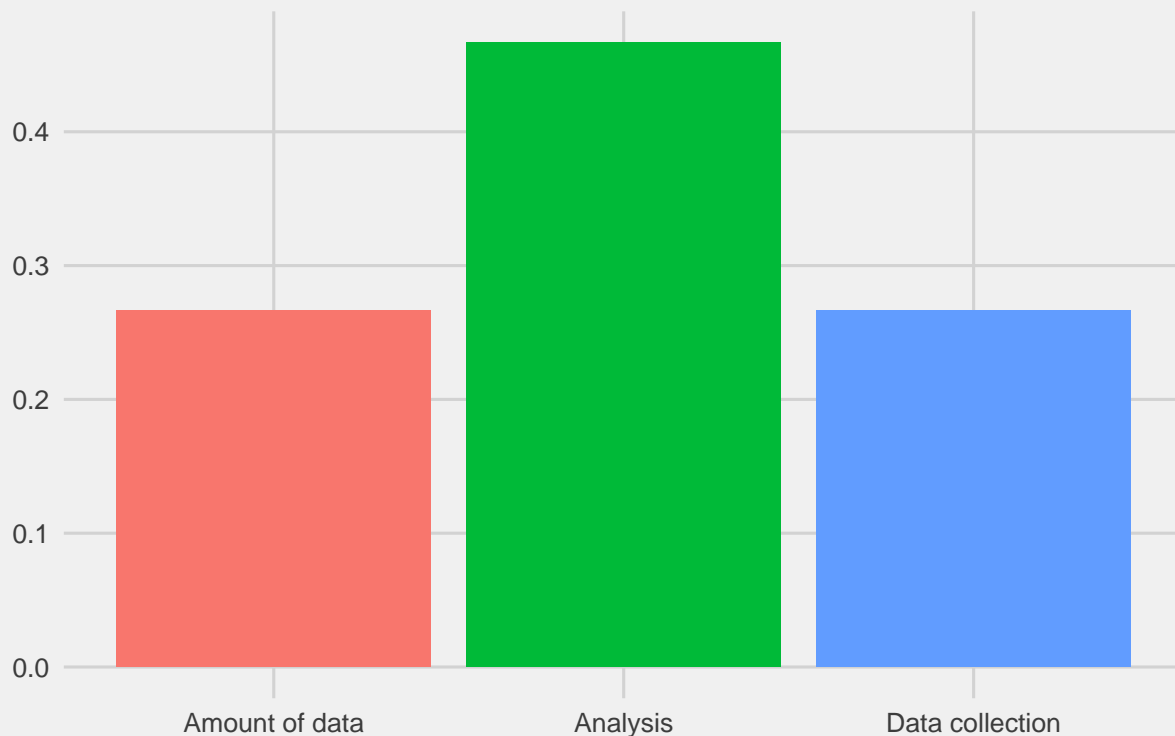
Group 2, method effectiveness


```
# What feature was most important to you in evaluating the effectiveness of Group
# 2's methods (either as strengths or weaknesses)?

ggplot(df, aes(x = factor(Q144), fill = factor(Q144))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 2, feature of method') +
  theme(legend.position = "none")
```

Group 2, feature of method


```
# What do you think Group 2 should do next?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q2e%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                                   t.info...cols.temp..
## Q2e_6                                                    Test or control other variables
## Q2e_4                                         Repeat the experiment with better equipment
## Q2e_12      Change the analysis (e.g., use a different fit line, incorporate systematic effects)
## Q2e_14                                           Compare their k-value to the expected value
## Q2e_15                                             Design a new experiment to test the results
## Q2e_16 Reduce uncertainty (e.g., more trials for the same masses, more bounces per trial, etc.)
## Q2e_17                                                   Check their work and write it up
## Q2e_34                               Repeat the experiment with more and different masses
## Q2e_11                                                             Other (Please describe)
```

```
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity')  +
  ggtitle('Group 2, what to do next') +
```

```r
theme(legend.position = "none",
      axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```
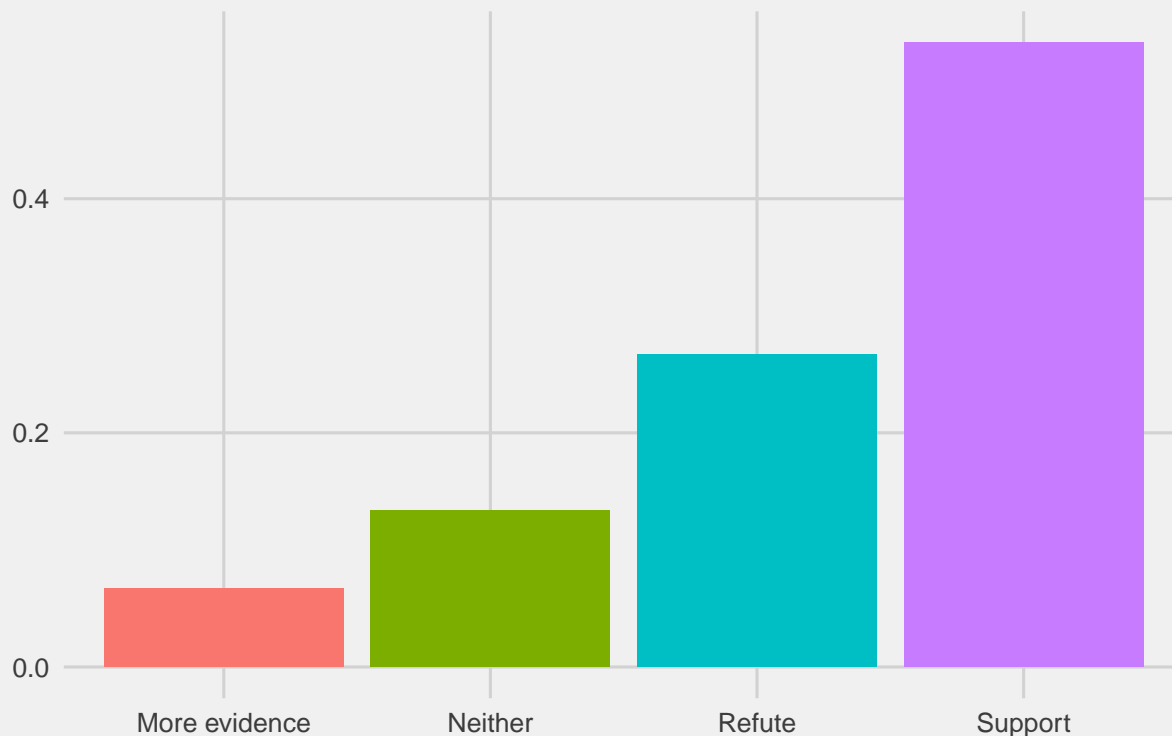
Group 2, what to do next
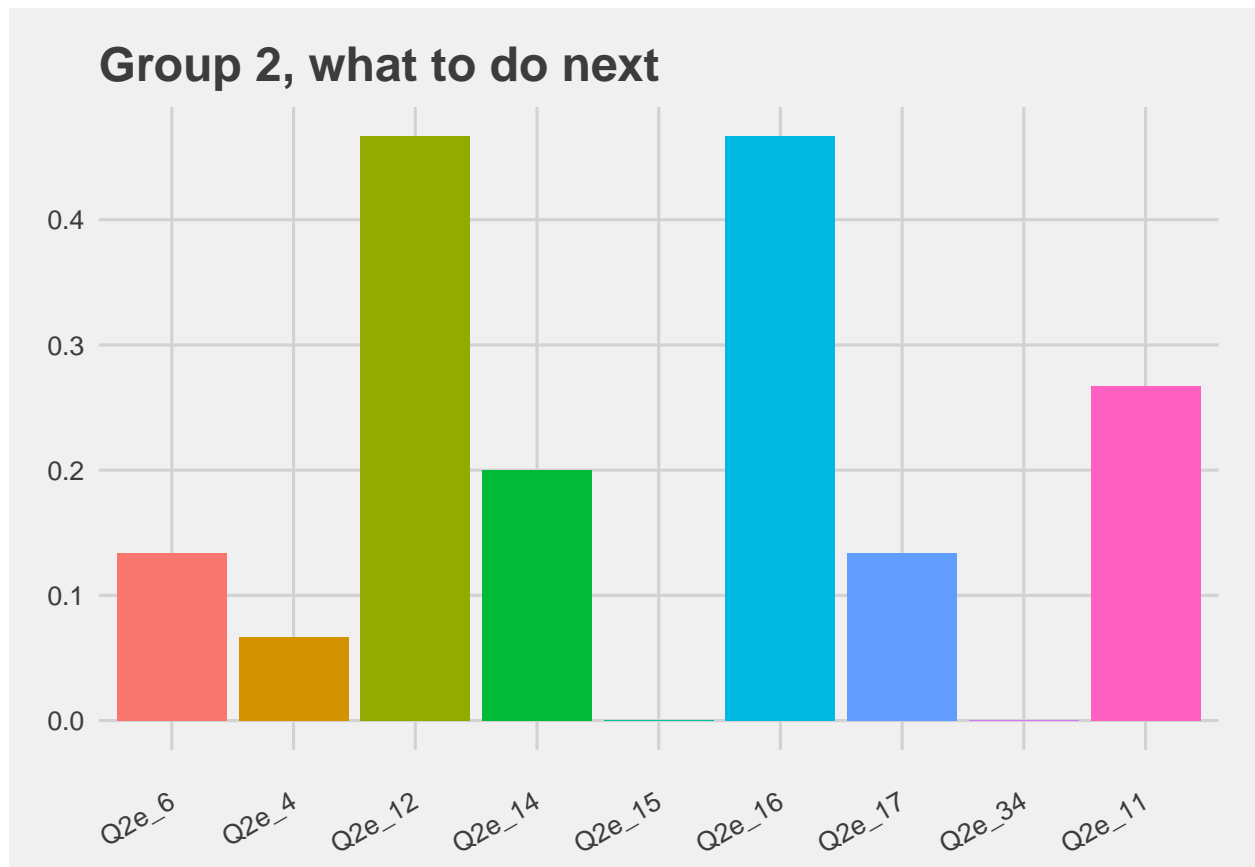
Page 3

```r
# What do you think Group 2 should say about whether their data agree with the
# model?

ggplot(df, aes(x = factor(Q3a), fill = factor(Q3a))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 2, data agree with the model') +
  theme(legend.position = "none")
```

Group 2, data agree with the model

```r
# Which statements below support your reasoning for whether Group 2's data agree
# with the model?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q145%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                        t.info...cols.temp..
## Q145_8                                   The points are close to the line
## Q145_11        The points are close to the line compared to the uncertainties
## Q145_21            There are equal numbers of points above and below the line
## Q145_2             The points are randomly scattered above and below the line
## Q145_9                                            There are very few outliers
## Q145_36                         The uncertainty in the points is small
## Q145_6     Many of the points cross the line (when considering their uncertainties)
## Q145_1                                   The data are approximately linear
## Q145_38                                              Other (Please describe)
## Q145_39                                    The points are far from the line
## Q145_40        The points are far from the line compared to the uncertainties
## Q145_41        There are not equal numbers of points above and below the line
```

```
## Q145_42 There is a pattern in how the points are scattered above and below the line
## Q145_43                                         There are too many outliers
## Q145_44                           The uncertainty in the points is large
## Q145_45       Too few points cross the line (when considering their uncertainties)
## Q145_46                               The data are curved (not linear)
```

```r
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity') +
  ggtitle('Group 2, reasoning') +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```
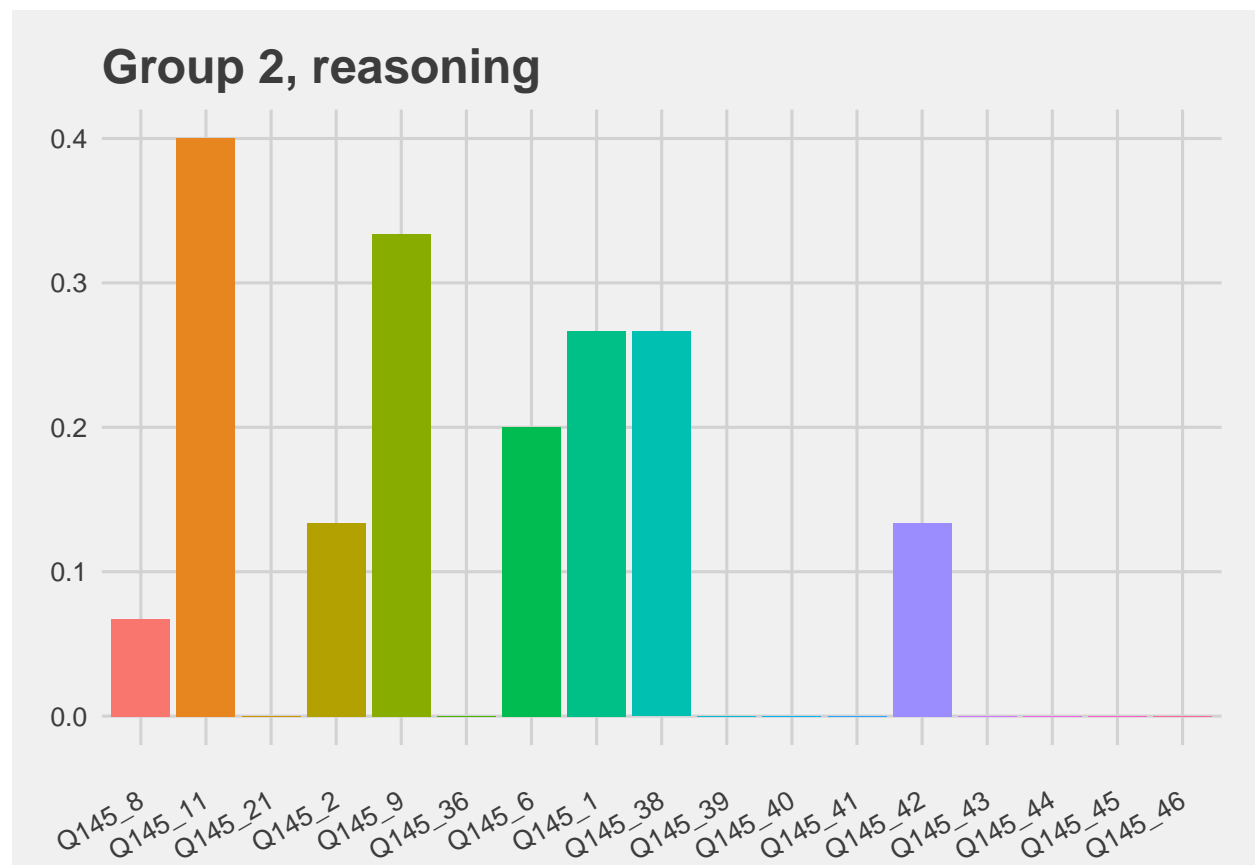


```r
# Which fit do you think Group 2 should use?

ggplot(df, aes(x = factor(Q3c), fill = factor(Q3c))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Group 2, which fit') +
  theme(legend.position = "none")
```

Group 2, which fit
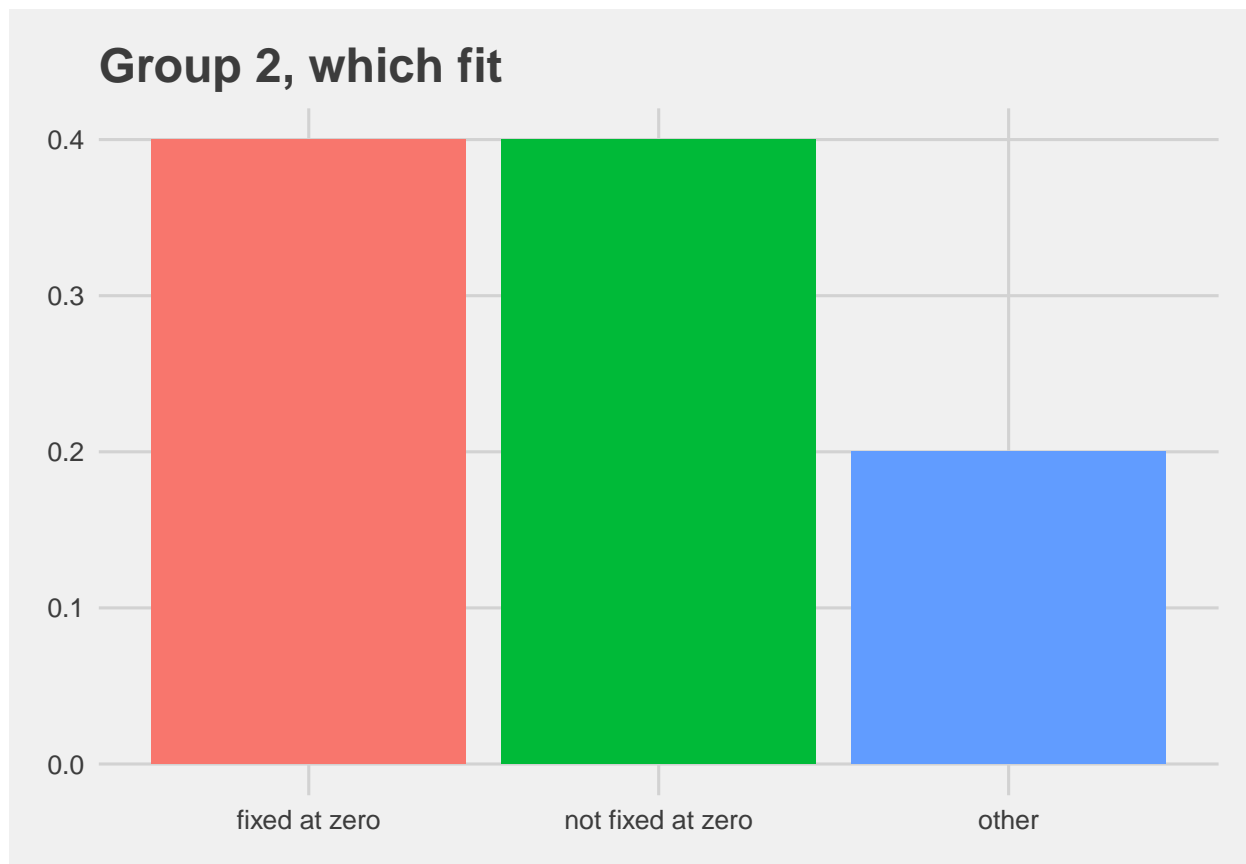

```r
# Others: Either fix the intercept at zero or change the model., both, I think
# the results of both fits should be discussed, including whether there is a
# natural explanation for adding an intercept in a revised model

# Which items below best reflect your reasoning?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q3d%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                           t.info...cols.temp..
## Q3d_8                                                              The fit is better
## Q3d_11                        The uncertainty in the spring constant value is smaller
## Q3d_1                               There may have been mistakes in the data collection
## Q3d_7                                     There may have been random uncertainty
## Q3d_9                                            There may have been human error
## Q3d_5            There may have been an invalid assumption (or something was ignored)
## Q3d_6                                The intercept is irrelevant in the investigation
## Q3d_4                              The ideal model doesn't really apply in the real world
## Q3d_3    The model says the intercept should be zero (and they can't just add a parameter)
## Q3d_29                                                                           Other
```

```r
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity')  +
  ggtitle('Group 2, reasoning for fit') +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```


Group 2, reasoning for fit
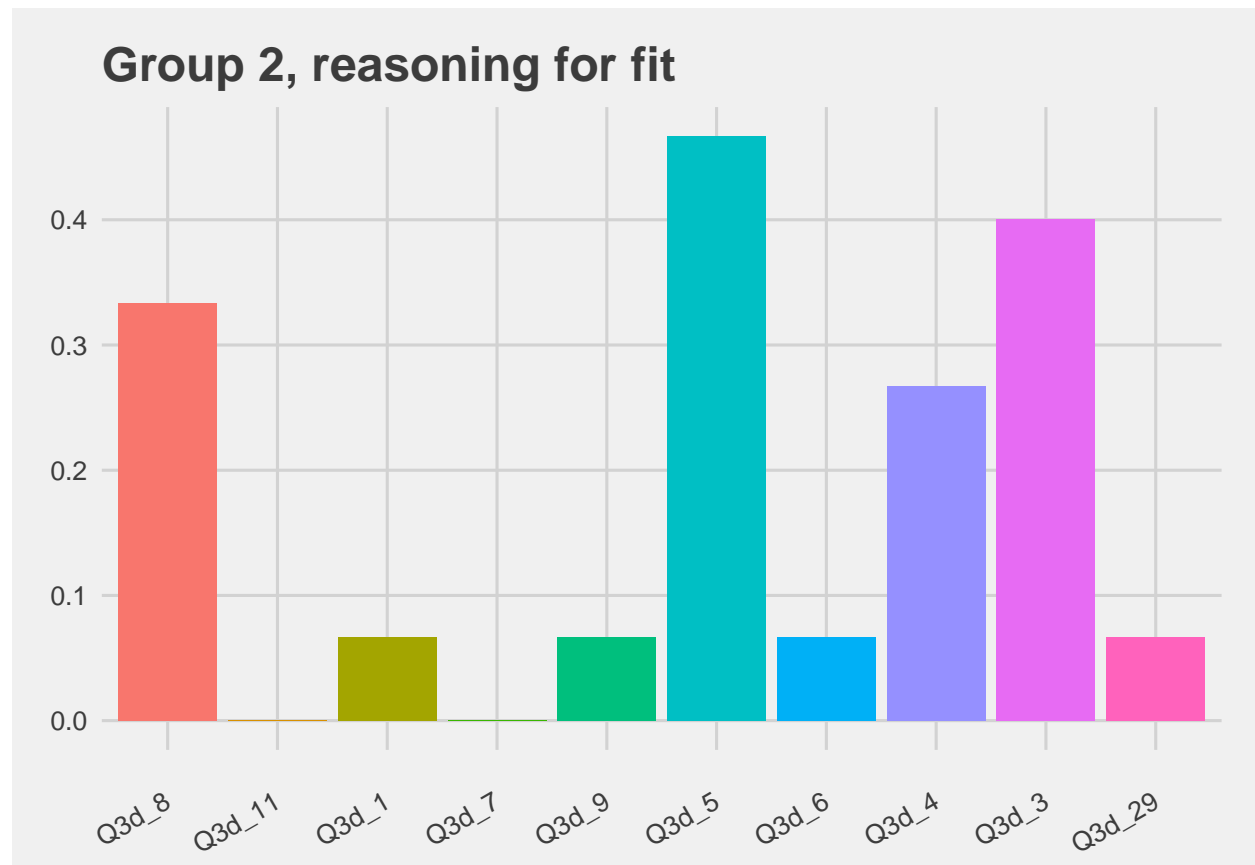
```r
# What do you think Group 2 should do next?

cols.temp <- colnames(df)[(colnames(df) %like% '%Q147%') &
                          !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(value = ifelse(value == 1, 1, 0))

data.frame(t(info[, cols.temp]))
```

```
##                                                              t.info...cols.temp..
## Q147_6                                                 Test or control other variables
## Q147_4                                              Repeat the experiment with better equipment
## Q147_12     Change the analysis (e.g., use a different fit line, incorporate systematic effects)
## Q147_14                                              Compare their k-value to the expected value
## Q147_15                                              Design a new experiment to test the results
## Q147_16 Reduce uncertainty (e.g., more trials for the same masses, more bounces per trial, etc.)
## Q147_17                                              Check their work and write it up
```

```
## Q147_34                              Repeat the experiment with more and different masses
## Q147_11                                                        Other (Please describe)
```

```
ggplot(df.temp, aes(x = variable, fill = variable, y = value/nrow(df))) +
  geom_bar(stat = 'identity')  +
  ggtitle('Group 2, what to do next') +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 30, vjust = 0.5, hjust = 1))
```

Group 2, what to do next
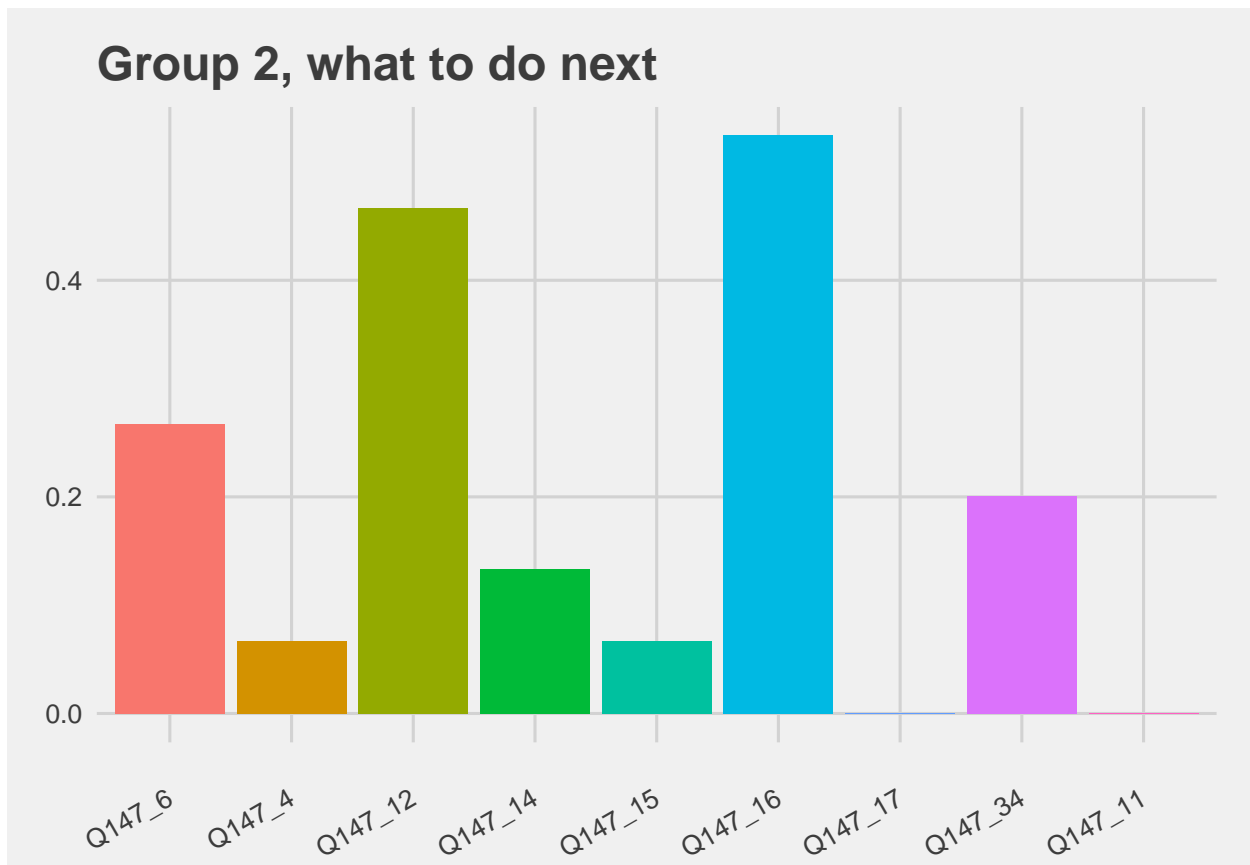
Page 4

```
# Which group do you think did a better job of testing the model?

ggplot(df, aes(x = factor(Q4a), fill = factor(Q4a))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Which group did a better job?') +
  theme(legend.position = "none")
```
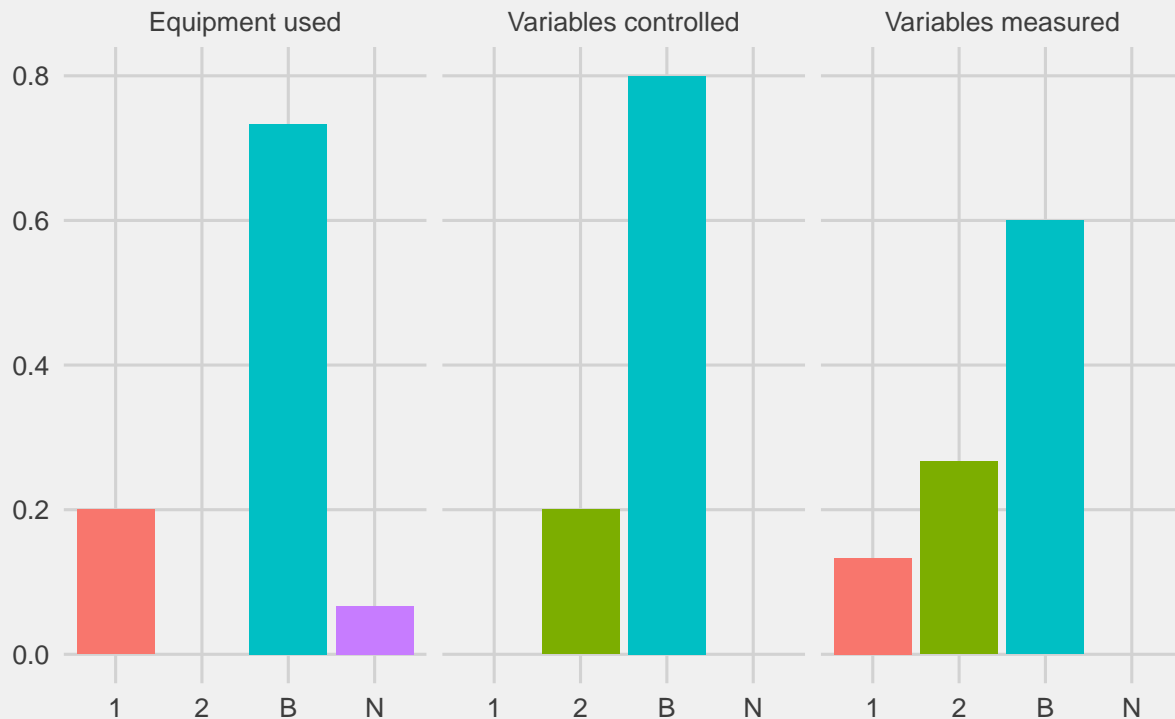
Which group did a better job?
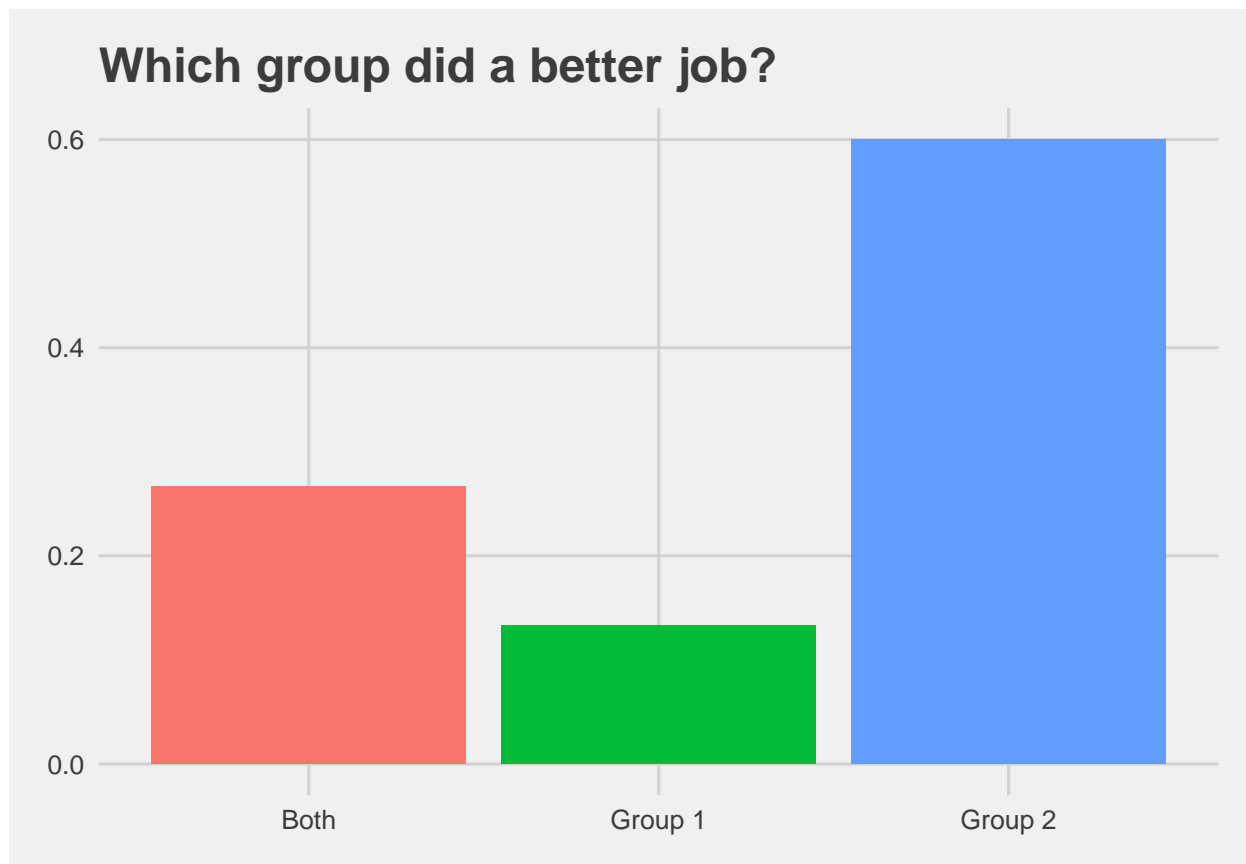

```r
cols.temp <- colnames(df)[(colnames(df) %like% '%152%') &
                            !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q152_1' ~ 'Equipment used',
    variable == 'Q152_2' ~ 'Variables measured',
    variable == 'Q152_3' ~ 'Variables controlled',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Compare data collection methods') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```

Compare data collection methods

| Equipment used | Variables controlled | Variables measured |


```
cols.temp <- colnames(df)[(colnames(df) %like% '%Q153%') &
                           !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q153_1' ~ 'N trials',
    variable == 'Q153_2' ~ 'N masses',
    variable == 'Q153_3' ~ 'N bounces',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Compare amount of data') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```

Compare amount of data
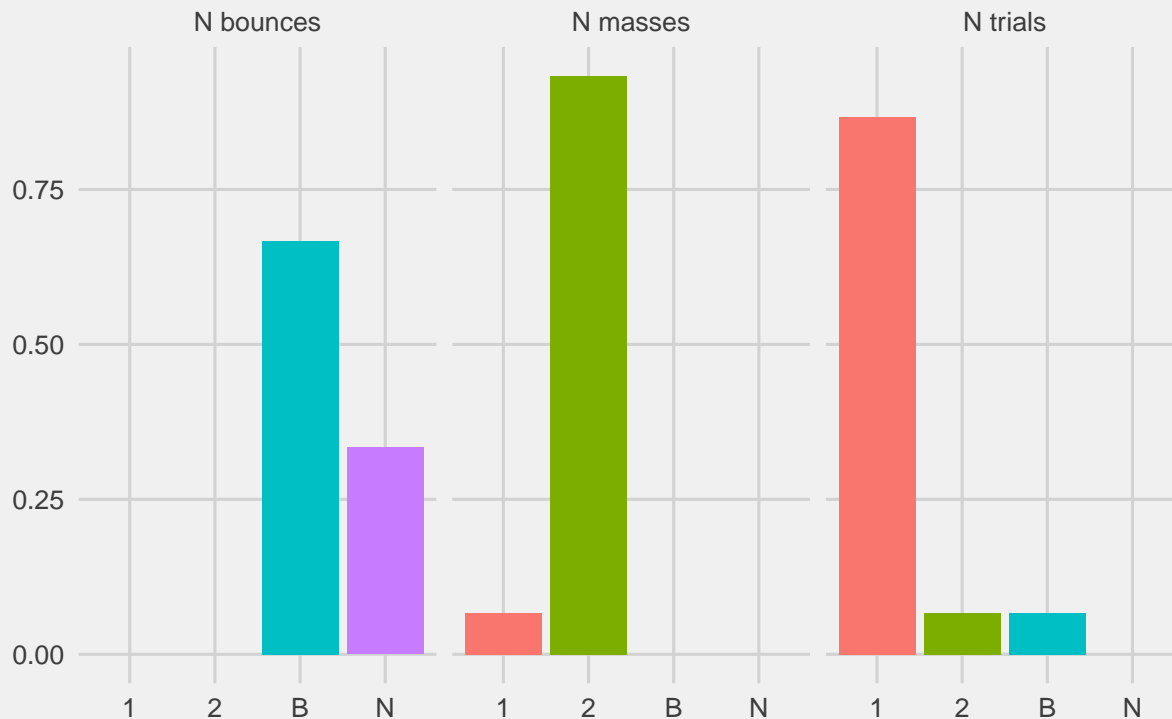
```
cols.temp <- colnames(df)[(colnames(df) %like% '%Q154%') &
                            !(colnames(df) %like% '%TEXT%')]

df.temp <- df[, cols.temp] %>%
  melt(., measure.vars = cols.temp) %>%
  mutate(variable = case_when(
    variable == 'Q154_1' ~ 'Explanations',
    variable == 'Q154_2' ~ 'Analysis',
    variable == 'Q154_3' ~ 'Similar k values',
    variable == 'Q154_4' ~ 'Uncertainty in data and k',
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(variable) & value != '')

ggplot(df.temp, aes(x = factor(value), fill = factor(value))) +
  geom_bar(aes(y = (..count..)/nrow(df))) +
  ggtitle('Compare analysis and outcomes') +
  theme(legend.position = "none") +
  facet_wrap(~variable)
```

Compare analysis and outcomes

Analysis

Explanations

Similar k values

Uncertainty in data and k

```
# What feature was most important to you for comparing the two groups?

ggplot(df, aes(x = factor(Q148), fill = factor(Q148))) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  ggtitle('Feature for comparing methods') +
  theme(legend.position = "none")
```

Feature for comparing methods