

Income Inequality vs Index of Health and Social Problems : ggplot

coop711

2018-05-09

Data

Equality Trust에서 기부금을 받고 제공하는 두 종류의 자료 중 23개 국가의 각종 지표를 비교한 자료에 World Bank에서 발표하는 GDP자료 ([https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(PPP\)_per_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita))를 추가하여 읽어들이면,

```
library(knitr)
library(RColorBrewer)
options(width = 132)
load("Inequality_Index_HS.RData")
dim(data_full)
```

```
## [1] 23 30
```

```
data_23 <- data_full[c("Country", "Income.inequality")]
```

소득불평등 (23개국)

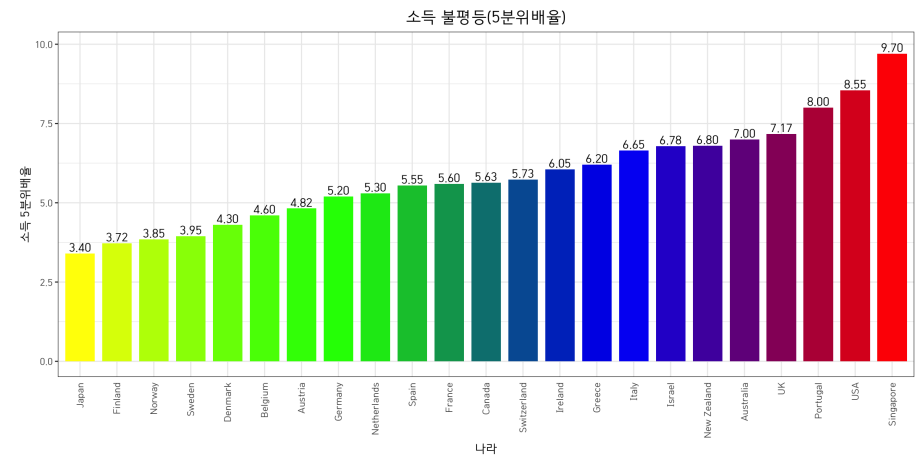
이 자료 중 소득불평등을 나타내는 지표는 5분위배율로서 두번째 컬럼에 `Income.inequality` 라는 이름으로 나와 있고, 건강과 사회문제 지표는 13번째 컬럼에 `Index.of.health...social_problems` 라는 이름으로 주어져 있다. 나라들은 `Country` 라는 변수명으로 첫번째 컬럼에 나와 있다.

Barplot(geom_bar)

```
# par(mai = c(2.0, 0.8, 0.8, 0.4) + 0.02)
library(ggplot2)
source("theme_kr.R")
# colfunc <- colorRampPalette(brewer.pal(9, "YlOrRd"))
# colfunc <- colorRampPalette(brewer.pal(11, "RdYlGn"))
colfunc <- colorRampPalette(c("Yellow", "Green", "Blue", "Red"))
fill_col <- colfunc(23)
o_ineq <- order(data_23$Income.inequality)
data_23$Country_o <- factor(data_23$Country, levels = data_23$Country[o_ineq])
```

```
g1 <- ggplot(data = data_23[c("Income.inequality", "Country_o")])
g2 <- g1 +
  geom_bar(mapping = aes(x = Country_o, y = Income.inequality),
    stat = "identity", width = 0.8, fill = fill_col)
# fill = rev(rainbow(23, start = 0, end = 2/3)))

# g2
g3 <- g2 +
  theme_bw()
g4 <- g3 +
# theme(axis.text.x = element_blank()) +
  theme(axis.ticks.x = element_blank()) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)) +
  theme.kr
# g4
g5 <- g4 +
  geom_text(aes(x = Country_o, y = Income.inequality + 0.2,
    label = format(Income.inequality, digits = 3, nsmall = 2)), family =
    "HCR Dotum LVT")
# g5
g6 <- g5 +
  labs(title = "소득 불평등(5분위배율)", x = "나라", y = "소득 5분위배율") +
  theme(plot.title = element_text(hjust = 0.5, size = 15))
g6
```



```
ggsave("../pics/Fifth_Ratio_ggplot.png", g6, width = 12, height = 6)
```

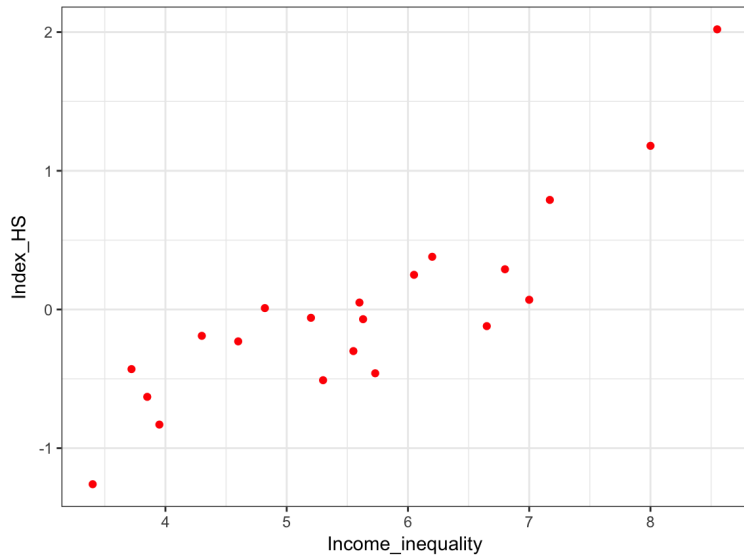
소득불평등과 건강 및 사회문제 (21개국)

Plots

우선 소득불평등과 건강 및 사회문제 지표의 관계를 대략적으로 살펴보면,

```
Country_kr <- c("호주", "오스트리아", "벨기에", "캐나다", "덴마크", "핀란드", "프랑스", "독일", "그리스", "아일랜드", "이탈리아", "일본", "네덜란드", "뉴질랜드", "노르웨이", "포르투갈", "스페인", "스웨덴", "스위스", "영국", "미국")
Index_inequality_df <- data_21[c("Country", "Income_inequality", "Index_HS")]
cor_1 <- cor(data_21["Income_inequality"], data_21["Index_HS"])
```

```
p1 <- ggplot(data = Index_inequality_df,
             mapping = aes(x = Income_inequality, y = Index_HS, label = Country_kr))
p1.1 <- p1 +
  geom_point(col = "red")
# p1.1
p1.2 <- p1.1 +
  geom_text(family = "HCR Dotum LVT")
# p1.2
p2 <- p1.1 +
  theme_bw()
# theme.kr
p2
```

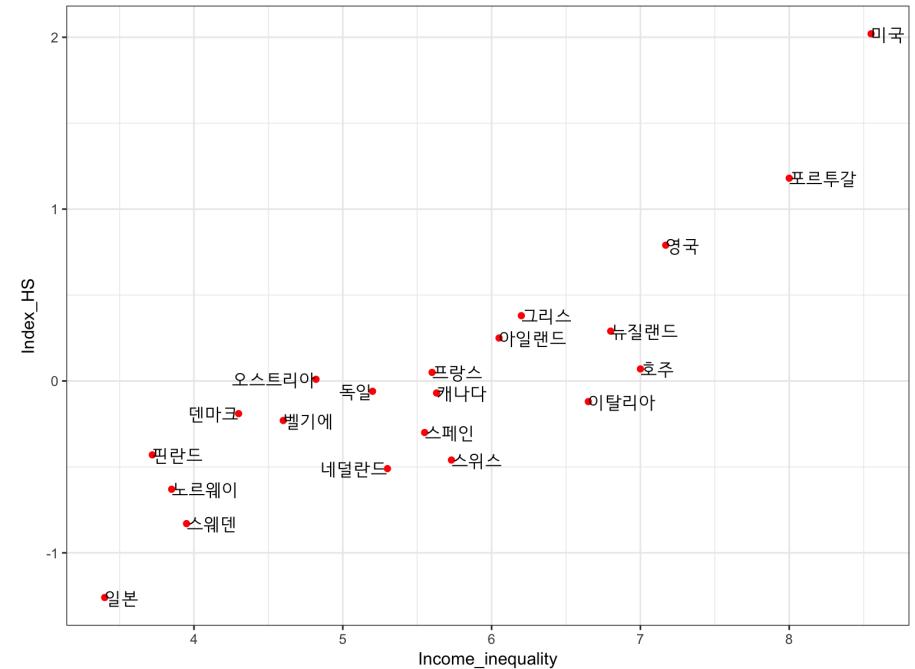


텍스트 위치를 옮겨 보자. 점의 왼쪽으로 옮겨야 할 나라들(ggplot에서는 hjust = "right")을 먼저 찾아보자.

```
Country <- data_21$Country
which(Country %in% c("Austria", "Denmark", "Germany", "Netherlands"))
```

```
## [1] 2 5 8 13
```

```
text_left <- which(Country %in% c("Austria", "Denmark", "Germany", "Netherlands"))
text_right <- setdiff(1:nrow(data_21), text_left)
hjust_text <- ifelse(1:nrow(data_21) %in% text_left, "right", "left")
p4 <- p2 +
  geom_text(hjust = hjust_text, family = "HCR Dotum LVT")
p4
```

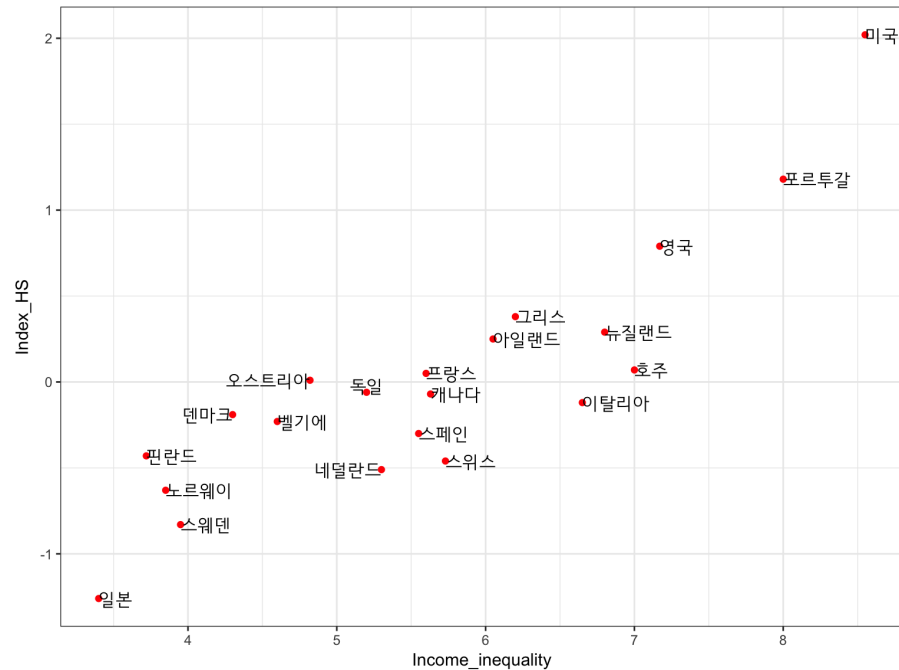


독일의 라벨을 위로 붙이면 보기가 나아질 것으로 생각되므로,

```
which(Country %in% "Germany")
```

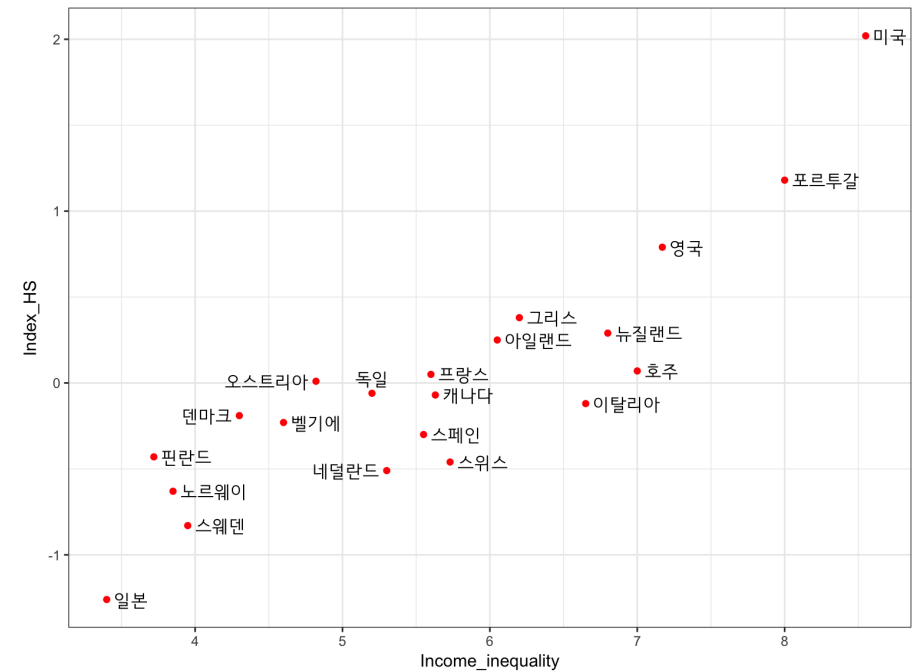
```
## [1] 8
```

```
text_up <- which(Country %in% "Germany")
text_left <- setdiff(1:nrow(data_21), c(text_right, text_up))
vjust_text <- ifelse(1:nrow(data_21) %in% text_up, "bottom", "center")
hjust_text <- ifelse(1:nrow(data_21) %in% text_up, "middle", hjust_text)
p5 <- p2 +
  geom_text(hjust = hjust_text, vjust = vjust_text, family = "HCR Dotum LVT")
p5
```



세부조정을 위해서 nudge_x, nudge_y 설정

```
nudge_y_text <- ifelse(vjust_text == "bottom", 0.05, 0)
nudge_x_text <- ifelse(hjust_text == "middle", 0, ifelse(hjust_text == "right", -0.05
, 0.05))
p6 <- p2 +
  geom_text(hjust = hjust_text, vjust = vjust_text,
    nudge_x = nudge_x_text, nudge_y = nudge_y_text, family = "HCR Dotum LVT")
p6
```

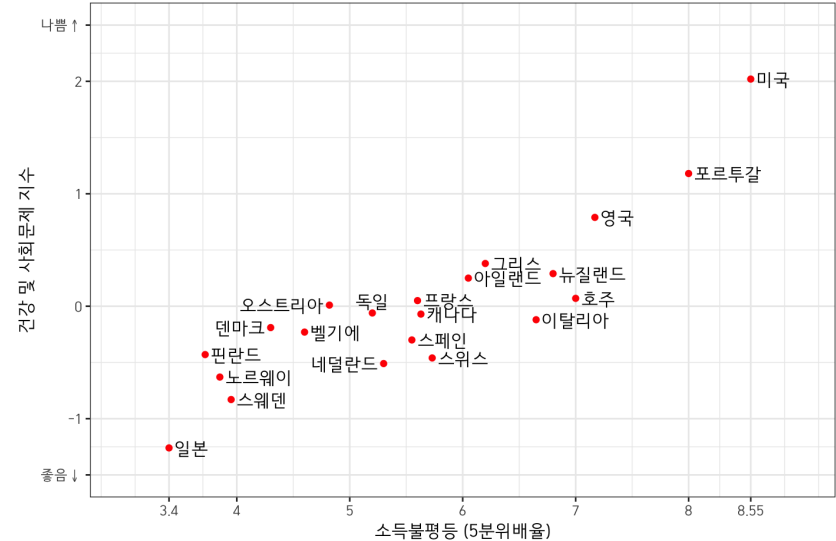


건강 및 사회문제 지표의 경우 어느 방향이 좋은지 알 수 없으므로 친절하게 도표의 주변에(margin)에 알려주고, 이제 조정된 text 외에 x-축과 y-축에 적절한 라벨과 메인 타이틀을 넣어보자.

```
main_title <- "소득불평등과 건강 및 사회문제 지수\n"
x_lab <- "소득불평등 (5분위배율)"
y_lab <- "건강 및 사회문제 지수"
lowest <- data_21$Income_inequality[Country == "Japan"]
highest <- data_21$Income_inequality[Country == "USA"]

p7 <- p6 + theme.kr +
  scale_x_continuous(name = x_lab, breaks = c(lowest, 4:8, highest), labels = c(lowest, 4:8, highest),
    limits = c(3, 9)) +
  scale_y_continuous(name = y_lab, breaks = c(-1.5, -1:2, 2.5),
    labels = c(expression("좋음" %down% ""), -1:2, expression("나쁨" %up% "")),
    limits = c(-1.5, 2.5)) +
  labs(title = main_title, family = "HCR Dotum LVT") +
  theme(plot.title = element_text(hjust = 0.5, vjust = -0.5, size = 15),
    plot.margin = margin(1, 1, 1, 1, "cm"))
p7
```

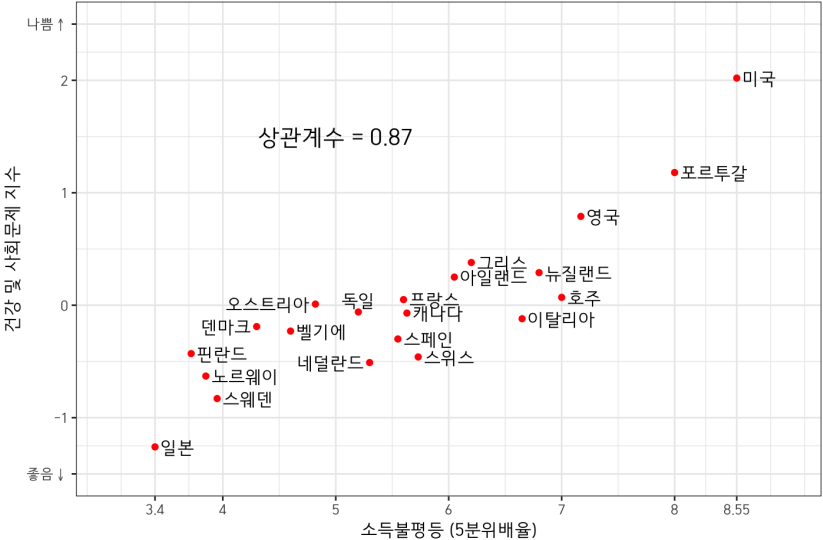
소득불평등과 건강 및 사회문제 지수



상관계수를 텍스트로 그림 안에 넣어주고 여기까지 작업한 내용을 별도의 파일로 저장해 놓으려면,

```
p8 <- p7 +
  annotate("text",
    x = 5,
    y = 1.5,
    label = paste("상관계수 =", round(cor_1, digits = 2)),
    size = 5,
    family = "HCR Dotum LVT")
p8
```

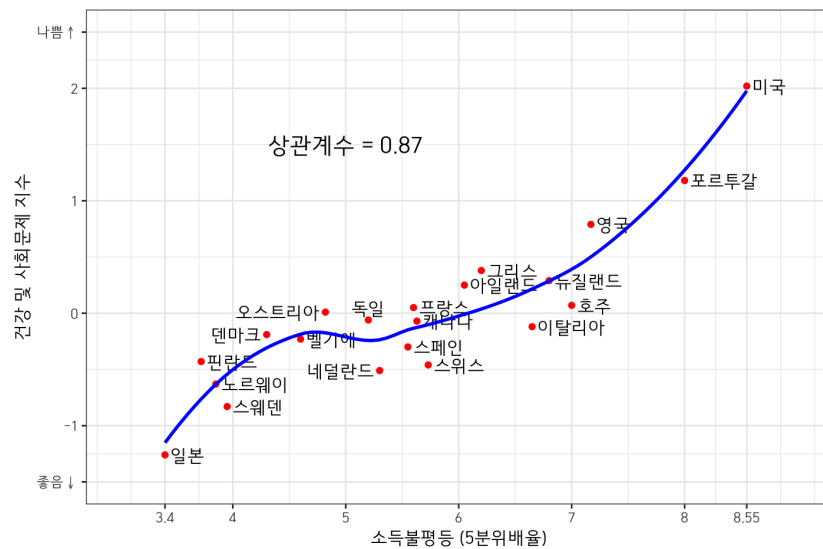
소득불평등과 건강 및 사회문제 지수



선형회귀선을 추가하여 대체적인 추세를 보려면 `lm()` 을 이용하되, `x` , `y` 의 순서를 제대로 바꿔야 함에 유의.

```
lm_ineq <- lm(Index_HS ~ Income_inequality, data = Index_inequality_df)
# lm_ineq <- lm(Index_inequality_df[2:1])
# p9 <- p8 + geom_abline(intercept = lm_ineq$coef[1], slope = lm_ineq$coef[2], colour = "blue")
p9 <- p8 +
  geom_smooth(method = "loess",
              se = FALSE,
              colour = "blue")
p9.2 <- p8 +
  geom_smooth(method = "loess",
              se = TRUE,
              colour = "blue")
p9
```

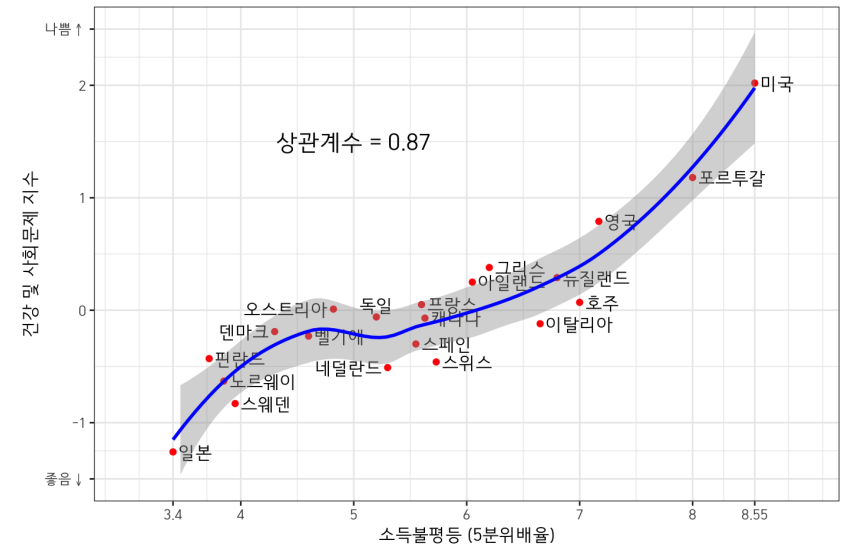
소득불평등과 건강 및 사회문제 지수



se 범위 추가

p9.2

소득불평등과 건강 및 사회문제 지수



ggsave("../pics/Inequality_vs_HS_Index_ggplot.png", p9, dpi = 75)

Saving 7 x 5 in image

ggsave("../pics/Inequality_vs_HS_Index_ggplot_se.png", p9.2, dpi = 75)

Saving 7 x 5 in image

GDP와 건강 및 사회문제 지수

Scatter Diagram

```
Index_GDP_df <- data_21[c("Country", "GDP_WB", "Index_HS")]
cor_2 <- cor(data_21["GDP_WB"], data_21["Index_HS"])
text_left_2 <- which(Country %in% c("Australia", "Belgium", "Canada", "France"))
text_right_2 <- setdiff(1:nrow(data_21), c(text_left_2))
hjust_text_2 <- ifelse(1:nrow(data_21) %in% text_left_2, "right", "left")
nudge_x_text_2 <- ifelse(hjust_text_2 == "right", -250, 250)
```

```
gd1 <- ggplot(data = Index_GDP_df,
              mapping = aes(x = GDP_WB,
                           y = Index_HS,
                           label = Country))

gd1.1 <- gd1 +
  geom_point(colour = "red")
# gd1.1
gd2 <- gd1.1 +
  theme_bw()
gd3 <- gd2 +
  geom_text(hjust = hjust_text_2)
# gd3
gd4 <- gd2 +
  geom_text(hjust = hjust_text_2,
            nudge_x = nudge_x_text_2)
# gd4
```

```
main_title_2 <- "GDP vs Index of Health and Social Problems"
x_lab_2 <- "GDP (Thousand Dollars)"
y_lab_2 <- "Index of Health and Social Problems"
gd5 <- gd4 +
  scale_x_continuous(name = x_lab_2,
                    breaks = seq(20000, 70000, by = 10000),
                    labels = seq(20, 70, by = 10),
                    limits = c(20000, 70000)) +
  scale_y_continuous(name = y_lab_2,
                    breaks = c(-1.5, -1:2, 2.5),
                    labels = c(expression("Better" %down% ""),
                                -1:2,
                                expression("Worse" %up% "")),
                    limits = c(-1.5, 2.5)) +
  labs(title = main_title_2)
# gd5
```

```
gd6 <- gd5 +
  annotate("text",
         x = 40000,
         y = 2,
         label = paste("r =", round(cor_2, digits = 2)),
         size = 5)
# gd6
gd7 <- gd6 +
  geom_smooth(colour = "blue",
             se = FALSE)
gd7
```

```
## `geom_smooth()` using method = 'loess'
```

