Severity by Specialty

Robert Schnitman

December 30, 2019

1. Setup

```r
libs <- c('foreign', 'tidyverse', 'knitr', 'kableExtra', 'magrittr')
lapply(libs, library, character.only = TRUE)

df <- as.data.frame(read.spss('backup.sav'))

df2 <- df %>% filter(HighestEdu != 5) # For some reason, there is a 5 in education.

df2$PatientType <- with(df2, ifelse(PatientType == 'Outpatient', 'Group1', 'Group2'))

specialties <- c('Specialty 1', 'Specialty 2', 'Specialty 3',
                 'Specialty 4', 'Specialty 5')

df2$Specialty <- sample(specialties, 447, replace = TRUE)
```

2. Average Severity by Specialty

```r
ggplot(filter(df2, is.na(Specialty) == FALSE)) +
  aes(y = as.numeric(Severity),
      x = as.character(Specialty),
      fill = as.character(Specialty)) +
  stat_summary(fun.y = 'mean', geom = 'bar', position = 'dodge', show.legend = FALSE) +
  facet_wrap(~ PatientType) +
  labs(title = 'Fig. 1 - Average Severity by Specialty',
       y     = 'Mean Severity Score',
       x     = 'Specialty',
       caption = 'Note: Specialties have been de-identified for this report.') +
  expand_limits(y = c(1, 5)) +     # Severity score is from 1 to 5.
  coord_flip() +
  theme_light() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major.y = element_blank())
```

Fig. 1 – Average Severity by Specialty
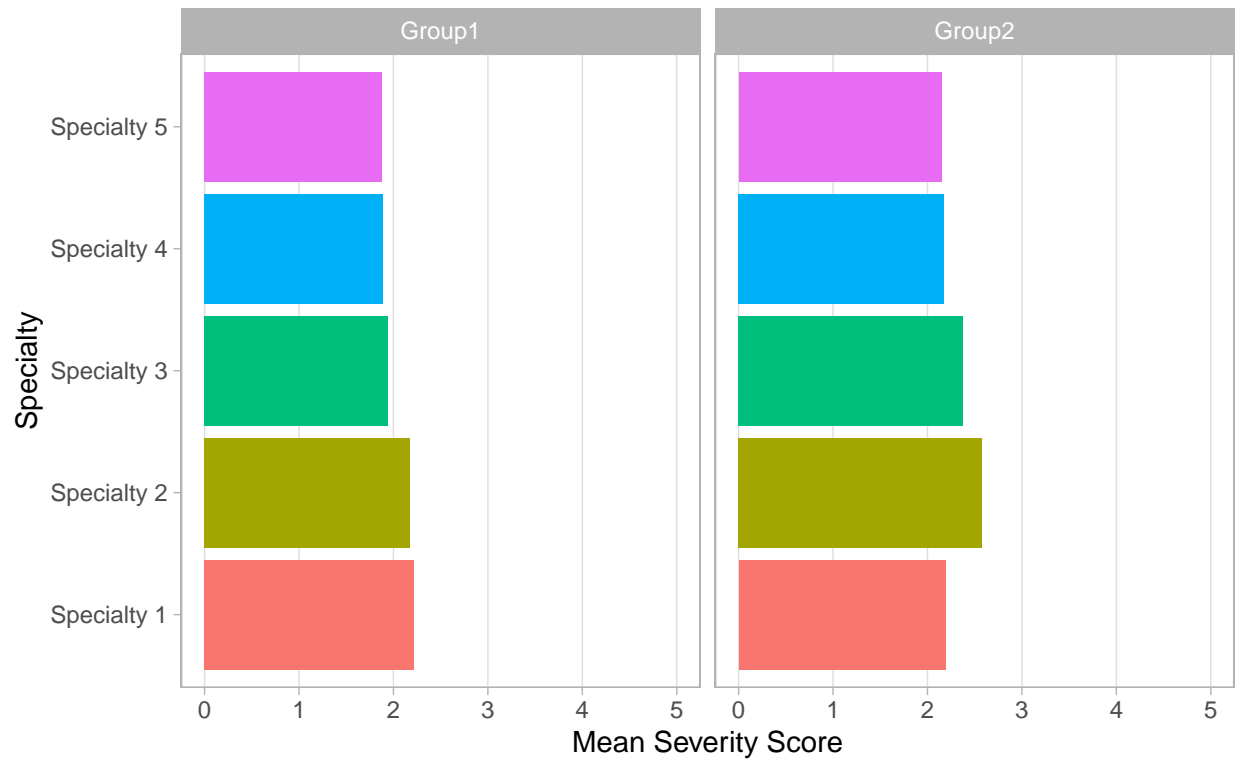
Note: Specialties have been de−identified for this report.

3. Regression Model

```r
m <- lm(data = df2, as.numeric(Severity) ~ factor(Specialty) + Age)

coefs <- coef(summary(m))

rownames(coefs) %<>% gsub('factor\\(Specialty\\)', '', .)

kable(coefs, booktabs = TRUE, linesep = '') %>%
  kable_styling(full_width = TRUE)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.8544163	0.1935964	14.7441583	0.0000000
Specialty 2	0.1216830	0.1511954	0.8048062	0.4213652
Specialty 3	-0.1320964	0.1579082	-0.8365388	0.4033049
Specialty 4	-0.2268491	0.1553737	-1.4600222	0.1449960
Specialty 5	-0.2444717	0.1518493	-1.6099624	0.1081216
Age	-0.0124156	0.0031210	-3.9781121	0.0000812

4. Model Diagnostics
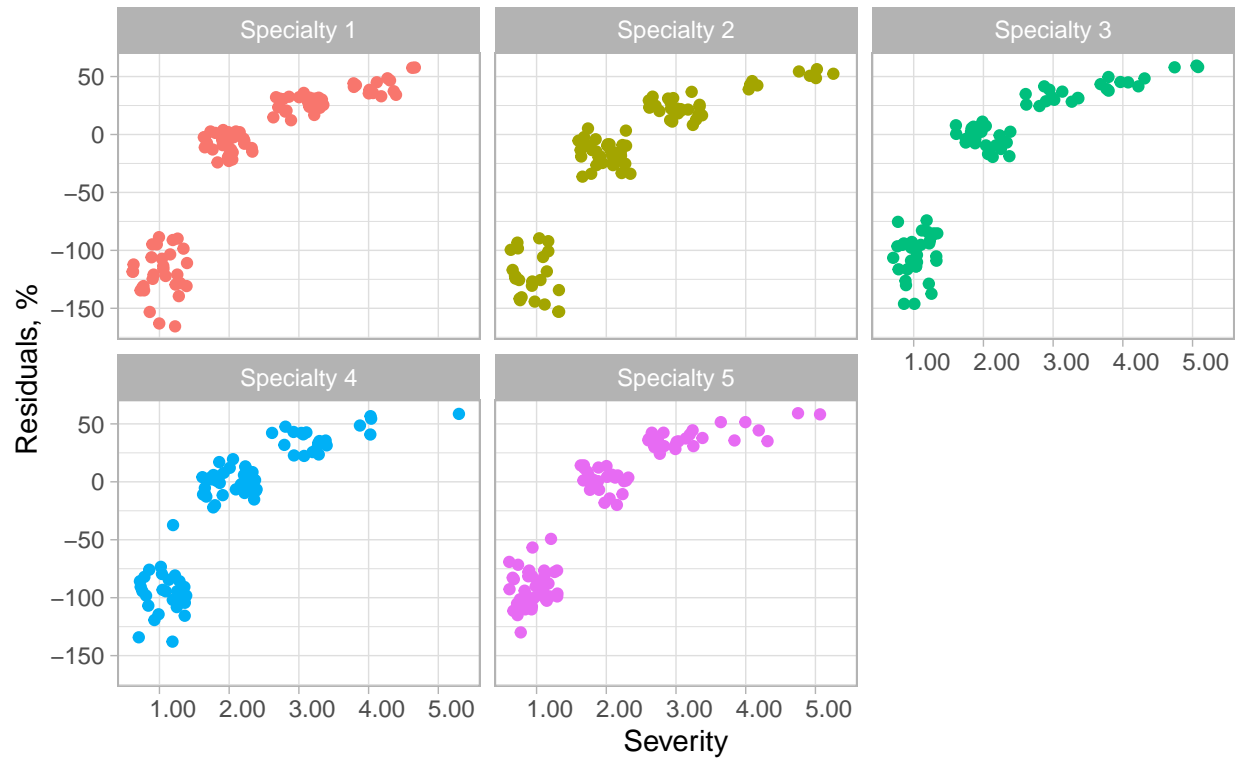
```r
df3 <- mutate(df2,
              fit = predict(m),
              res = resid(m),
              res_pct = res/(fit + res)) # Residuals %

ggplot(df3) +
  aes(y = Severity, x = Age, color = Specialty) +
  geom_jitter(show.legend = FALSE) +
  facet_wrap(~ Specialty) +
  theme_light() +
  labs(title = 'Fig. 2 - Severity vs. Age by Specialty',
       caption = 'Note: Specialties have been de-identified for this report.')
```

Fig. 2 – Severity vs. Age by Specialty

Note: Specialties have been de−identified for this report.
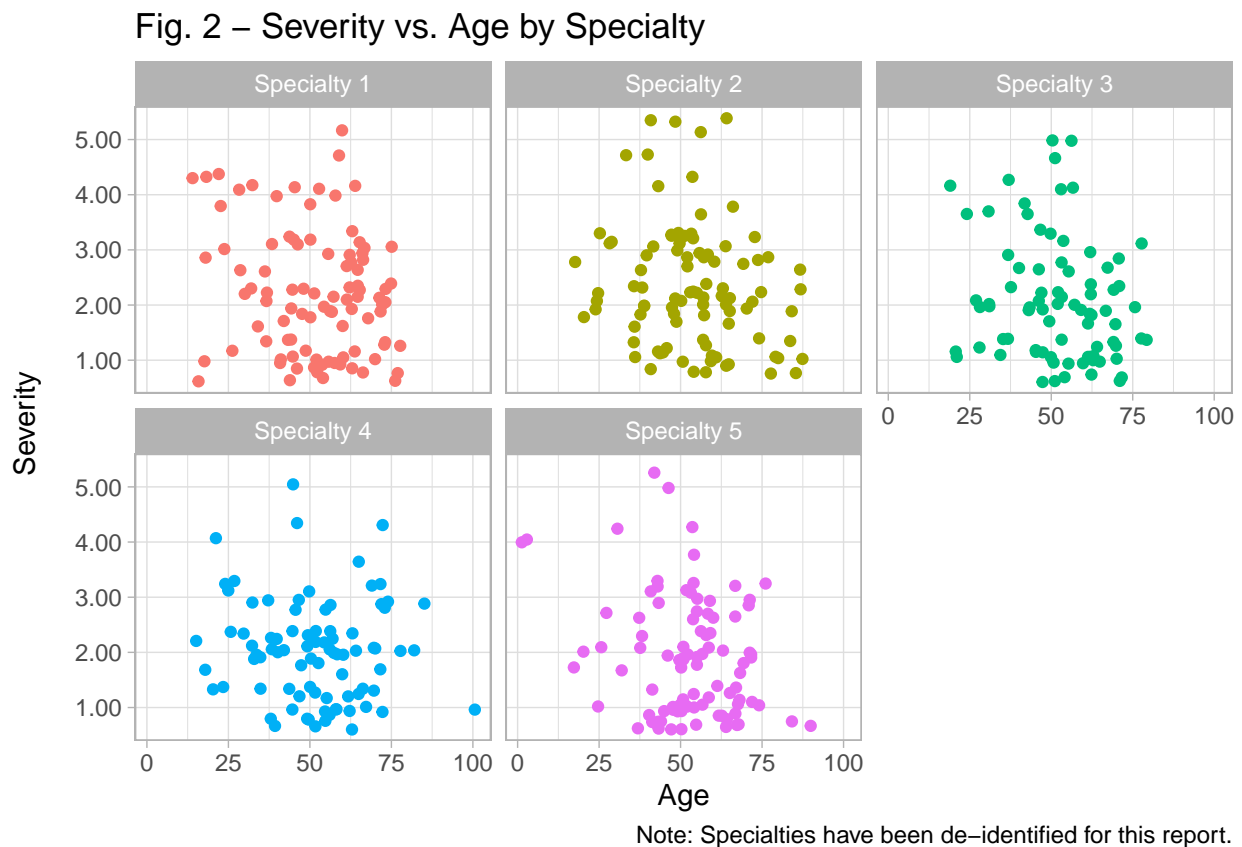
```
ggplot(df3) +
  aes(y = res_pct*100, x = Severity, color = Specialty) +
  geom_jitter(show.legend = FALSE) +
  facet_wrap(~ Specialty) +
  theme_light() +
  labs(title = 'Fig. 3 - Residuals Analysis',
       caption = 'Note: Specialties have been de-identified for this report.',
       y = 'Residuals, %')
```

Fig. 3 – Residuals Analysis

Note: Specialties have been de−identified for this report.