# FAHA – Assignment 02

## Question 1: Understanding

I drew the graph in the R software. Here is the code used for the creation, followed by the graph itself.

```r
library(tibble)
library(ggplot2)
corrs = c(rep(0, 29), seq(0, 0.62, 0.02), seq(0.64, 0.355, -0.015))
df = tibble(years = 1940:2020, trend = corrs + 0.05, corrs = corrs + runif(81, 0, 0.1))

ggplot(df, aes(x = years, y = corrs)) +
  geom_line(linetype = 1) +
  geom_line(aes(y = trend), color = 'blue', linetype = 2, size = 1.5 ) +
  geom_vline(xintercept = c(1969, 2001), color = "red", linetype = 3) +
  annotate("text", x = 1969, y = 0.75, label = "Since 1969 the correlation increases.",
           color = "darkgreen") +
  annotate("text", x = 2001, y = 0.05, label = "Since 2001 the correlation decreases.",
           color = "darkgreen") +
  annotate("text", x = 1954, y = 0.25, label = "Indiference period\n(1940-1969)") +
  annotate("text", x = 1980, y = 0.55, label = "Differentiation period\n(1970-2000)") +
  annotate("text", x = 2010, y = 0.25, label = "Equitation period\n(2001-2020)") +
  labs(title = "Changes in average correlation between the gender of author and
the presence of selected signalling words in blurbs over the course of time",
    x = "Year of books' publication",
    y = "Correlation between the gender of author
and the presence of signaling words",
caption = "Note: The dashed blue line indicates longitudinal trend in average correlation.
The solid black line indicates average correlation for respective year.") +
  theme_minimal()
```
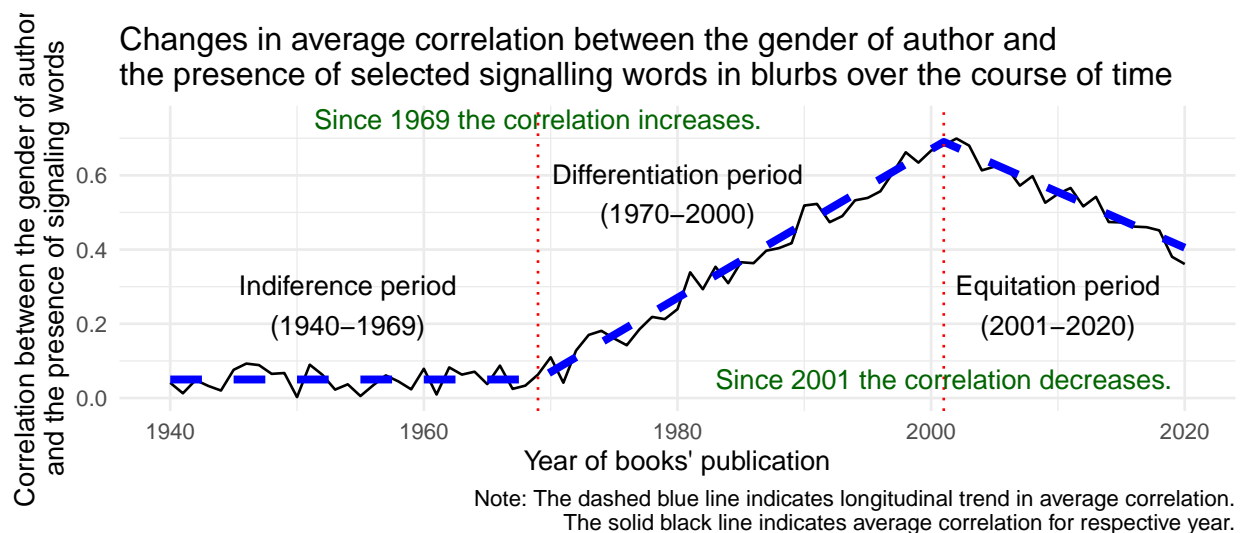
# Question 2: Application

**Question 2.1 – Novelty in the set of blurbs**