DataVis Challenge

Randy Boyes

2/17/2019

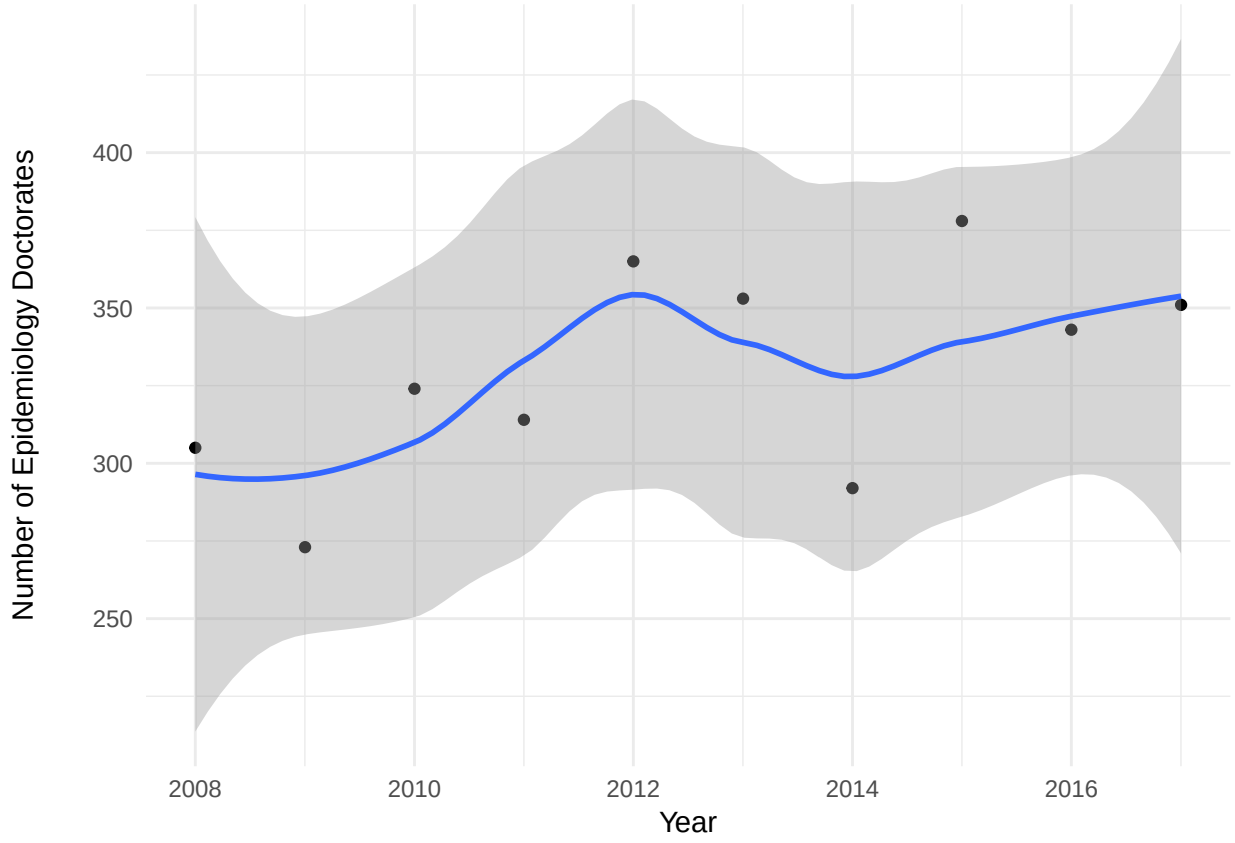
```
years <- list()


for(i in 1:(ncol(tab013) - 1)){
  temp <- data.frame(tab013$`Fine field of study`, tab013[, (i+1)])
  temp$Year <- colnames(temp)[2] %>% substring(2) %>% paste0("-01-01") %>% as.Date(format = "%Y-%m-%d")
  colnames(temp) <- c("Field", "Grads", "Year")
  temp$Grads <- temp$Grads %>% as.numeric()
  years[[i]] <- temp
}

Clean_Data <- do.call(rbind.data.frame, years)
head(Clean_Data)
```

```
##                                         Field Grads       Year
## 1                                 All fields 48777 2008-01-01
## 2                              Life sciences 11086 2008-01-01
## 3   Agricultural sciences and natural resources  1198 2008-01-01
## 4                       Agricultural economics   111 2008-01-01
## 5 Agricultural and horticultural plant breeding    28 2008-01-01
## 6                 Agricultural animal breeding     3 2008-01-01
```

```
labels <- c("Epidemiologyd" = "Epidemiology",
            "Public health" = "Public Health",
            "Biometrics and biostatistics" = "Biometrics and Biostatistics",
            "Health systems administration" = "Health Systems Administration",
            "Statistics (mathematics)" = "Statistics",
            "Computational biology" = "Computational Biology")

epi_plot <- Clean_Data %>% filter(Field == "Epidemiologyd") %>%
  ggplot(aes(x = Year, y = Grads)) +
  geom_point() +
  geom_smooth() +
  labs(x = "Year", y = "Number of Epidemiology Doctorates") +
  theme_minimal() +
  theme(axis.title.y = element_text(margin = margin(t = 0, r = 20, b = 0, l = 0)))

comparison_plot <- Clean_Data %>% filter(Field %in% c("Epidemiologyd",
                                    "Public health",
                                    "Biometrics and biostatistics",
                                    "Health systems administration",
                                    "Statistics (mathematics)",
                                    "Computational biology")) %>%
  ggplot(aes(x = Year, y = Grads, color = Field)) +
  geom_point() +
  geom_smooth() +
```

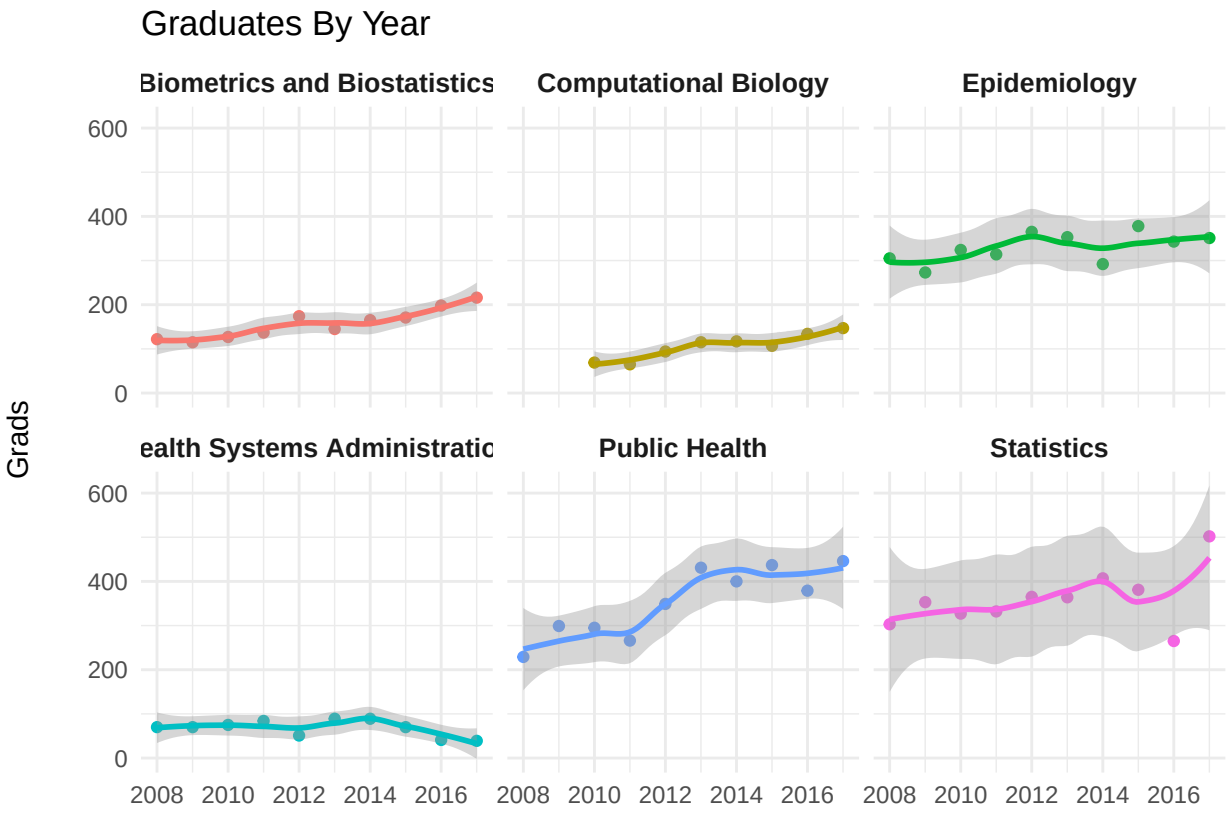
```
  labs(title = "Graduates By Year", x = "") +
  facet_wrap(~ Field, ncol = 3, labeller = labeller(Field = labels)) +
  theme_minimal() +
  theme(legend.position="none", strip.text.x = element_text(size=10, face="bold"), axis.title.y = eleme


epi_plot
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
comparison_plot
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggsave(epi_plot, file = "epi.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
ggsave(comparison_plot, file = "comparison.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```