```r
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
```

```r
library(magrittr)
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library(MatchIt)
```

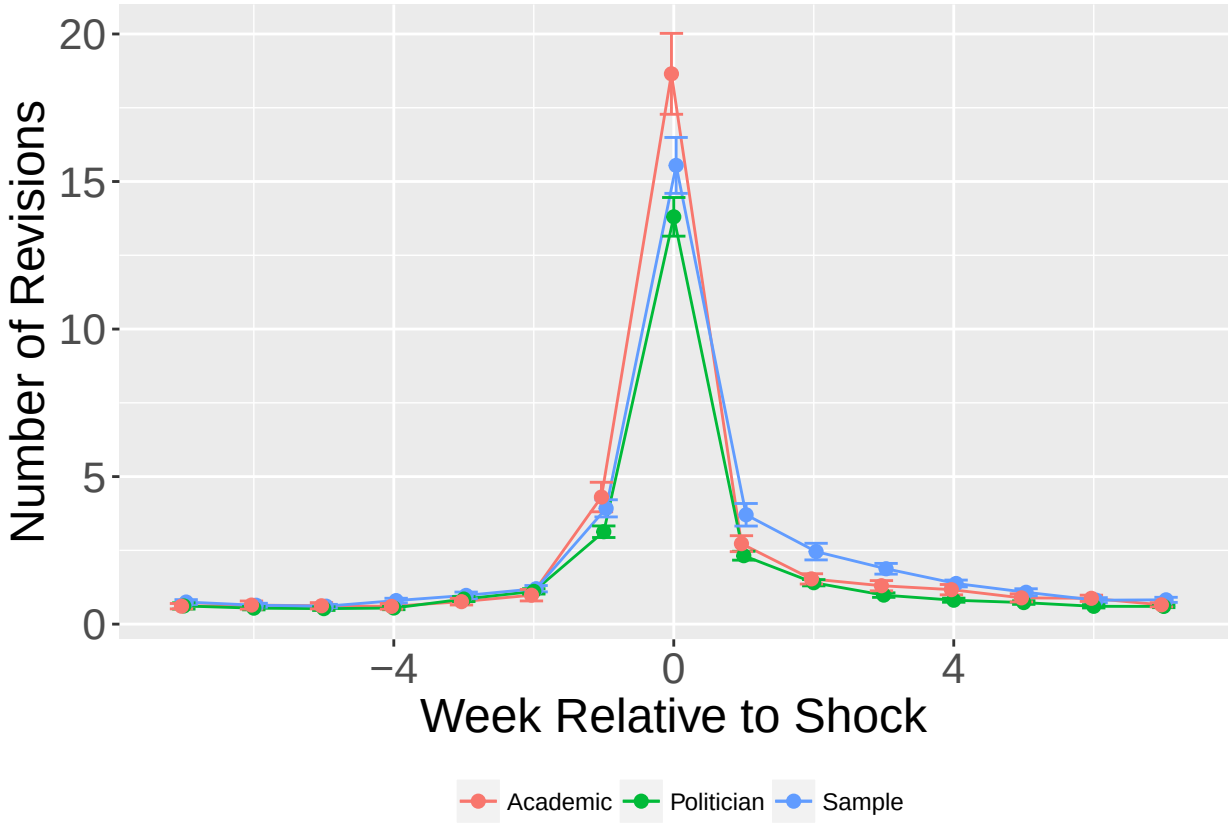
```
## Warning: package 'MatchIt' was built under R version 3.4.3
```

```r
library(sem)
source("/Users/arkzhang/Documents/customized_functions/R/plot_mean_errorbar.R")
source("/Users/arkzhang/Documents/customized_functions/R/plot_median.R")
knitr::opts_knit$set(root.dir = "/Users/arkzhang/Documents/research/shock_wiki_2018")
```

```r
# data preparation
df_main <- read.csv("data/all_treated_main_metric.csv")
df_main <- merge(x = df_main,
                 y = read.csv("data/all_treated_info.csv") %>%
                   dplyr::select(ArticleId, Type),
                 by = "ArticleId")
df_main %<>% mutate(FracRevertedNew = ifelse(NumRevNew == 0, NA, NumRevertedNew / NumRevNew),
                    IsRevertedNew = ifelse(is.na(FracRevertedNew), NA, ifelse(FracRevertedNew > 0, 1, 0
                    FracReverted = ifelse(NumRev == 0, NA, (NumRevertedNew + NumRevertedOld) / NumRev))

# combine with talk page data
df_talk <- rbind(
  read.csv("data/wiki_talk_stats_academics_final.csv"),
  read.csv("data/wiki_talk_stats_politicians_final.csv"),
  read.csv("data/wiki_talk_stats_sample_final.csv"))
df_main <- merge(x = df_main, y = df_talk, by = c("ArticleId", "RelWeek"), all.x = T)
df_main %<>% mutate(total_comment_na_replaced = ifelse(is.na(X.total_comment), 0, X.total_comment),
                    editis_per_neweditor_na_replaced = ifelse(NumEditorNew == 0, NA, ifelse(is.na(X.edi
                    log_talk_main_ratio = ifelse(NumRev == 0, NA, log(1+total_comment_na_replaced) / log

# combine with retention data
df_retention <- read.csv("data/retention.csv")
df_main <- merge(x = df_main, y = df_retention,
                 by = c("ArticleId", "RelWeek"), all.x = T)
```

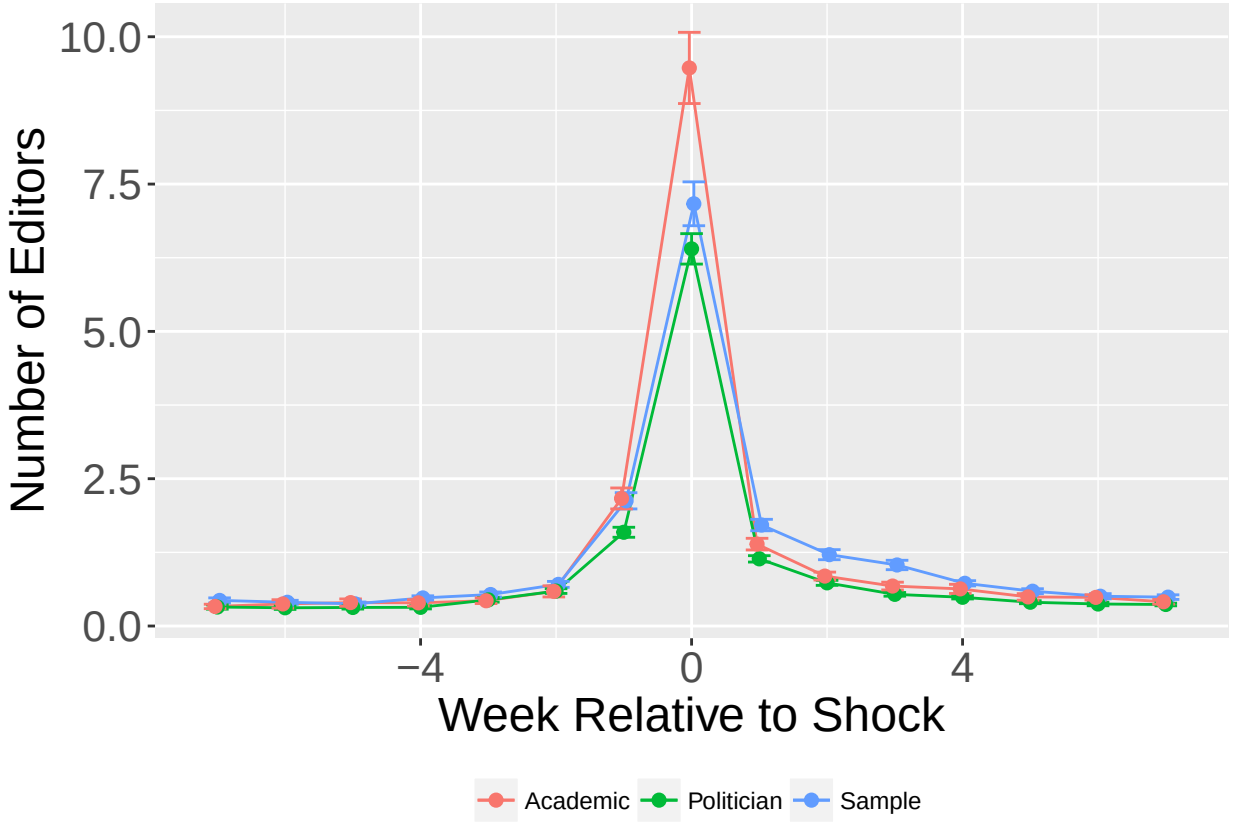
```r
df_main %<>% mutate(FracCumulRevNew = ifelse(NumCumulRev == 0, NA, NumCumulRevNew / NumCumulRev),
                    FracCumulRevOld = ifelse(NumCumulRev == 0, NA, NumCumulRevOld / NumCumulRev))

# combine with retention on all wikipedia
df_main <- merge(x = df_main,
                 y = read.csv("data/all_treated_allwikireten_currentweek.csv") %>%
                   dplyr::select(ArticleId, RelWeek, MeanNewEditorRetenAllWiki, MeanOldEditorRetenAllWi
df_main %<>% mutate(LogMeanNewEditorRetenAllWiki = log(1 + MeanNewEditorRetenAllWiki),
                    LogMeanOldEditorRetenAllWiki = log(1 + MeanOldEditorRetenAllWiki))
```

```r
# plot number of revisions on focal articles (Fig.2a)
plot_numrev <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, NumRev, Type))
plot_numrev <- plot_numrev +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_numrev$labels$x <- "Week Relative to Shock"
plot_numrev$labels$y <- "Number of Revisions"
ggsave("figures/numrev.pdf", plot = plot_numrev,
       width = 8, height = 8, units = "in", device = "pdf")
plot_numrev
```



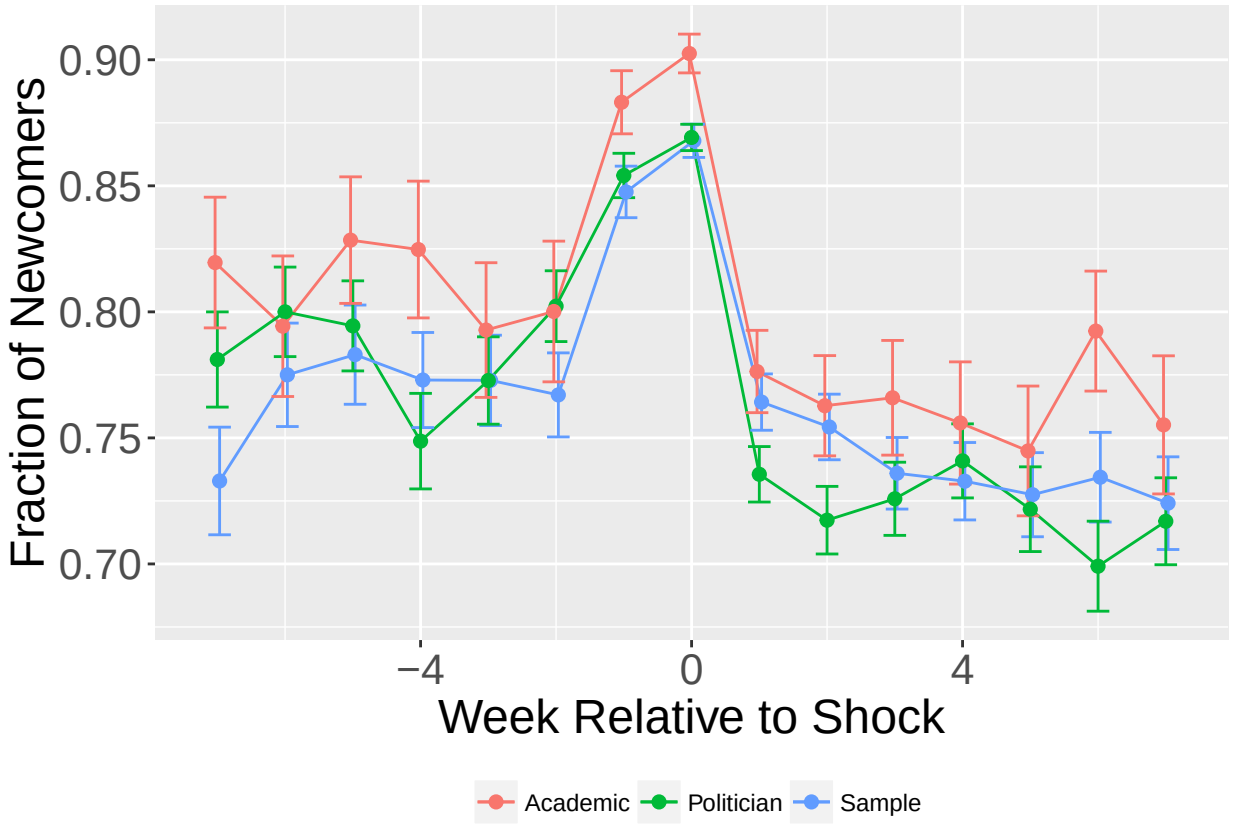

```r
# plot number of editors on focal articles (Fig.2b)
plot_numeditor <- plot_mean_errorbar(
```

```
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, NumEditor, Type))
plot_numeditor <- plot_numeditor +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_numeditor$labels$x <- "Week Relative to Shock"
plot_numeditor$labels$y <- "Number of Editors"
ggsave("figures/numeditor.pdf", plot = plot_numeditor,
       width = 8, height = 8, units = "in", device = "pdf")
plot_numeditor
```

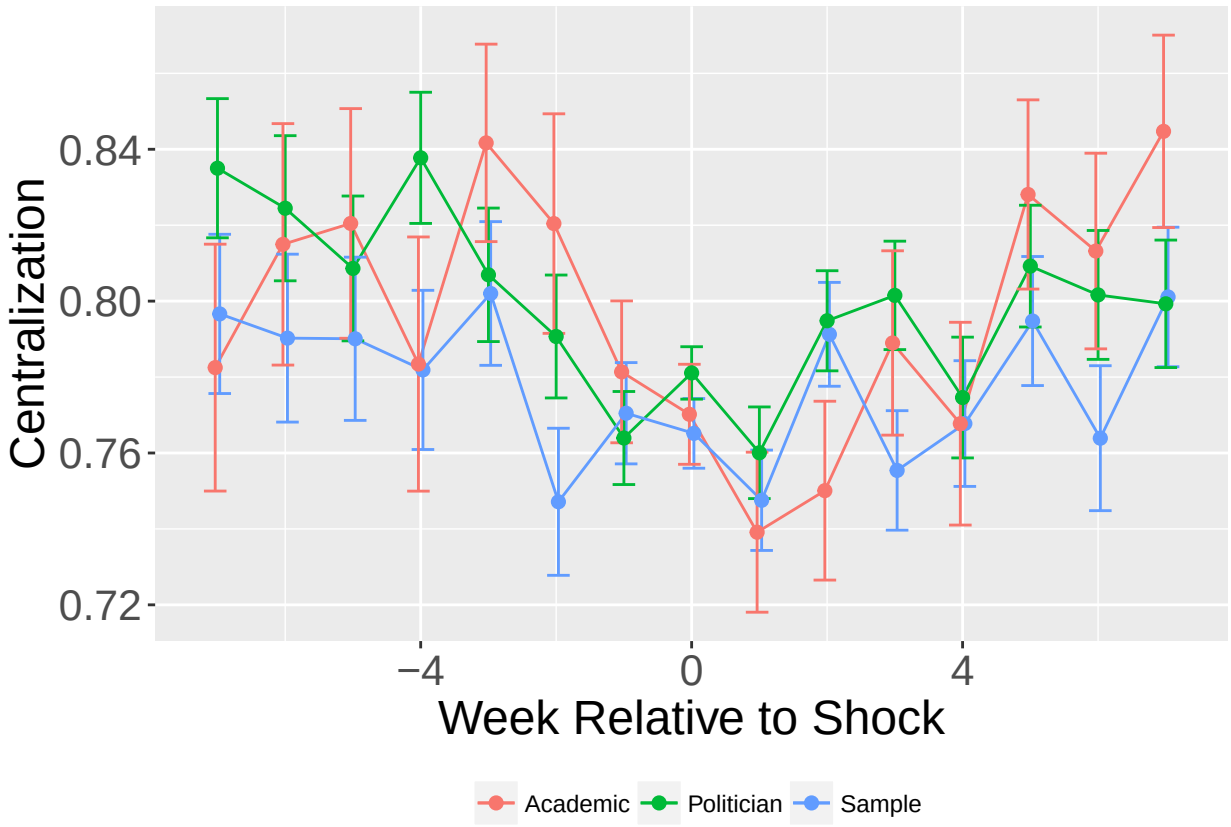



```
# plot fraction of newcomers on focal articles (Fig.3a)
df_main %<>% mutate(FracNewEditor = ifelse(NumEditor == 0, NA, NumEditorNew / NumEditor))
plot_fracnewcomer <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, FracNewEditor, Type))
plot_fracnewcomer <- plot_fracnewcomer +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_fracnewcomer$labels$x <- "Week Relative to Shock"
plot_fracnewcomer$labels$y <- "Fraction of Newcomers"
ggsave("figures/frac_newcomer.pdf", plot = plot_fracnewcomer,
       width = 8, height = 8, units = "in", device = "pdf")
```

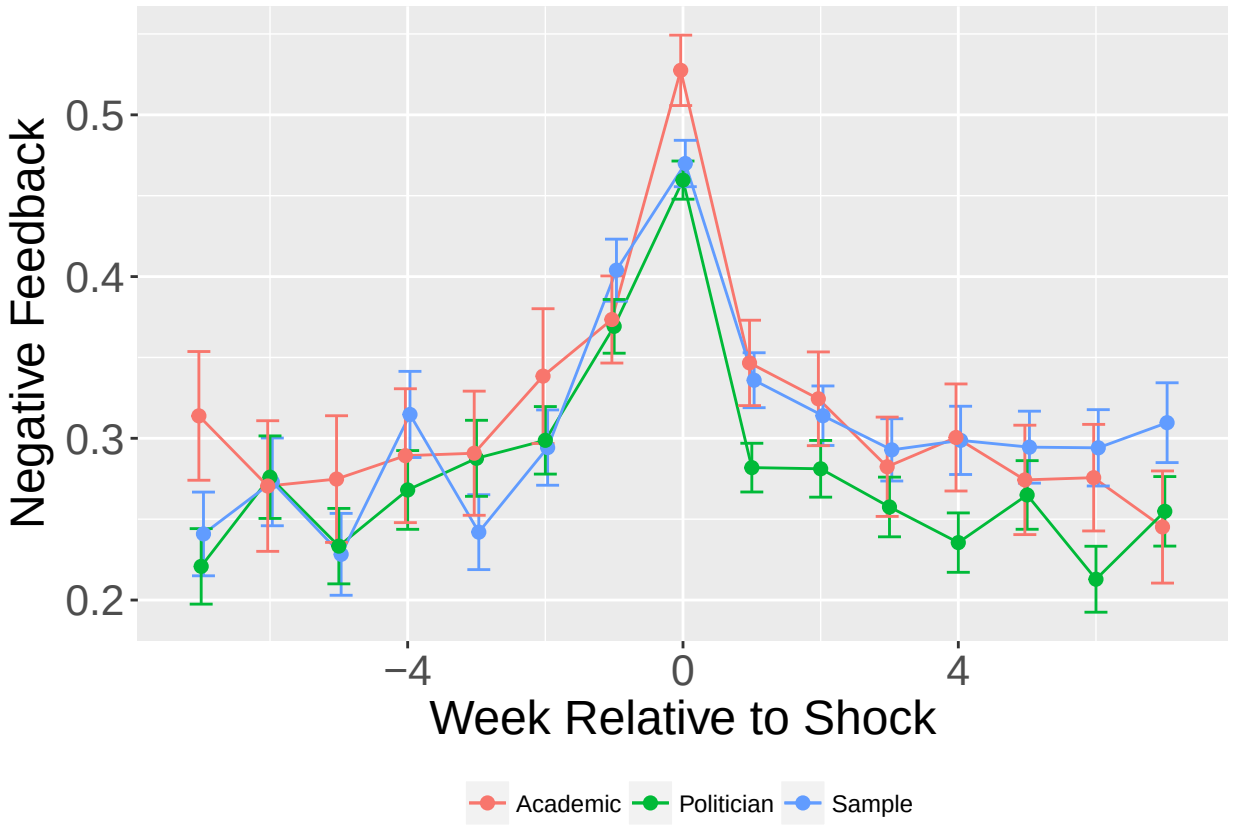
```
plot_fracnewcomer
```



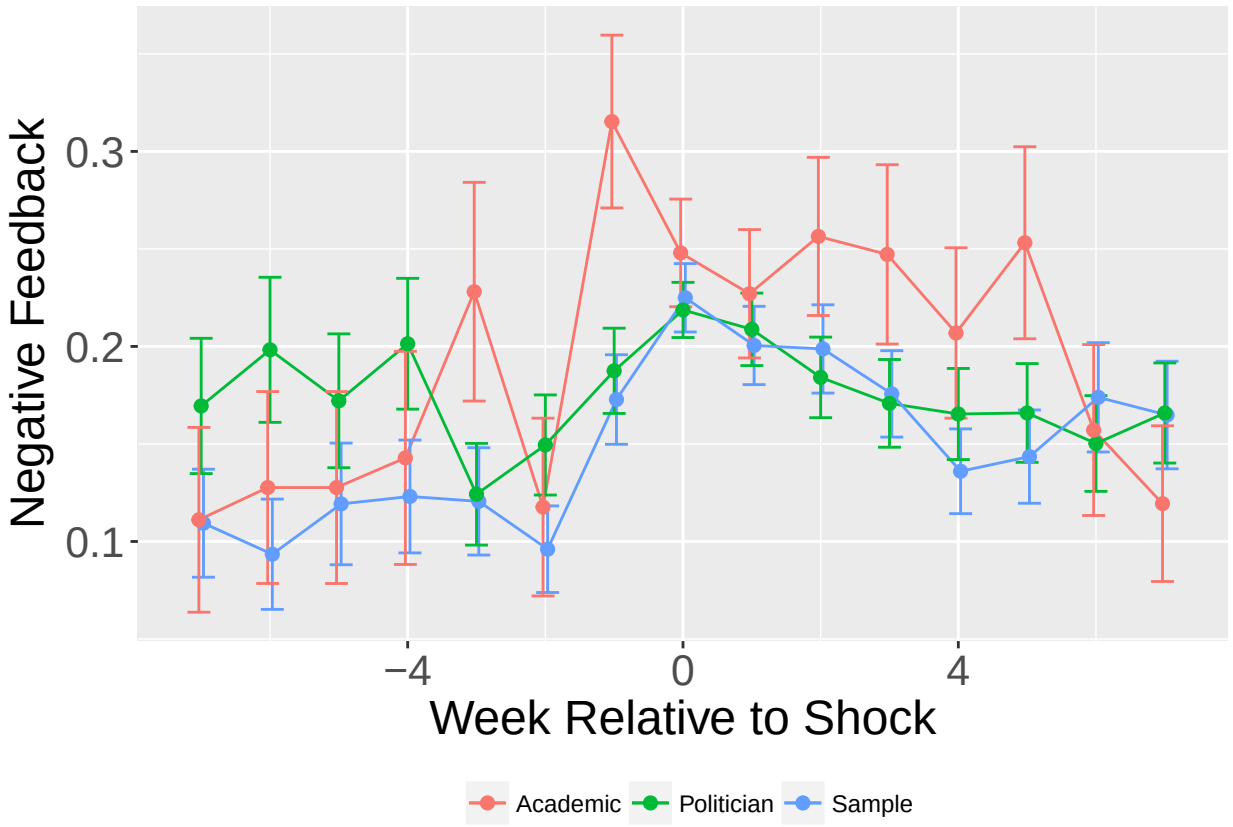

```
# plot gini on focal articles (Fig.3b)
plot_gini <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, Gini, Type))
plot_gini <- plot_gini +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_gini$labels$x <- "Week Relative to Shock"
plot_gini$labels$y <- "Centralization"
ggsave("figures/gini.pdf", plot = plot_gini,
       width = 8, height = 8, units = "in", device = "pdf")
plot_gini
```



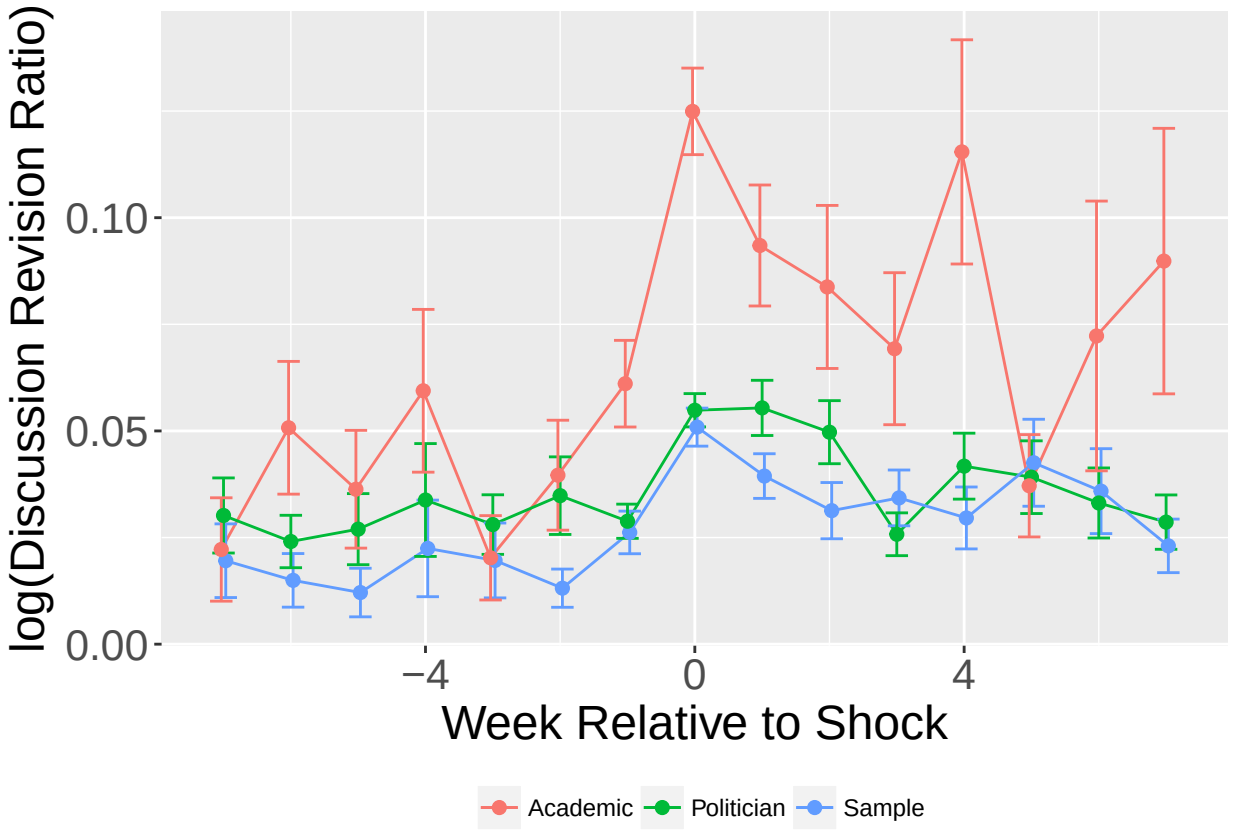

```
# plot negative feedback on new comers (Fig.4a)
plot_negfb_new <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, IsRevertedNew, Type))
plot_negfb_new <- plot_negfb_new +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_negfb_new$labels$x <- "Week Relative to Shock"
plot_negfb_new$labels$y <- "Negative Feedback"
ggsave("figures/negfb_new.pdf", plot = plot_negfb_new,
       width = 8, height = 8, units = "in", device = "pdf")
plot_negfb_new
```



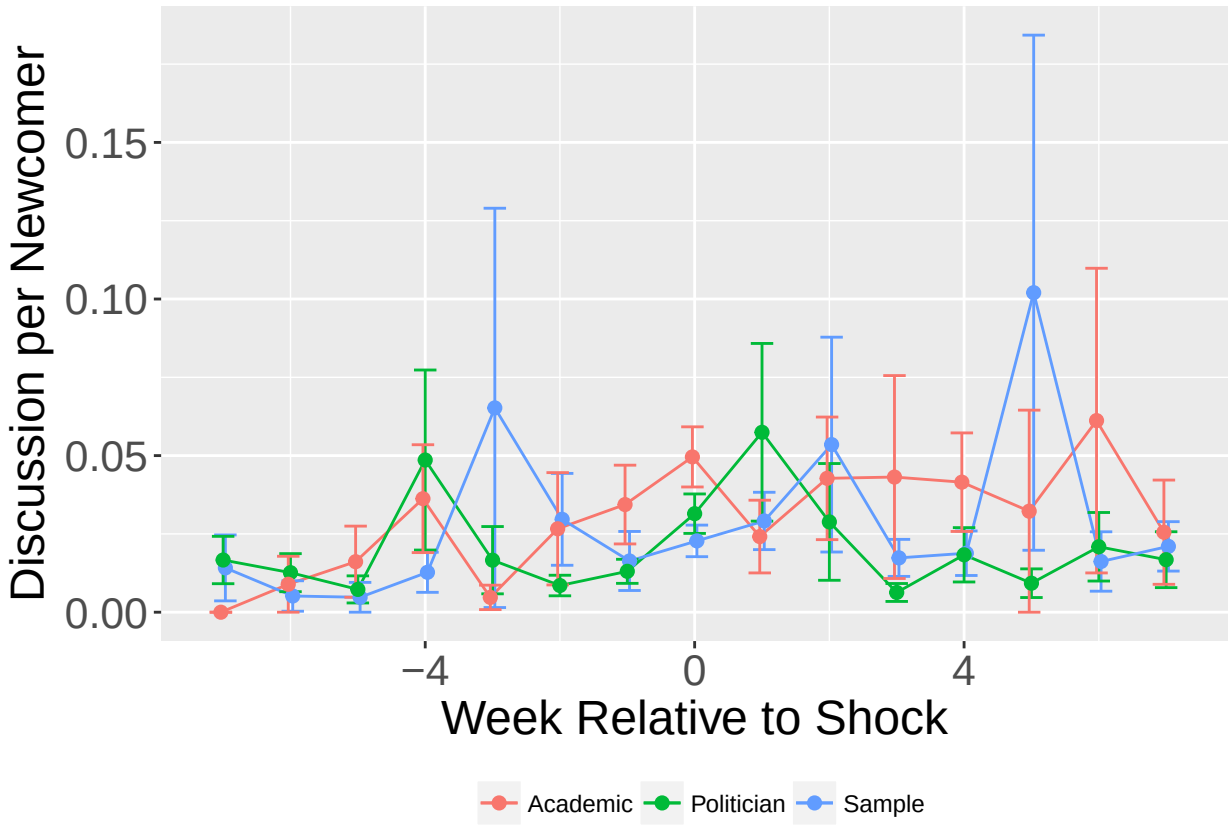

```
# plot negative feedback on incumbents (Fig.4b)
df_main %<>%
  mutate(FracRevertedOld = ifelse(NumRevOld == 0, NA, NumRevertedOld / NumRevOld),
         IsRevertedOld = ifelse(is.na(FracRevertedOld), NA, ifelse(FracRevertedOld > 0, 1, 0)))
plot_negfb_old <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, IsRevertedOld, Type))
plot_negfb_old <- plot_negfb_old +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_negfb_old$labels$x <- "Week Relative to Shock"
plot_negfb_old$labels$y <- "Negative Feedback"
ggsave("figures/negfb_old.pdf", plot = plot_negfb_old,
       width = 8, height = 8, units = "in", device = "pdf")
plot_negfb_old
```



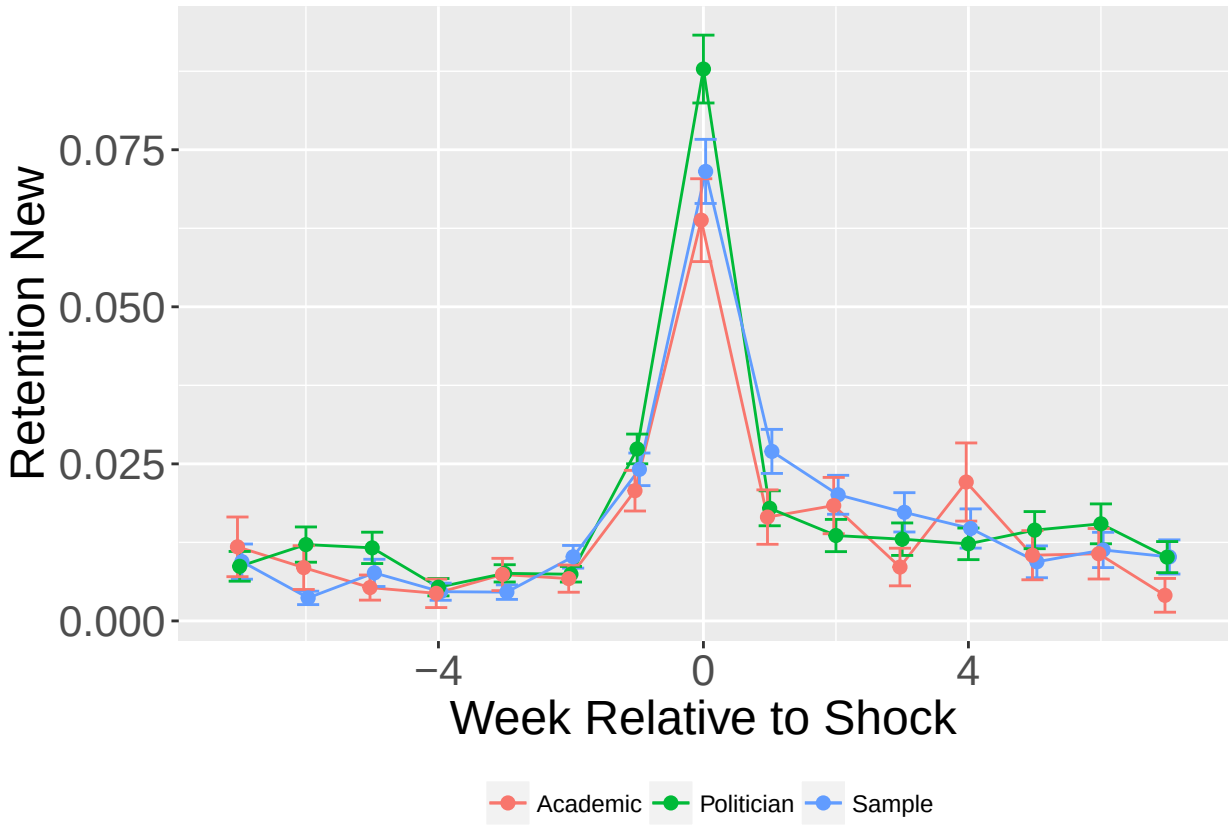

```
# plot talk_main ratio (Fig.5a)
plot_talk_main_ratio <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, log_talk_main_ratio, Type))
plot_talk_main_ratio <- plot_talk_main_ratio +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_talk_main_ratio$labels$x <- "Week Relative to Shock"
plot_talk_main_ratio$labels$y <- "log(Discussion Revision Ratio)"
ggsave("figures/talk_main_ratio.pdf", plot = plot_talk_main_ratio,
       width = 8, height = 8, units = "in", device = "pdf")
plot_talk_main_ratio
```



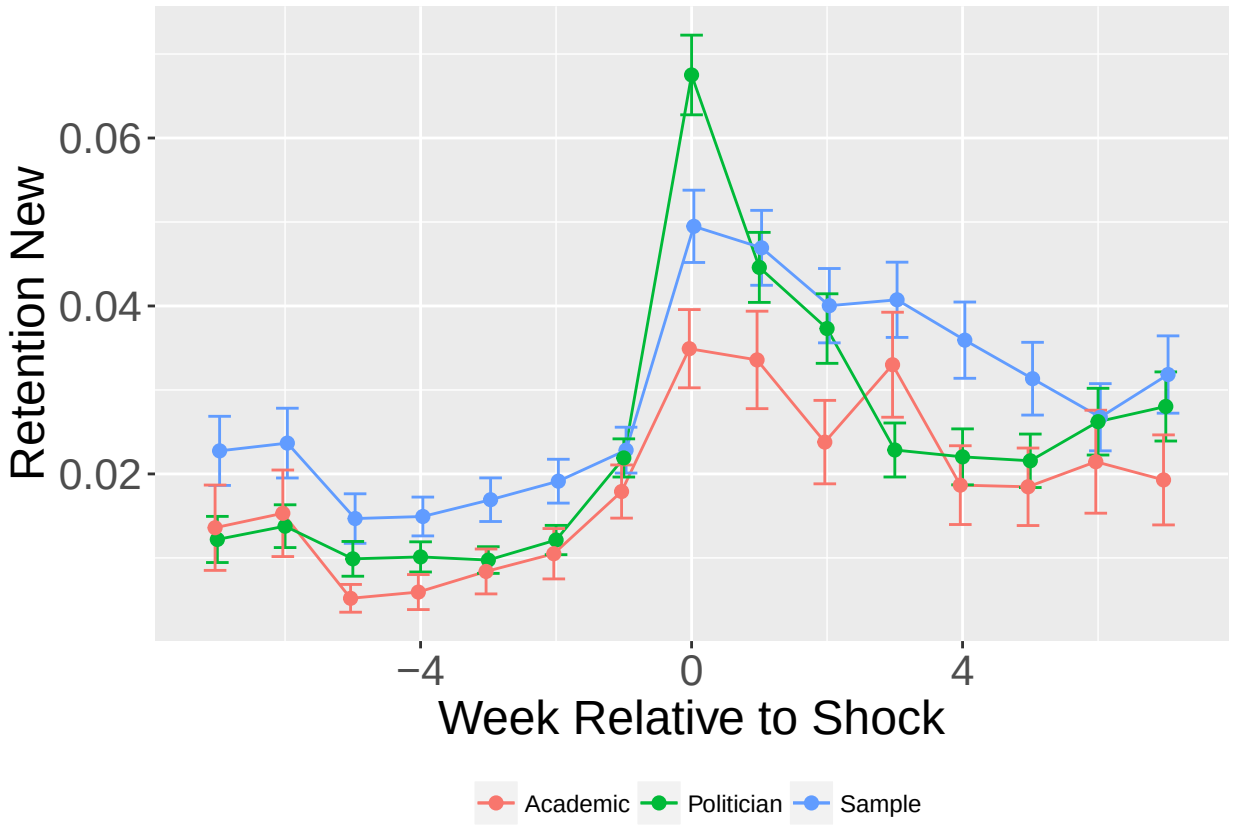

```
# plot fraction of edits on talk page per newcomer (Fig.5b)
plot_avg_comment <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, X.editis_per_neweditor, Type))
plot_avg_comment <- plot_avg_comment +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_avg_comment$labels$x <- "Week Relative to Shock"
plot_avg_comment$labels$y <- "Discussion per Newcomer"
ggsave("figures/avgcomment.pdf", plot = plot_avg_comment,
       width = 8, height = 8, units = "in", device = "pdf")
plot_avg_comment
```



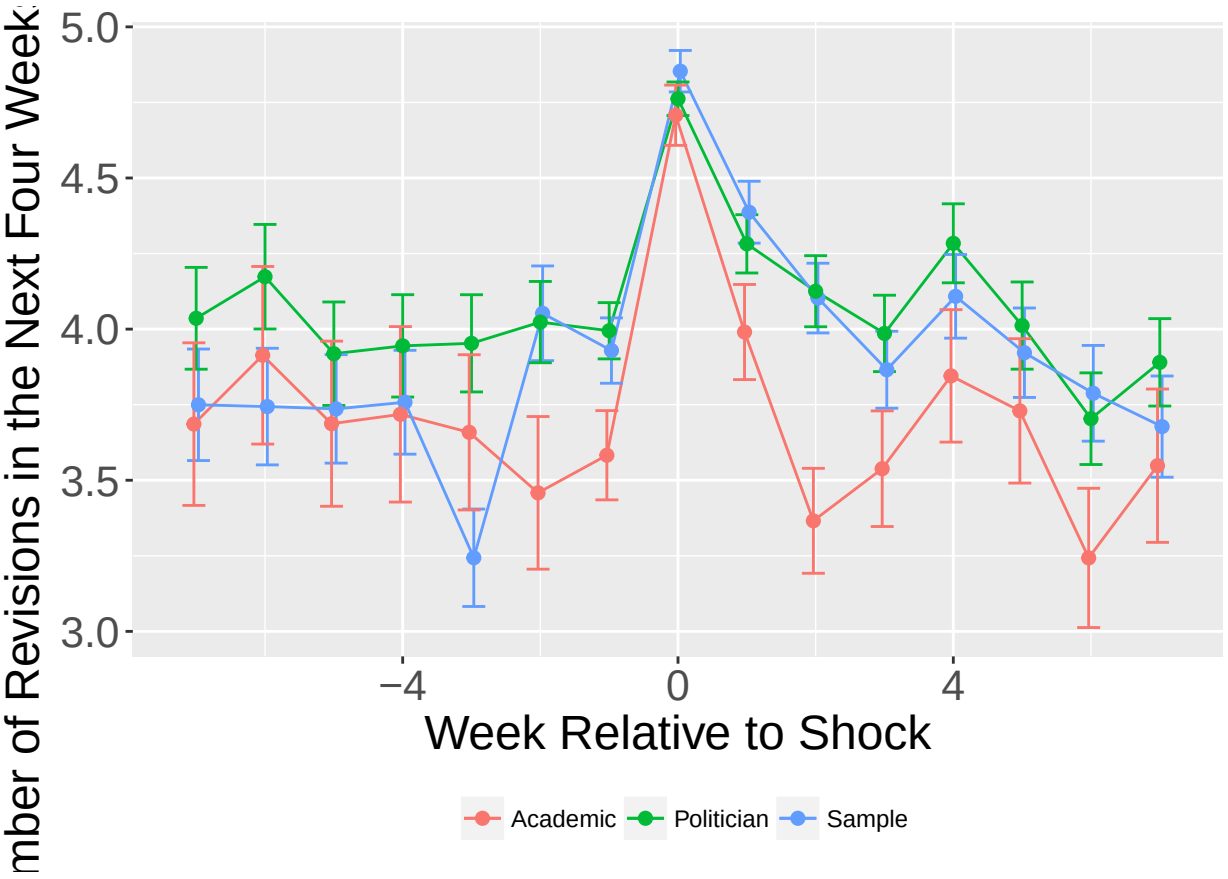

```
# plot cumulative retention for newcomers (Fig.6a)
plot_cumulrevnew <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, FracCumulRevNew, Type))
plot_cumulrevnew <- plot_cumulrevnew +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_cumulrevnew$labels$x <- "Week Relative to Shock"
plot_cumulrevnew$labels$y <- "Retention New"
ggsave("figures/cumulrevnew.pdf", plot = plot_cumulrevnew,
       width = 8, height = 8, units = "in", device = "pdf")
plot_cumulrevnew
```



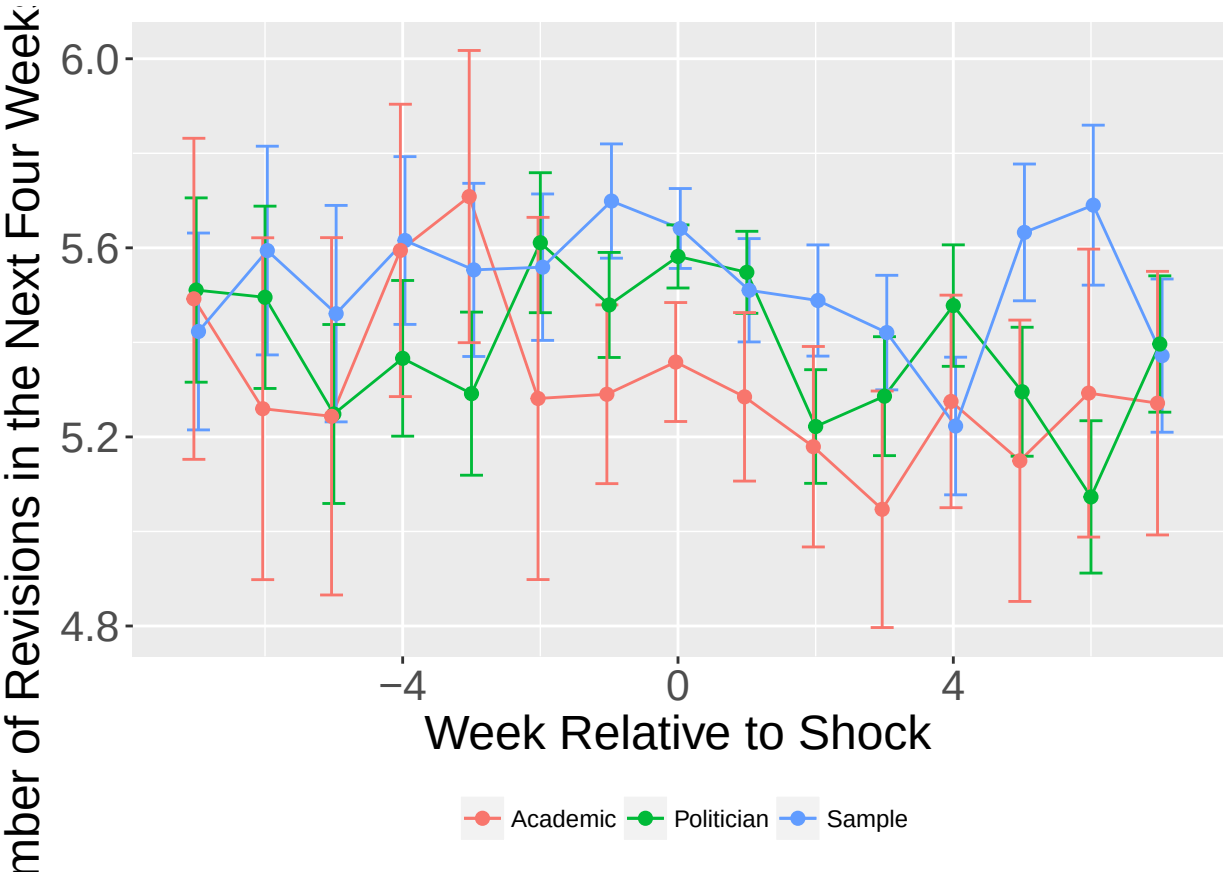

```
# plot cumulative retention for incumbents (Fig.6a)
plot_cumulrevold <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, FracCumulRevOld, Type))
plot_cumulrevold <- plot_cumulrevold +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_cumulrevold$labels$x <- "Week Relative to Shock"
plot_cumulrevold$labels$y <- "Retention New"
ggsave("figures/cumulrevnew.pdf", plot = plot_cumulrevnew,
       width = 8, height = 8, units = "in", device = "pdf")
plot_cumulrevold
```




```
# plot retention over all wikipedia for newcomers (Fig.7a)
plot_logallwikiretennew <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, LogMeanNewEditorRetenAllWiki, Type))
plot_logallwikiretennew <- plot_logallwikiretennew +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_logallwikiretennew$labels$x <- "Week Relative to Shock"
plot_logallwikiretennew$labels$y <- "Log of Number of Revisions in the Next Four Weeks over All Wikipedi
ggsave("figures/logallwikiretennew.pdf", plot = plot_logallwikiretennew,
       width = 8, height = 8, units = "in", device = "pdf")
plot_logallwikiretennew
```




```
# plot retention over all wikipedia for incumbents (Fig.7b)
plot_logallwikiretenold <- plot_mean_errorbar(
  df_main %>% filter(RelWeek >= -7 & RelWeek <= 7)
     %>% dplyr::select(RelWeek, LogMeanOldEditorRetenAllWiki, Type))
plot_logallwikiretenold <- plot_logallwikiretenold +
  theme(legend.position="bottom",
        axis.text=element_text(size=16),
        axis.title=element_text(size=18)) +
  labs(color = "")
plot_logallwikiretenold$labels$x <- "Week Relative to Shock"
plot_logallwikiretenold$labels$y <- "Log of Number of Revisions in the Next Four Weeks over All Wikipedi
ggsave("figures/logallwikiretenold.pdf", plot = plot_logallwikiretenold,
       width = 8, height = 8, units = "in", device = "pdf")
plot_logallwikiretenold
```


mber of Revisions in the Next Four Week
6.0
5.6
5.2
4.8
−4
0
4
Week Relative to Shock
Academic
Politician
Sample