ggplot Line Plot Multiple Categorical Variables With Continuous Variable

Fan Wang

2021-09-12

Contents

1 ggplot Line Plot

Go to the **RMD**, **R**, **PDF**, or **HTML** version of this file. Go back to fan's REconTools Package, R Code Examples Repository (bookdown site), or Intro Stats with R Repository (bookdown site).

1.1 Continuous Y and X Variables, Three Categories, One is Subplot

Visualize one continuous variable, along the x-axis, given three categorical variables, with 12 combined categories $3 \times 2 \times 2 = 12$:

- one as subplot (productivity type), 3 unique values
- one as line-color (gamma levels), 2 unique values
- one as line-type (GE vs PE), 2 unique values

The outcome is continuous CEV, generated for results with different productivity types (subplot), generated for PE vs GE (linetype), and at different parameter specifications (lower and higher gamma). X-axis is continuous. The graphs rely on this csv file cev_data.csv.

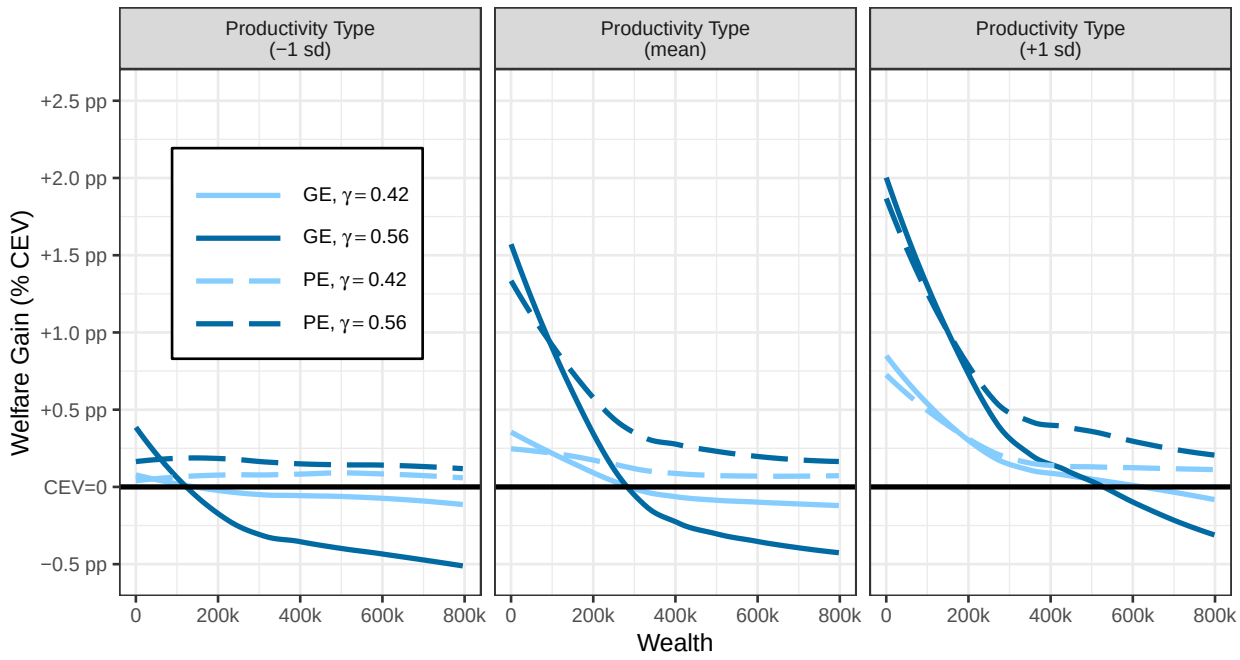
```
# Libraries
# library(tidyverse)

# Load in CSV
bl_save_img <- TRUE
spt_csv_root <- c("G:/repos/R4Econ/tabgraph/ggline/_file/")
spt_img_root <- c("G:/repos/R4Econ/tabgraph/ggline/_file/")
spn_cev_data <- paste0(spt_csv_root, "cev_data.csv")
spn_cev_graph <- paste0(spt_img_root, "cev_graph.png")
spn_cev_graph_eps <- paste0(spt_img_root, "cev_graph.eps")
df_cev_graph <- as_tibble(read.csv(spn_cev_data)) %>% select(-X)

# Dataset subsetting ------

# Line Patterns and Colors ------
# ar_st_age_group_leg_labels <- c("\nGE\n\u03B3=0.42\n", "\nGE\n\u03B3=0.56\n",
```

```
#                                 "\nPE\n\u03B3=0.42\n", "\nPE\n\u03B3=0.42\n")
ar_st_age_group_leg_labels <- c(
  bquote("GE," ~ gamma == .(0.42)),
  bquote("GE," ~ gamma == .(0.56)),
  bquote("PE," ~ gamma == .(0.42)),
  bquote("PE," ~ gamma == .(0.56))
)
ar_st_colours <- c("#85ccff", "#026aa3", "#85ccff", "#026aa3")
ar_st_linetypes <- c("solid", "solid", "longdash", "longdash")

# Labels and Other Strings -------
st_title <- ""
st_x <- "Wealth"
st_y <- "Welfare Gain (% CEV)"
st_subtitle <- paste0(
  "https://fanwangecon.github.io/",
  "R4Econ/tabgraph/ggline/htmlpdfr/fs_ggline_mgrp_ncts.html"
)

# ar_st_age_group_leg_labels <- c("C\u2013Optimal", "V\u2013Optimal")

prod_type_recode <- c(
  "Productivity Type\n(-1 sd)" = "8993",
  "Productivity Type\n(mean)" = "10189",
  "Productivity Type\n(+1 sd)" = "12244"
)

x_labels <- c("0", "200k", "400k", "600k", "800k")
x_breaks <- c(
  0,
  5,
  10,
  15,
  20
)
x_min <- 0
x_max <- 20

# y_labels <- c('-0.01',
#               '\u2191\u2191\nWelfare\nGain\n\nCEV=0\n\nWelfare\nLoss\n\u2193\u2193',
#               '+0.01', '+0.02', '+0.03', '+0.04','+0.05')
y_labels <- c(
  "-0.5 pp",
  "CEV=0",
  "+0.5 pp", "+1.0 pp", "+1.5 pp", "+2.0 pp", "+2.5 pp"
)
y_breaks <- c(-0.01, 0, 0.01, 0.02, 0.03, 0.04, 0.05)
y_min <- -0.011
y_max <- 0.051

# data change -------
df_cev_graph <- df_cev_graph %>%
  filter(across(counter_policy, ~ grepl("70|42", .))) %>%
```

```
  mutate(prod_type_lvl = as.factor(prod_type_lvl)) %>%
  mutate(prod_type_lvl = fct_recode(prod_type_lvl, !!!prod_type_recode))

# graph ------
pl_cev <- df_cev_graph %>%
  group_by(prod_type_st, cash_tt) %>%
  ggplot(aes(
    x = cash_tt, y = cev_lvl,
    colour = counter_policy, linetype = counter_policy, shape = counter_policy
  )) +
  facet_wrap(~prod_type_lvl, nrow = 1) +
  geom_smooth(method = "auto", se = FALSE, fullrange = FALSE, level = 0.95)

# labels
pl_cev <- pl_cev +
  labs(
    x = st_x,
    y = st_y,
    subtitle = st_subtitle
  )

# set shapes and colors
pl_cev <- pl_cev +
  scale_colour_manual(values = ar_st_colours, labels = ar_st_age_group_leg_labels) +
  scale_shape_discrete(labels = ar_st_age_group_leg_labels) +
  scale_linetype_manual(values = ar_st_linetypes, labels = ar_st_age_group_leg_labels) +
  scale_x_continuous(
    labels = x_labels, breaks = x_breaks,
    limits = c(x_min, x_max)
  ) +
  scale_y_continuous(
    labels = y_labels, breaks = y_breaks,
    limits = c(y_min, y_max)
  )

# Horizontal line
pl_cev <- pl_cev +
  geom_hline(yintercept = 0, linetype = "solid", colour = "black", size = 1)
# geom_hline(yintercept=0, linetype='dotted', colour="black", size=2)

# theme
pl_cev <- pl_cev +
  theme_bw() +
  theme(
    text = element_text(size = 10),
    legend.title = element_blank(),
    legend.position = c(0.16, 0.65),
    legend.background = element_rect(
      fill = "white",
      colour = "black",
      linetype = "solid"
    ),
    legend.key.width = unit(1.5, "cm")
```

```
  )

# Print Images to Screen -----
print(pl_cev)
```

https://fanwangecon.github.io/R4Econ/tabgraph/ggline/htmlpdfr/fs_ggline_mgrp_ncts.html




```
# Save Image Outputs -----
if (bl_save_img) {
  png(spn_cev_graph,
    width = 160,
    height = 105, units = "mm",
    res = 150, pointsize = 7
  )
  ggsave(
    spn_cev_graph_eps,
    plot = last_plot(),
    device = "eps",
    path = NULL,
    scale = 1,
    width = 200,
    height = 100,
    units = c("mm"),
    dpi = 150,
    limitsize = TRUE
  )
  print(pl_cev)
  dev.off()
}
```

```
## pdf
##   2
```

1.2 Continuous Y and X Variables, Two Categories, One is Subplot

In contrast to the first line plot, in this second example, we use both varying line color as well as line shape and scatter type to distinguish categories of one categorical variable. Visualize one continuous variable, along the x-axis, given three categorical variables, with 10 combined categories $2 \times 5 = 10$:

- one as subplot (GE vs PE), 2 unique values
- one with line-color, line-color and scatter shape joint variation (counterfactual type), 5 unique values

The outcome is change in male and female labor participation gaps, generated under partial and general equilibrium (subplot), generated for different counterfactual policies (linetype). X-axis is calendar year. Features:

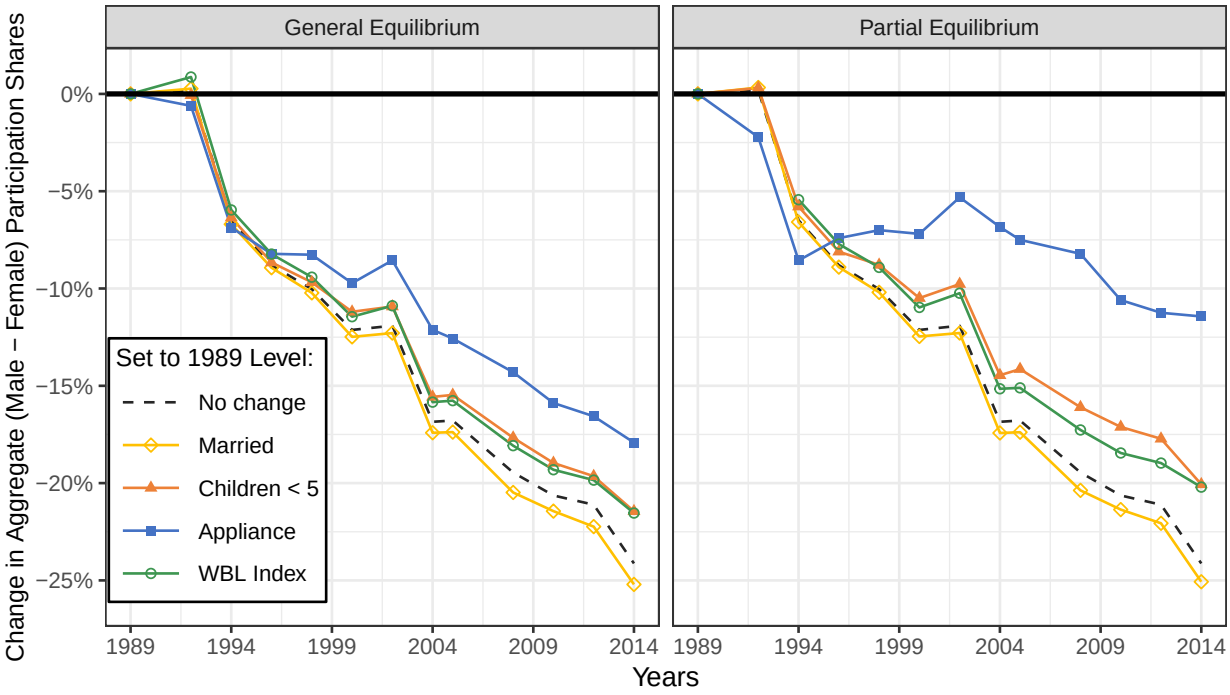
- Calendar year as x-axis
- Line + scatter with varying line patterns and scatter shapes
- Scatter shapes
- Show five lines together, with 2 lines stand out more, and 4 lines overall different than 1
- Legend box area with longer legend text, transparent and no border

For data processing, converts all possible numerical variables to numeric.

```
# Load in CSV
bl_save_img <- TRUE
spt_csv_root <- c("G:/repos/R4Econ/tabgraph/ggline/_file/")
spt_img_root <- spt_csv_root
spn_flfp_sklocc_data <- paste0(spt_csv_root, "flfp_data.csv")
spn_flfp_sklocc_graph <- paste0(spt_img_root, "flfp_sam2fshr_graph.png")
spn_flfp_sklocc_graph_eps <- paste0(spt_img_root, "flfp_sam2fshr_graph.eps")

# Load data
# Convert all convertable numeric columns from string to numeric
# https://stackoverflow.com/a/49054046/8280804
is_all_numeric <- function(x) {
  !any(is.na(suppressWarnings(as.numeric(na.omit(x))))) & is.character(x)
}
df_flfp <- as_tibble(read.csv(spn_flfp_sklocc_data)) %>%
  mutate_if(is_all_numeric, as.numeric) %>%
  filter(year <= 2014)

# Dataset subsetting ------

# Line Patterns and Colors ------
ctr_var_recode <- c(
  "No change" = "1",
  "Married" = "31",
  "Children < 5" = "32",
  "Appliance" = "33",
  "WBL Index" = "34"
)

# https://www.rgbtohex.net/
ar_st_colours <- c("#262626", "#FFC001", "#ED8137", "#4472C4", "#3E9651")
# http://www.sthda.com/english/wiki/ggplot2-line-types-how-to-change-line-types-of-a-graph-in-r-softwar
ar_st_linetypes <- c("dashed", "solid", "solid", "solid", "solid")
# http://sape.inf.usi.ch/quick-reference/ggplot2/shape
# 32 is invisible shape
ar_it_shapes <- c(32, 5, 17, 15, 1)
```

```r
# Labels and Other Strings -------
st_title <- ""
st_x <- "Years"
st_y <- "Change in Aggregate (Male - Female) Participation Shares"
st_subtitle <- paste0(
  "https://fanwangecon.github.io/",
  "R4Econ/tabgraph/ggline/htmlpdfr/fs_ggline_mgrp_ncts.html"
)

# ge_pe_recode <- c(
#   "General Equilibrium\n(Adjust Wages)" = "GE",
#   "Partial Equilibrium\n(Wage as Observed)" = "PE"
# )
ge_pe_recode <- c(
  "General Equilibrium" = "GE",
  "Partial Equilibrium" = "PE"
)
# x.breaks <- c(1989, seq(1992, 2004, by = 2), 2005, seq(2008, 2014, by = 2))
# x.labels <- paste(x.breaks[1:13])
x.breaks <- seq(1989, 2014, by = 5)
x.labels <- paste(x.breaks[1:6])
x.min <- 1989
x.max <- 2014

y.breaks <- round(seq(-0.30, 0.05, by = 0.05), 2)
y.labels <- paste0(paste(y.breaks[1:length(y.breaks)] * 100), "%")

y.min <- -0.26
y.max <- 0.01

# data change -------
df_flfp_sklocc_graph <- df_flfp %>%
  filter(ctr_var_idx %in% c(1, 31, 32, 33, 34) & category == "C001") %>%
  mutate(
    ge_pe = as.factor(ge_pe),
    ctr_var_idx = as.factor(ctr_var_idx)
  ) %>%
  mutate(ge_pe = fct_recode(ge_pe, !!!ge_pe_recode)) %>%
  mutate(ctr_var_idx = fct_recode(ctr_var_idx, !!!ctr_var_recode)) %>%
  select(year, ctr_var_idx, ge_pe, part_yeargender_shr_m2f_dfv1st)

# graph ------
pl_flfp_agg <- df_flfp_sklocc_graph %>%
  ggplot(aes(
    x = year, y = part_yeargender_shr_m2f_dfv1st,
    colour = ctr_var_idx, linetype = ctr_var_idx, shape = ctr_var_idx
  )) +
  facet_wrap(~ge_pe, nrow = 1) +
  geom_line() +
  geom_point()

# labels
pl_flfp_agg <- pl_flfp_agg +
```

```
  labs(
    x = st_x,
    y = st_y
  )
# subtitle = st_subtitle

# set shapes and colors
# scale_colour_manual(values = ar_st_colours, labels = ctr_var_recode) +
# scale_shape_manual(values=ar_it_shapes, labels = ctr_var_recode) +
# scale_linetype_manual(values = ar_st_linetypes, labels = ctr_var_recode) +
pl_flfp_agg <- pl_flfp_agg +
  scale_colour_manual(values = ar_st_colours) +
  scale_shape_manual(values = ar_it_shapes) +
  scale_linetype_manual(values = ar_st_linetypes) +
  scale_x_continuous(
    labels = x.labels, breaks = x.breaks,
    limits = c(x.min, x.max)
  ) +
  scale_y_continuous(
    labels = y.labels, breaks = y.breaks,
    limits = c(y.min, y.max)
  )

# Horizontal line
pl_flfp_agg <- pl_flfp_agg +
  geom_hline(yintercept = 0, linetype = "solid", colour = "black", size = 1)
# geom_hline(yintercept=0, linetype='dotted', colour="black", size=2)

# theme
pl_flfp_agg <- pl_flfp_agg +
  theme_bw() +
  theme(
    text = element_text(size = 11),
    legend.title = element_text(size = 10),
    legend.margin = margin(c(0.1, 0.1, 0.1, 0.1), unit = "cm"),
    legend.position = c(0.10, 0.27),
    legend.background = element_rect(
      fill = "white",
      colour = "black",
      linetype = "solid"
    ),
    legend.key.width = unit(1.0, "cm"),
    axis.title.y = element_text(size = 10)
  ) +
  guides(
    color = guide_legend(title = "Set to 1989 Level:"),
    linetype = guide_legend(title = "Set to 1989 Level:"),
    shape = guide_legend(title = "Set to 1989 Level:")
  )

# Print Images to Screen
print(pl_flfp_agg)
```




```
# Save Image Outputs -----
if (bl_save_img) {
  png(spn_flfp_sklocc_graph,
    width = 200,
    height = 100, units = "mm",
    res = 150, pointsize = 7
  )
  ggsave(
    spn_flfp_sklocc_graph_eps,
    plot = last_plot(),
    device = "eps",
    path = NULL,
    scale = 1,
    width = 200,
    height = 100,
    units = c("mm"),
    dpi = 150,
    limitsize = TRUE
  )
  print(pl_flfp_agg)
  dev.off()
}
```

```
## pdf
##   2
```

1.3 Continuous Y and X Variables, Four Categories, Three for Subplot

In contrast to the line plot above, in this third example, we have three categorical variables that will be visualized via plots and subplots. We have four categoricalv variables overall, for the fourth categorical variable, as in the second example, we continue to use both varying line color as well as line shape and scatter

type to distinguish categories of this fourth categorical variable. Visualize one continuous variable, along the x-axis, given four categorical variables, with 60 combined categories $3 \times 2 \times 2 \times 3 = 36$:

- one as plot, generate three different plots, 3 unique values, achieved by saving a function and running the function three times with variable conditioning.
- one as *facet_grid* row group, 2 unique values.
- one as *facet_grid* column group, 2 unique values.
- one with line-color, line-color and scatter shape joint variation (counterfactual type), 3 unique values

Following the example above, continue to analyze female labor participation. Generated under partial and general equilibrium (subplot), and skill and occupational groups. , generated for different counterfactual policies (linetype). X-axis is calendar year.

Features:

- *facet_grid*: Multiple rows and columns for faceting, row and column labels
- No spacing for empty title line
- Graph as function with simple variable and parameter adjustment
- No minor grid
- Do not show ylabel.

First define the graphing function:

```
# The graphing function with limited parameter options.
ff_grhlfp_gepeedu_byocc <-
  function(bl_save_img = TRUE,
           st_occ = "Manual",
           y_breaks = round(seq(0.08, 0.18, by = 0.02), 2),
           y_min = 0.08,
           y_max = 0.19,
           ar_leg_position = c(0.29, 0.50),
           it_width = 160, it_height = 105,
           st_subtitle = paste0(
             "https://fanwangecon.github.io/",
             "R4Econ/tabgraph/ggline/htmlpdfr/fs_ggline_mgrp_ncts.html"
           )) {

    # Load in CSV
    spt_csv_root <- c("G:/repos/R4Econ/tabgraph/ggline/_file/")
    spt_img_root <- spt_csv_root
    spn_flfp_sklocc_data <- paste0(spt_csv_root, "flfp_data.csv")
    spn_flfp_sklocc_graph <- paste0(
      spt_img_root,
      paste0("flfp_gepe_colhigh_", tolower(st_occ), "_graph.png")
    )
    spn_flfp_sklocc_graph_eps <- paste0(
      spt_img_root,
      paste0("flfp_gepe_colhigh_", tolower(st_occ), "_graph.eps")
    )

    # Load data
    # Convert all convertable numeric columns from string to numeric
    # https://stackoverflow.com/a/49054046/8280804
    is_all_numeric <- function(x) {
      !any(is.na(suppressWarnings(as.numeric(na.omit(x))))) & is.character(x)
    }
    df_flfp <- as_tibble(read.csv(spn_flfp_sklocc_data)) %>%
```

```
  mutate_if(is_all_numeric, as.numeric) %>%
  filter(year <= 2014)

# Dataset subsetting ------

# Line Patterns and Colors ------
ctr_var_recode <- c(
  "Prediction no Counterfactual" = "1",
  "Married at 1989 Levels" = "31",
  "Children < 5 at 1989 Levels" = "32",
  "Appliance at 1989 Levels" = "33",
  "WBL Index at 1989 Levels" = "34"
)

# https://www.rgbtohex.net/
# ar_st_colours <- c("#262626", "#FFC001", "#ED8137", "#4472C4", "#3E9651")
ar_st_colours <- c("#262626", "#ED8137", "#4472C4")
# http://www.sthda.com/english/wiki/ggplot2-line-types-how-to-change-line-types-of-a-graph-in-r-sof
ar_st_linetypes <- c("dashed", "solid", "solid")
# http://sape.inf.usi.ch/quick-reference/ggplot2/shape
# 32 is invisible shape
# ar_it_shapes <- c(32, 5, 17, 15, 1)
ar_it_shapes <- c(32, 17, 15)

# Labels and Other Strings -------
st_x <- "Years"
st_y <- paste0("Female ", st_occ, " Occupation Participation Shares")

# ge_pe_recode <- c(
#   "General Equilibrium\n(Adjust Wages)" = "GE",
#   "Partial Equilibrium\n(Wage as Observed)" = "PE"
# )
ge_pe_recode <- c(
  "General Equilibrium" = "GE",
  "Partial Equilibrium" = "PE"
)
# ge_pe_recode <- c(
#   "GE" = "GE",
#   "PE" = "PE"
# )

skilled_unskilled_recode <- c(
  "College Women" = "skilled",
  "Secondary Women" = "unskilled"
)

# x_breaks <- seq(1989, 2014, by = 5)
x_breaks <- c(1990, 1995, 2000, 2005, 2010)
x_labels <- paste(x_breaks[1:length(x_breaks)])
x_min <- 1989
x_max <- 2014

# y_breaks <- round(seq(0.08, 0.18, by = 0.02), 2)
```

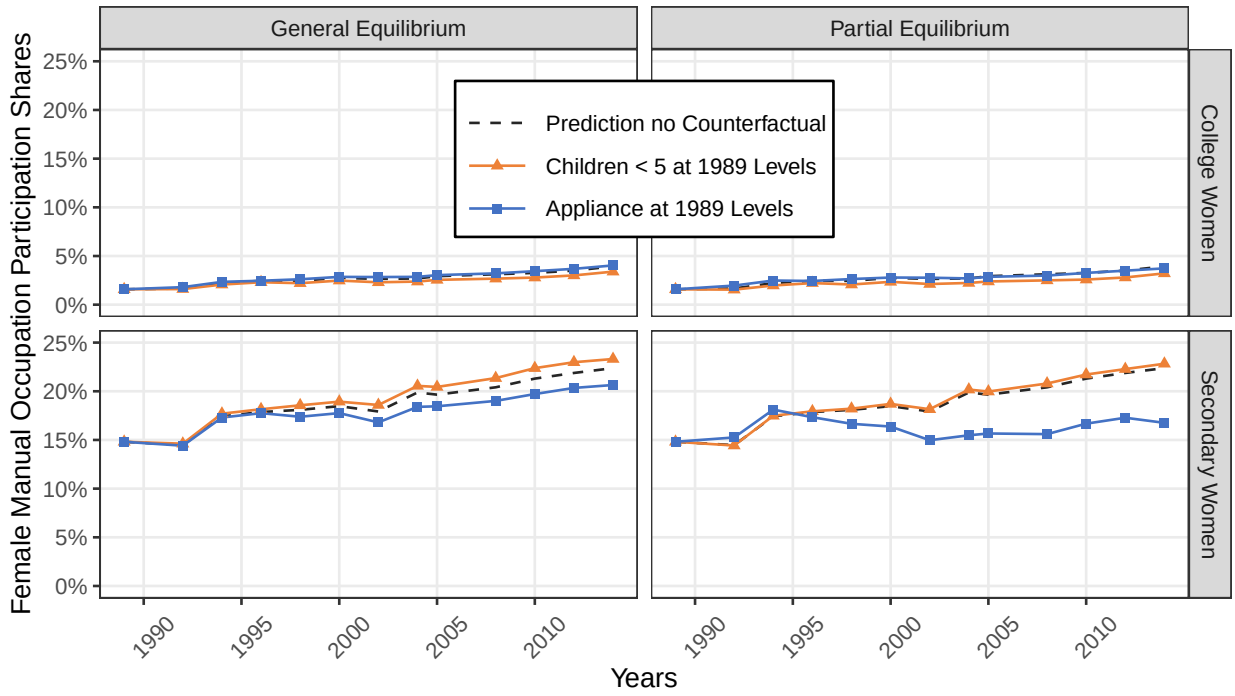
```r
y_labels <- paste0(paste(y_breaks[1:length(y_breaks)] * 100), "%")

# y_min <- 0.08
# y_max <- 0.19

# data change -------
df_flfp_sklocc_graph <- df_flfp %>%
  filter(ctr_var_idx %in% c(1, 32, 33) &
    gender == "Female" &
    occupation %in% c(st_occ)) %>%
  mutate(
    ge_pe = as.factor(ge_pe),
    ctr_var_idx = as.factor(ctr_var_idx)
  ) %>%
  mutate(
    ge_pe = fct_recode(ge_pe, !!!ge_pe_recode),
    skill = fct_recode(skill, !!!skilled_unskilled_recode),
    ctr_var_idx = fct_recode(ctr_var_idx, !!!ctr_var_recode)
  ) %>%
  select(year, skill, occupation, ctr_var_idx, ge_pe, genskl_part_share)

# graph ------
pl_flfp_sklocc <- df_flfp_sklocc_graph %>%
  ggplot(aes(
    x = year, y = genskl_part_share,
    colour = ctr_var_idx, linetype = ctr_var_idx, shape = ctr_var_idx
  )) +
  facet_grid(skill ~ ge_pe) +
  geom_line() +
  geom_point()

# labels
if (st_subtitle == "") {
  pl_flfp_sklocc <- pl_flfp_sklocc +
    labs(
      x = st_x,
      y = st_y
    )
} else {
  pl_flfp_sklocc <- pl_flfp_sklocc +
    labs(
      x = st_x,
      y = st_y,
      subtitle = st_subtitle
    )
}

# set shapes and colors
pl_flfp_sklocc <- pl_flfp_sklocc +
  scale_colour_manual(values = ar_st_colours) +
  scale_shape_manual(values = ar_it_shapes) +
  scale_linetype_manual(values = ar_st_linetypes) +
  scale_x_continuous(
```

```
    labels = x_labels, breaks = x_breaks,
    limits = c(x_min, x_max)
  ) +
  scale_y_continuous(
    labels = y_labels, breaks = y_breaks,
    limits = c(y_min, y_max)
  )

# theme
pl_flfp_sklocc <- pl_flfp_sklocc +
  theme_bw() +
  theme(
    text = element_text(size = 11),
    panel.grid.minor = element_blank(),
    legend.title = element_blank(),
    legend.position = ar_leg_position,
    legend.margin = margin(c(0.1, 0.1, 0.1, 0.1), unit = "cm"),
    legend.background = element_rect(
      fill = "white",
      colour = "black",
      linetype = "solid"
    ),
    legend.key.width = unit(1.0, "cm"),
    axis.text.x = element_text(angle = 45, vjust = 0.1, hjust = 0.1)
    # axis.text.y = element_text(angle = 90, hjust = 0.4)
  )
# element_text(angle = 90, hjust = 0.4)
# axis.title.y = element_blank(), # no y-label

# Save Image Outputs -----
if (bl_save_img) {
  ggsave(
    spn_flfp_sklocc_graph,
    plot = last_plot(),
    device = "png",
    path = NULL,
    scale = 1,
    width = it_width,
    height = it_height,
    units = c("mm"),
    dpi = 150,
    limitsize = TRUE
  )
  ggsave(
    spn_flfp_sklocc_graph_eps,
    plot = last_plot(),
    device = "eps",
    path = NULL,
    scale = 1,
    width = it_width,
    height = it_height,
    units = c("mm"),
    dpi = 150,
```

```
        limitsize = TRUE
      )
      # dev.off()
    }

    return(pl_flfp_sklocc)
  }
```

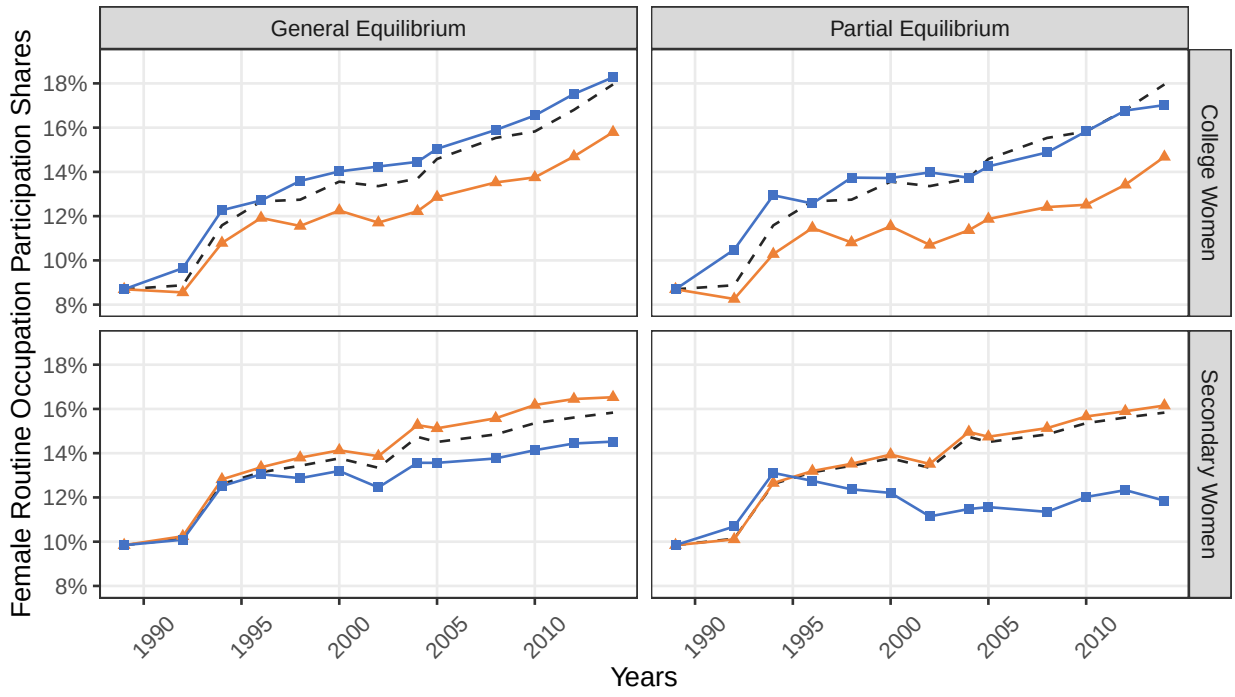
Second, run the function, for Manual, Routine and Analytical Works Separately.

```
it_width <- 100
it_height <- 100
st_subtitle <- paste0(
  "https://fanwangecon.github.io/",
  "R4Econ/tabgraph/ggline/htmlpdfr/fs_ggline_mgrp_ncts.html"
)
st_subtitle <- ""
# Manual,
pl_flfp_sklocc_manual <- ff_grhlfp_gepeedu_byocc(
  bl_save_img = TRUE,
  st_occ = "Manual",
  y_breaks = round(seq(0.00, 0.25, by = 0.05), 2),
  y_min = 0.00, y_max = 0.25,
  ar_leg_position = c(0.50, 0.80),
  it_width = it_width, it_height = it_height,
  st_subtitle = st_subtitle
)
print(pl_flfp_sklocc_manual)
```

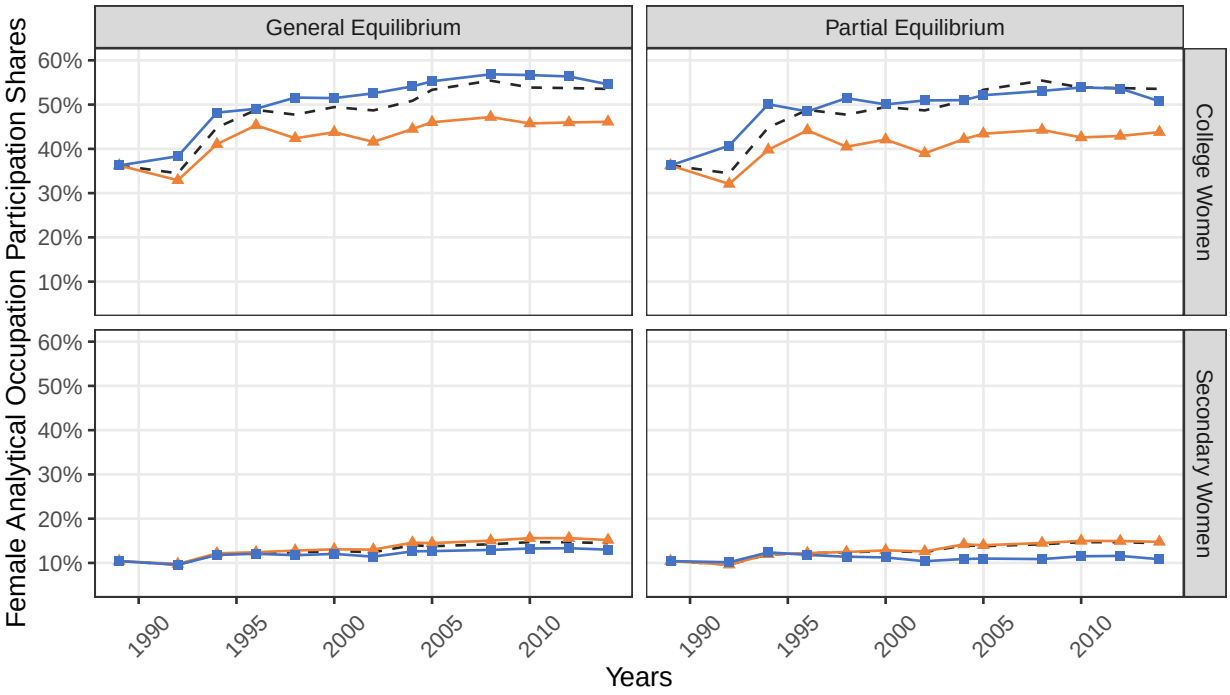



```
# Routine
pl_flfp_sklocc_routine <- ff_grhlfp_gepeedu_byocc(
```

```
  bl_save_img = TRUE,
  st_occ = "Routine",
  y_breaks = round(seq(0.08, 0.18, by = 0.02), 2),
  y_min = 0.08, y_max = 0.19,
  ar_leg_position = "none",
  it_width = it_width, it_height = it_height,
  st_subtitle = st_subtitle
)
print(pl_flfp_sklocc_routine)
```




```
# Analytical
pl_flfp_sklocc_analytical <- ff_grhlfp_gepeedu_byocc(
  bl_save_img = TRUE,
  st_occ = "Analytical",
  y_breaks = round(seq(0.10, 0.60, by = 0.10), 2),
  y_min = 0.05, y_max = 0.60,
  ar_leg_position = "none",
  it_width = it_width, it_height = it_height,
  st_subtitle = st_subtitle
)
print(pl_flfp_sklocc_analytical)
```


General Equilibrium
Partial Equilibrium
College Women
Secondary Women
Female Analytical Occupation Participation Shares
60%
50%
40%
30%
20%
10%
1990
1995
2000
2005
2010
Years