Homework 8

Enter your name and EID here

Connor Hanna

cdh3663

This homework is due on April 11, 2022 at 11:00am. Please submit as a pdf file on Canvas.

Problem 1: (6 pts) Use the dataset `BA_degrees` but focus only on the degree fields "Agriculture and natural resources", "Communication, journalism, and related programs", and "Education".

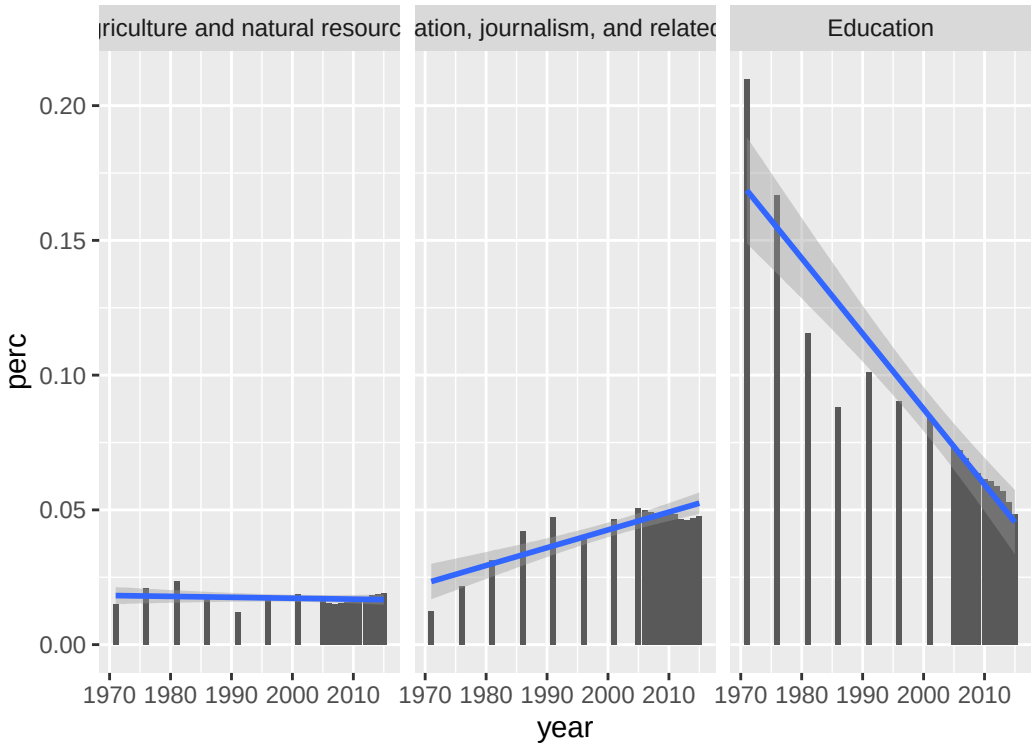
Make a single plot that satisfies these three criteria:

(a) Plot a time series of the proportion of degrees (colum `perc`) in each field over time and create a separate panel per degree field.
(b) Add a straight line fit to each panel.
(c) Order the panels by the difference between the maximum and the minimum proportion (i.e., the range of the data).

```
#filtering to the desired fields
list <- c("Agriculture and natural resources",
                      "Communication, journalism, and related programs",
                      "Education")
BA_degrees <-
  BA_degrees |>
    filter(field %in% list)

#plot generation
BA_degrees |>
  ggplot(
    aes(year, perc)
  ) +
  geom_col() +
  facet_wrap(vars(field)) +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Problem 2: (4 pts) Create a single pipeline that fits a linear model to each of the three fields from Problem 1 and outputs results in a tidy linear model summary table. The first column of the table should be `field` and the remaining columns should contain the linear model summary statistics such as `r.squared` for each field. Display the resulting table below.

```
BA_degrees |>
  nest(data = -field) |>
  mutate(
    fit = map(data, ~lm(year ~ perc, data = .x)),
    glance_out = map(fit, glance)
  ) |>
  select(field, glance_out) |>
  unnest(cols = glance_out)
```

```
## # A tibble: 3 x 13
##   field r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##   <chr>     <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1 Agri~    0.0342       -0.0262 14.1      0.566 4.63e-1     1  -72.1  150.  153.
## 2 Comm~    0.749         0.734   7.16    47.8   3.49e-6     1  -59.9  126.  129.
## 3 Educ~    0.857         0.848   5.41    96.0   3.63e-8     1  -54.9  116.  118.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```