SPEC REU R Resources: Visualizing Regression Results with dot-and-whisker Plots – Homework

Yeiyoung Choo and Claudia Salas Gimenez

March 2025

Welcome to the final assignment of Data Visualization III and the SPEC Data Training Modules! Throughout these modules, you have progressed from learning the basics of R—defining vectors and performing simple arithmetic—to developing advanced skills in data management, visualization, and regression analysis. These skills are designed to prepare you for future research analysis and publication-ready quantitative analysis.

In this homework, you will continue practicing generating dot-and-whisker plots using the `ggcoefstats` function from the `ggstatsplot` package. We'll use these plots to visualize the regression results to answer the research question: Does a land-oriented economic structure increase the likelihood that a country engages in territorial disputes? Additionally, you will extract regression outputs using the broom package to create further customized visualizations beyond dot-and-whisker plots. By the end of this assignment, you will have mastered generating and customizing dot-and-whisker plots, extracting and managing regression outputs in tidy dataframes, and comparing multiple regression models—skills that will help you communicate complex statistical findings effectively.

Save your responses in your personal subfolder in the `412_413` shared Google Drive folder. The R script should be titled `HW_DM3_[YOUR INITIALS]`.

Initial Setup

Begin by setting your working directory to the location of your data files, and load the required libraries and dataset. For this assignment, we will work with the `ProductivePacifists_Data.RDATA` dataset.

For reference, the `ProductivePacifists_Data.RDATA` data was introduced in the research paper *"Productive Pacifists: The Rise of Production-Oriented States and Decline of Profit-Motivated Conquest"* by Markowitz, Mulesky, Graham and Fariss (2020). You can find the research paper and the online appendix here.

```
# Set working directory
#setwd("YourFolderPath")

# Load required libraries
library(tidyverse)
library(broom)
library(ggstatsplot)
## Install ggcoefstats from CRAN if not already installed

# Load the data
load("ProductivePacifists_Data.RDATA")
```

Exploring the Relationship Between Land-oriented Economic Structure and Territorial Disputes

In this assignment, we will analyze whether a country's land-oriented economic structure influences its propensity to initiate or escalate territorial disputes. The research question is: Does a land-oriented economic structure increase the likelihood that a country engages in territorial disputes? The prediction is that countries with a more land-oriented economic structure are more likely to engage in territorial disputes—especially when resource-rich areas are at stake—because such an economic orientation may drive expansionist policies.

Exercise 1: Run Linear Regressions

Exercise 1.1: Estimating the Simple Regression Model

First, run a simple regression model (Model 1) that predicts territorial dispute initiation `tc_dummy_ICOW`, a binary outcome variable, as dependent variable and `land_oriented_medium_continuous`, a continuous measure of economic orientation as independent variable.

Note: Given that the dependent variables are binary, logistic regression is generally preferred over OLS. Logistic regression naturally constrains predicted probabilities to between 0 and 1, capturing the nonlinear relationship between predictors and the likelihood of an event, whereas OLS might produce predictions outside that range and not accurately reflect the underlying probability structure.

```r
# Estimate simple regression model (Model 1)
m1 <- glm(tc_dummy_ICOW ~ land_oriented_medium_continuous,
          family = binomial(link = "logit"), data = dat)

# Display regression results
summary(m1)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_continuous,
##     family = binomial(link = "logit"), data = dat)
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -3.22280    0.08363 -38.534   <2e-16 ***
## land_oriented_medium_continuous -0.05914    0.02982  -1.983   0.0473 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3977.7  on 13570  degrees of freedom
## Residual deviance: 3973.4  on 13569  degrees of freedom
##   (2028 observations deleted due to missingness)
## AIC: 3977.4
##
## Number of Fisher Scoring iterations: 6
```

Exercise 1.2: Estimating the Multivariate Regression Model

Next, expand Model 1 by adding additional control variables. For instance, include variables that capture regime type (`autocracy_BX`) as it could affect a country's propensity to initiate conflicts, and military expenditures (`milex_constant2010us_AFM`) to capture military capacity, influencing the feasibility and likelihood of engaging in militarized disputes. Save this model as Model 2.

Note: Use the logged version of military expenditures to reduce skewness, normalize distribution, and facilitate clearer comparisons across countries.

```
# Create logged military expenditures variable to normalize distribution
dat$milex_logged <- log(dat$milex_constant2010us_AFM + 1)

# Estimate multiple regression model (Model 2)
m2 <- glm(tc_dummy_ICOW ~ land_oriented_medium_continuous + autocracy_BX +
          milex_logged,
          family = binomial(link = "logit"), data = dat)

# Display regression results
summary(m2)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_continuous +
##     autocracy_BX + milex_logged, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -17.80333    1.67682 -10.617  < 2e-16 ***
## land_oriented_medium_continuous   0.03638    0.03252   1.119    0.263
## autocracy_BX                      0.50826    0.12608   4.031 5.55e-05 ***
## milex_logged                      4.60612    0.53350   8.634  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3285.7  on 10466  degrees of freedom
## Residual deviance: 3199.3  on 10463  degrees of freedom
##   (5132 observations deleted due to missingness)
## AIC: 3207.3
##
## Number of Fisher Scoring iterations: 6
```

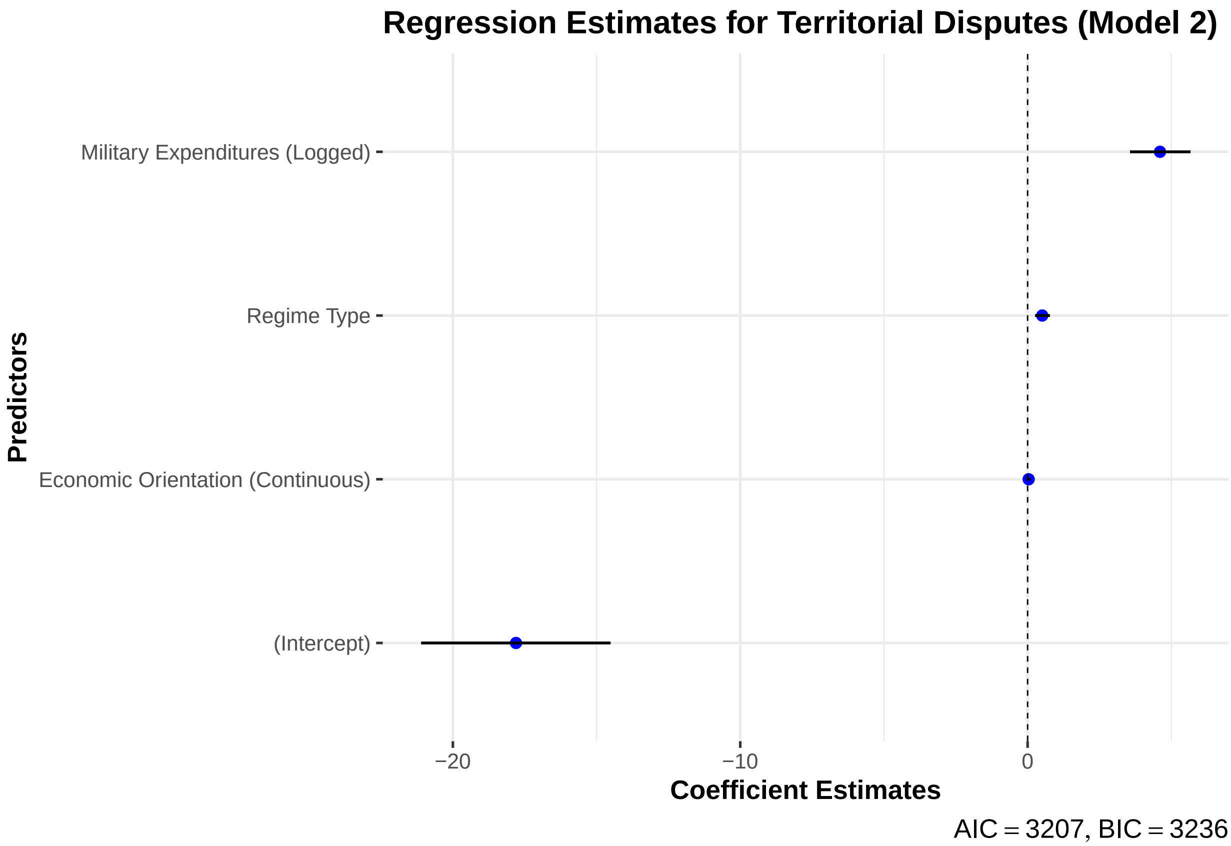
Exercise 2: Generate Dot-and-Whisker Plots with ggstatsplot

Exercise 2.1: Dot-and-Whisker Plot for a Single Model

Now, let's create a dot-and-whisker plot for Model 2 using the `ggcoefstats` function. Challenge yourself to enhance the clarity of your plot by upgrading it beyond the basics.

```
# Create dot-and-whisker plot using ggcoefstats
ggcoefstats(x = m2,
            stats.labels = FALSE,
            ## Remove detailed text labels (coefficients, t-statistics, p-values)
            point.args = list(size = 1.5, color = "blue", na.rm = TRUE),
            vline.args = list(linewidth = 0.25, linetype = "dashed"),
            xlab = "Coefficient Estimates",
```

```r
          ylab = "Predictors",
          title = "Regression Estimates for Territorial Disputes (Model 2)") +
  scale_y_discrete(labels = c(
    "land_oriented_medium_continuous" = "Economic Orientation (Continuous)",
    "autocracy_BX" = "Regime Type",
    "milex_logged" = "Military Expenditures (Logged)",
    "land_CONT" = "Land Borders",
    "cinc_MC" = "Military Capabilities"))
```

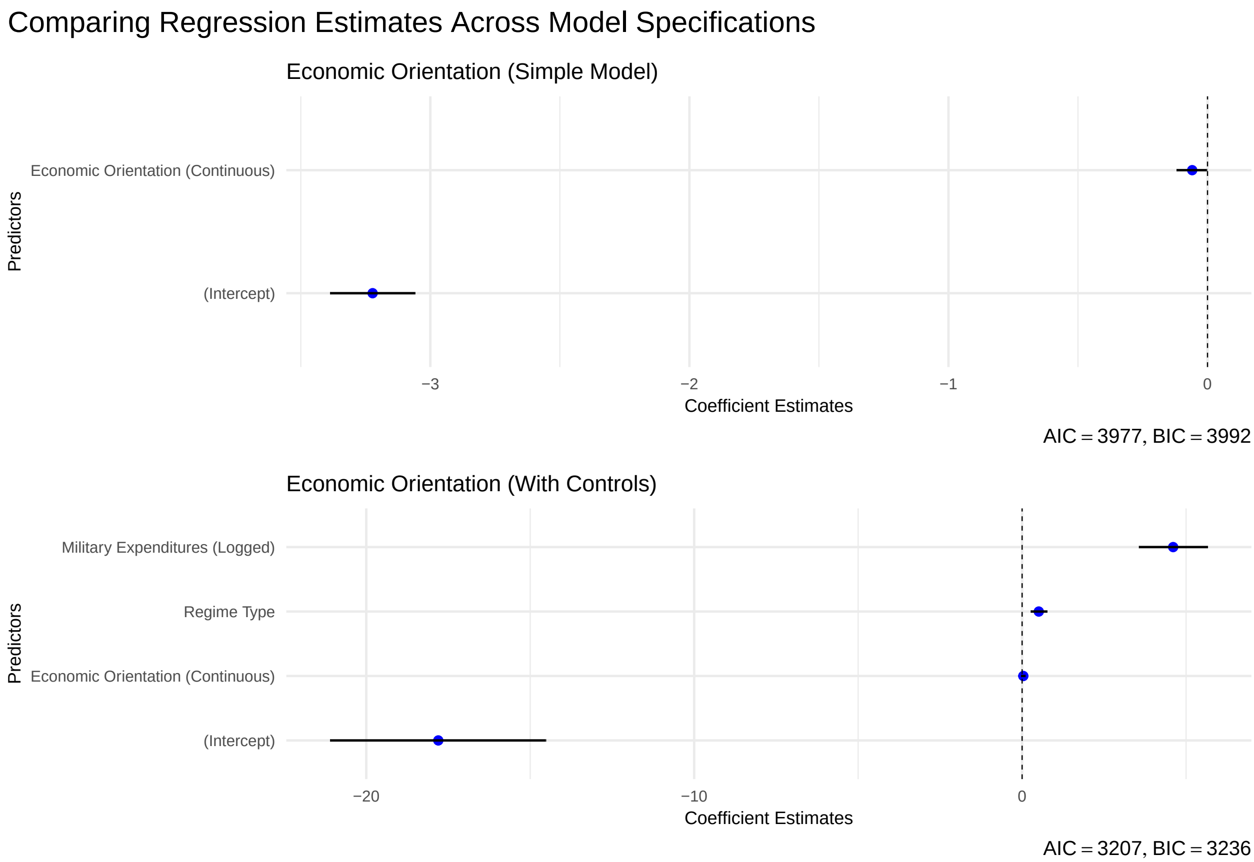


Exercise 2.3: Combined Dot-and-Whisker Plot for Multiple Models

Combine the dot-and-whisker plots for Models 1 and 2 into a single figure to compare how the inclusion of different control variables affects the estimated impact of economic orientation on territorial disputes.

```r
combine_plots(
  plotlist = list(
    # Plot for Model 1: Simple Model
    ggcoefstats(x = m1,
                stats.labels = FALSE,
                point.args = list(size = 1.5, color = "blue", na.rm = TRUE),
                vline.args = list(linewidth = 0.25, linetype = "dashed"),
                xlab = "Coefficient Estimates",
                ylab = "Predictors",
                title = "Economic Orientation (Simple Model)") +
      scale_y_discrete(labels = c(
        "land_oriented_medium_continuous" = "Economic Orientation (Continuous)"
```

```r
    )) +
    theme_minimal() +
    theme(plot.title = element_text(size = 10),
          axis.title = element_text(size = 8),
          axis.text = element_text(size = 7)),
  # Plot for Model 2: With Control Variables
  ggcoefstats(x = m2,
              stats.labels = FALSE,
              point.args = list(size = 1.5, color = "blue", na.rm = TRUE),
              vline.args = list(linewidth = 0.25, linetype = "dashed"),
              xlab = "Coefficient Estimates",
              ylab = "Predictors",
              title = "Economic Orientation (With Controls)") +
    scale_y_discrete(labels = c(
      "land_oriented_medium_continuous" = "Economic Orientation (Continuous)",
      "autocracy_BX" = "Regime Type",
      "milex_logged" = "Military Expenditures (Logged)",
      "land_CONT" = "Land Borders",
      "cinc_MC" = "Military Capabilities"
    )) +
    theme_minimal() +
    theme(plot.title = element_text(size = 10),
          axis.title = element_text(size = 8),
          axis.text = element_text(size = 7))),
plotgrid.args = list(nrow = 2),
annotation.args =
  list(title = "Comparing Regression Estimates Across Model Specifications"))
```



Exercise 3: Extracting and Customizing Tidy Regression Results

While `ggcoefstats` provides clear visualizations, you might wish to further customize your output or create alternative visualizations. Extract the regression results into tidy dataframes using the `broom` package. Use the `tidy()` function to obtain dataframes with regression coefficients, standard errors, and p-values, filtering out the intercept.

```r
# Extract tidy regression results for each model, excluding the intercept
m1df <- tidy(m1) %>% filter(term != "(Intercept)")
m2df <- tidy(m2) %>% filter(term != "(Intercept)")
```

Bonus Exercise

As a bonus question, extend Model 2 by adding additional control variable(s) from the dataset to create a new regression model (Model 3). Other relevant factors to control for could include the number of neighboring countries sharing a land border (`land_CONT`) and the overall measure of military capabilities (`cinc_MC`) to ensure that the associations we observe between economic orientation and territorial disputes are not confounded by regime characteristics or military capabilities. Then, create a dot-and-whisker plot to compare Models 2 and 3, and provide a brief discussion on how the inclusion of the new control variable(s) affects the coefficient estimates.

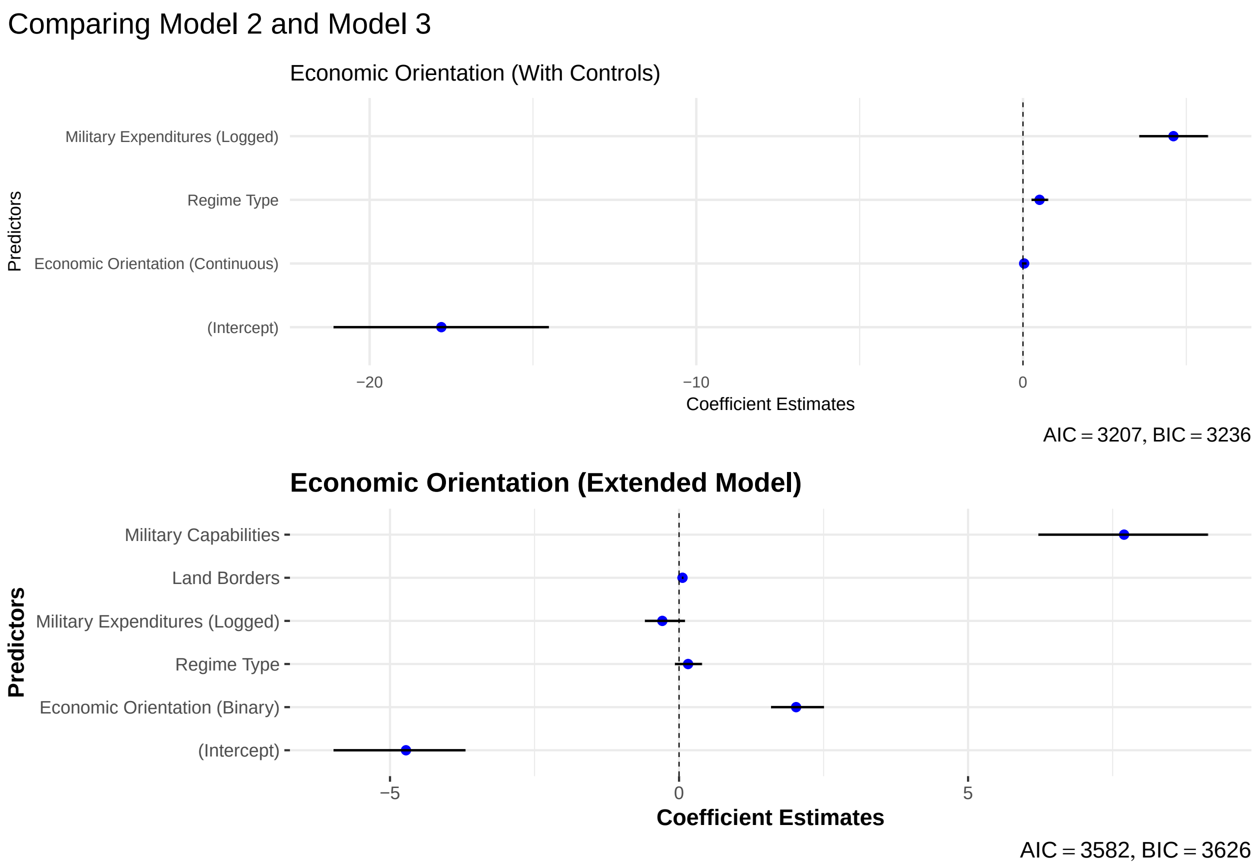
```r
# Estimate expanded regression model (Model 3)
m3 <- glm(tc_dummy_ICOW ~ land_oriented_medium_binary + autocracy_BX +
          milex_logged + land_CONT + cinc_MC,
          family = binomial(link = "logit"), data = dat)
```

```r
# Display regression results
summary(m3)
```

```
##
## Call:
## glm(formula = tc_dummy_ICOW ~ land_oriented_medium_binary + autocracy_BX +
##     milex_logged + land_CONT + cinc_MC, family = binomial(link = "logit"),
##     data = dat)
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -4.723818   0.571017  -8.273   <2e-16 ***
## land_oriented_medium_binary  2.025144   0.230502   8.786   <2e-16 ***
## autocracy_BX                 0.156065   0.117754   1.325   0.1851
## milex_logged                -0.289341   0.171593  -1.686   0.0918 .
## land_CONT                    0.059805   0.005191  11.521   <2e-16 ***
## cinc_MC                      7.698385   0.745846  10.322   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3989.4  on 11578  degrees of freedom
## Residual deviance: 3570.0  on 11573  degrees of freedom
##   (4020 observations deleted due to missingness)
## AIC: 3582
##
## Number of Fisher Scoring iterations: 7
```

```r
# Combine dot-and-whisker plots for Models 2 and 3
combine_plots(
  plotlist = list(
    # Plot for Model 2: With Control Variables
    ggcoefstats(x = m2,
                stats.labels = FALSE,
                point.args = list(size = 1.5, color = "blue", na.rm = TRUE),
                vline.args = list(linewidth = 0.25, linetype = "dashed"),
                xlab = "Coefficient Estimates",
                ylab = "Predictors",
                title = "Economic Orientation (With Controls)") +
      scale_y_discrete(labels = c(
        "land_oriented_medium_continuous" = "Economic Orientation (Continuous)",
        "autocracy_BX" = "Regime Type",
        "milex_logged" = "Military Expenditures (Logged)",
        "land_CONT" = "Land Borders",
        "cinc_MC" = "Military Capabilities"
      )) +
      theme_minimal() +
      theme(plot.title = element_text(size = 10),
            axis.title = element_text(size = 8),
            axis.text = element_text(size = 7)),
    # Plot for Model 3: Extended Model
    ggcoefstats(x = m3,
                stats.labels = FALSE,
```

```
                 point.args = list(size = 1.5, color = "blue", na.rm = TRUE),
                 vline.args = list(linewidth = 0.25, linetype = "dashed"),
                 xlab = "Coefficient Estimates",
                 ylab = "Predictors",
                 title = "Economic Orientation (Extended Model)") +
      scale_y_discrete(labels = c(
        "land_oriented_medium_binary" = "Economic Orientation (Binary)",
        "autocracy_BX" = "Regime Type",
        "milex_logged" = "Military Expenditures (Logged)",
        "land_CONT" = "Land Borders",
        "cinc_MC" = "Military Capabilities"))),
  plotgrid.args = list(nrow = 2),
  annotation.args = list(title = "Comparing Model 2 and Model 3"))
```



Conclusion

In this homework, we examined whether a country's land-oriented economic structure affects its propensity to engage in territorial disputes by running multiple regression analyses and visualizing the results with dot-and-whisker plots using the `ggcoefstats` function from the `ggstatsplot` package.

This assignment, along with the SPEC Data Training Modules, has equipped you with advanced skills in generating and customizing plots, managing datasets, and running and interpreting regression analyses. These techniques will serve as valuable tools for conducting and communicating robust quantitative research in your future projects.