

Visualizing regression data

EPSE 581 FINAL PRESENTATION

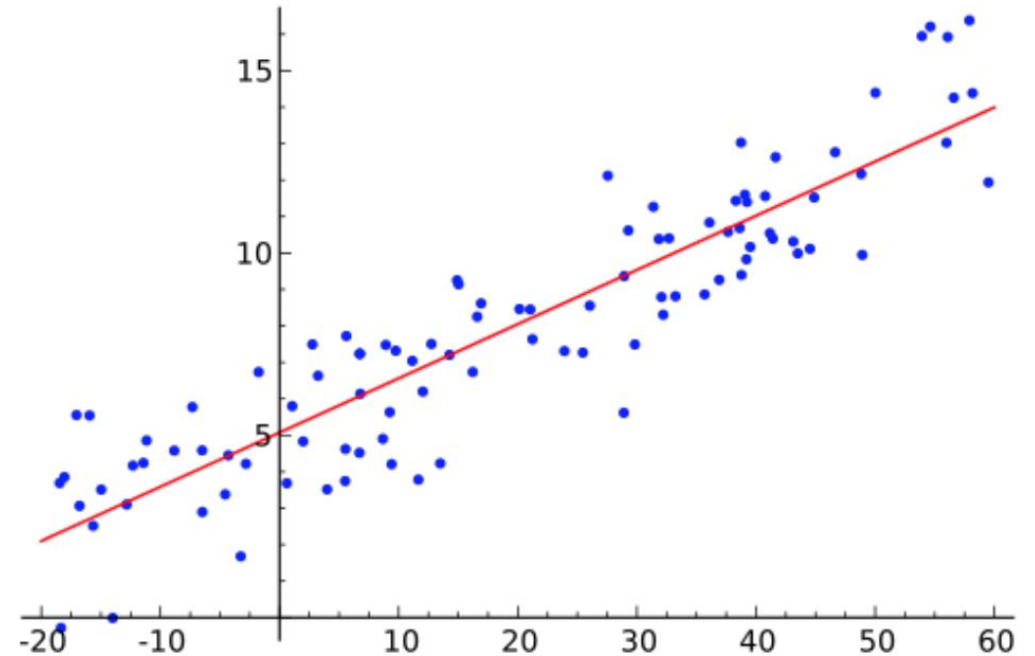
Qinxin Lin

Xuyan Tang

Mindy Chiang

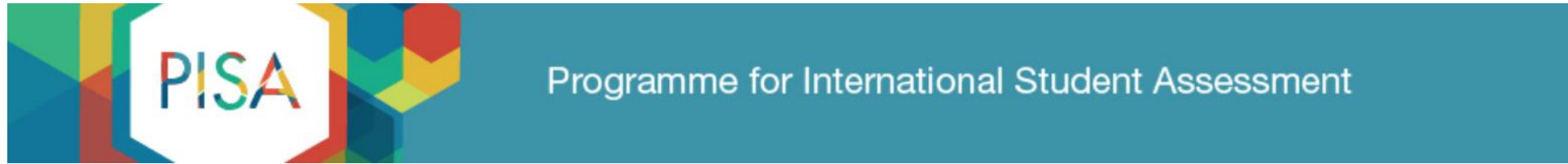
Purpose of our presentation

- To demonstrate effective ways to present **regression** data
- Using tools we learned in class: **Shiny R** and **Tableau**



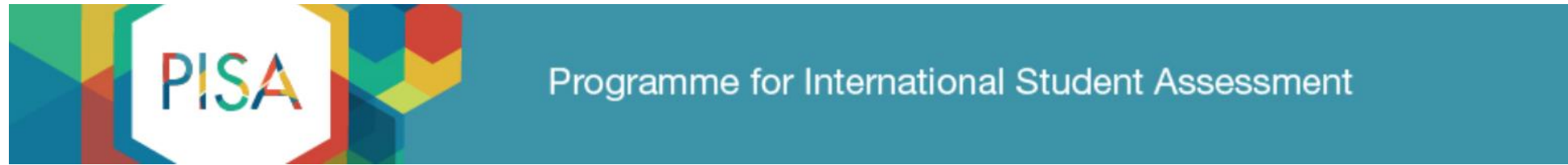
<https://stats.stackexchange.com/questions/89747/how-to-describe-or-visualize-a-multiple-linear-regression-model>

We used OECD PISA data



- PISA is OECD's **Programme for International Student Assessment**
- Every **3 years** it tests **15-year-old** students worldwide in reading, mathematics and science. In 2015, ~540,000 students completed the assessments in 72 countries
- PISA offers insights for **education policy** and practice & helps **monitor trends** in students' performance across countries and in demographic subgroups within a country

We analyzed what factors at home predicted Canadian student performance



- **2015** PISA data for **Canada** analyzed
- **Research question:** What **factors at home** predict Canadian 15-year-old students' **performance in math, science and reading?**
- We tested **linear regression** models where:
 - DV:** Student performance in math, science or reading
 - IVs:** Parental occupation, Parental education, Home education resources, Gender and Region

Using Shiny R to present regression results



Multiple Regression Analysis

Name

Select the input for the dependent variable

Dependent Variables

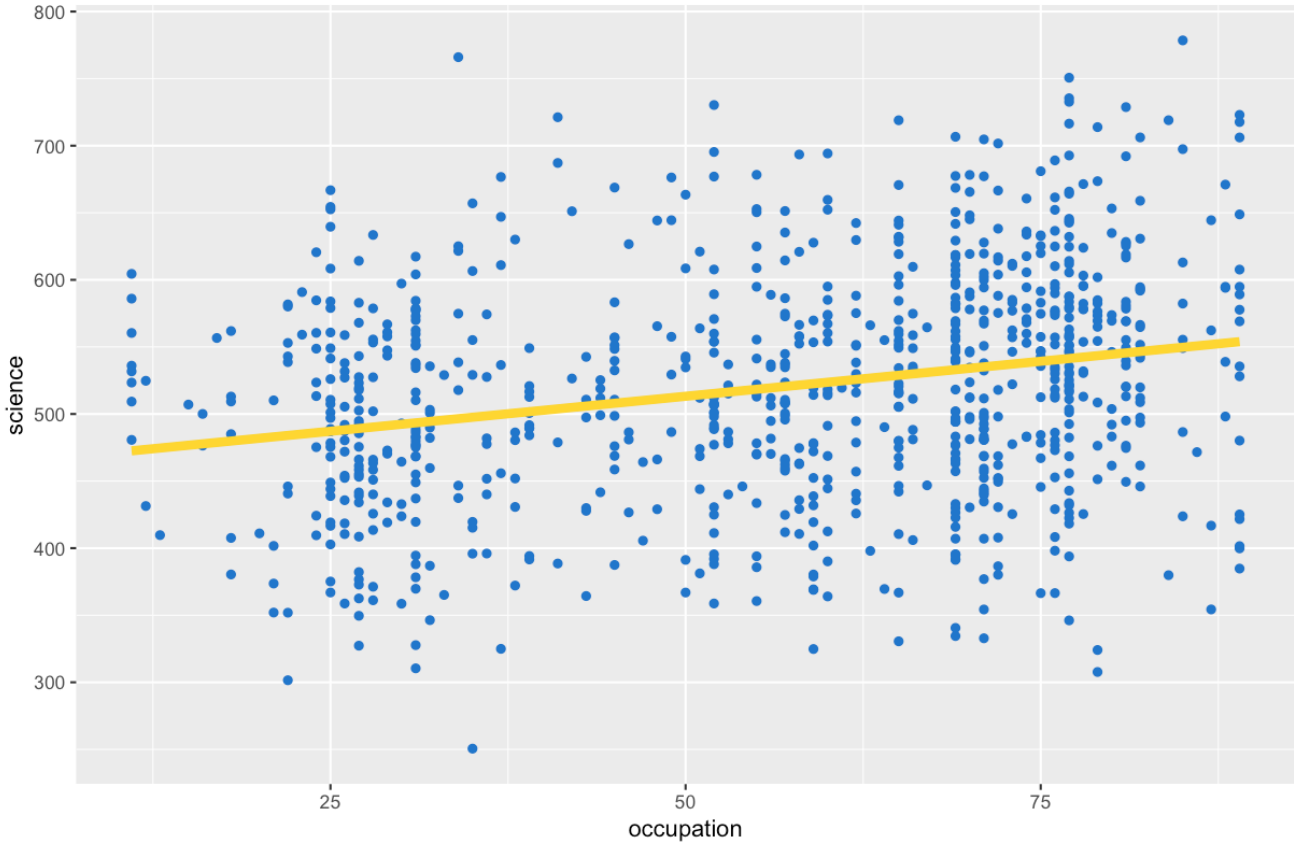
Select the inputs for the independent variables

Independent Variables

- Parent Occupation Status
- Parent Education
- Education Resources

- Data
 - Summary Statistics
 - Histograms
 - Scatter Plot**
 - Correlations
-
- LINE Assumptions (1)
 - LINE Assumptions (2)
 - Multicollinearity
 - Model

scatter plots



How to adapt the code to use your dataset

```
sidebarLayout(  
  sidebarPanel(  
    textInput("name", label = h5("Name"), value = "Name"),  
    HTML('</br>'),  
    selectInput(inputId = "dv", label = "Dependent Variables", choices = list("Math Score"="math", "Reading Score"="reading", "Science Score"="science"),  
               selected=c("math")),  
    HTML('</br>'),  
    p("Independent Variables"),  
    checkboxGroupInput(inputId = "iv", label = "Continuous Variables",  
                      choices = list("Parent Occupation Status"="occupation", "Parent Education"="education", "Education Resources"="resources"),  
                      selected=c("occupation")),  
    HTML('</br>'),  
    selectInput(inputId = "categ", label = "Categorical Variables",  
               choices = list("Gender" = "gender", "Region" = "region", "None" = "none"), selected = "gender")),  
  )  
)
```

How to adapt the code to use your dataset

```
server <- function(input, output) {  
  # read dataset  
  dataset<-read.csv(file="581 project 2.csv", header=TRUE, sep=",")  
  
  iv <- reactive({  
    if(is.null(input$iv)){  
      return(NULL)  
    }  
    if("occupation" %in% input$iv){  
      if("education" %in% input$iv){  
        if("resources" %in% input$iv){  
          c("occupation", "education", "resources")  
        }else{  
          c("occupation", "education")  
        }  
      }else{  
        if("resources" %in% input$iv){  
          c("occupation", "resources")  
        }else{  
          c("occupation")  
        }  
      }  
    }else{  
      if("education" %in% input$iv){  
        if("resources" %in% input$iv){  
          c("education", "resources")  
        }else{  
          c("education")  
        }  
      }else{  
        c("resources")  
      }  
    }  
  })  
}
```


How to adapt the code to use your dataset

```
output$distPlot_iv <- renderPlot({
  par(mfrow=c(1,3))
  if(length(iv())==1){
    hist(dataset[,iv()[1]],col='skyblue',border = 'white', main="Independent Variable", xlab=iv()[1], lwd=2)
  }else if(length(iv())==2){
    hist(dataset[,iv()[1]],col='skyblue',border = 'white', main="Independent Variable", xlab=iv()[1], lwd=2)
    hist(dataset[,iv()[2]],col='skyblue',border = 'white', main="Independent Variable", xlab=iv()[2], lwd=2)
  }else{
    hist(dataset[,iv()[1]],col='skyblue',border = 'white', main="Independent Variable", xlab=iv()[1], lwd=2)
    hist(dataset[,iv()[2]],col='skyblue',border = 'white', main="Independent Variable", xlab=iv()[2], lwd=2)
    hist(dataset[,iv()[3]],col='skyblue',border = 'white', main="Independent Variable", xlab=iv()[3], lwd=2)
  }
})
```

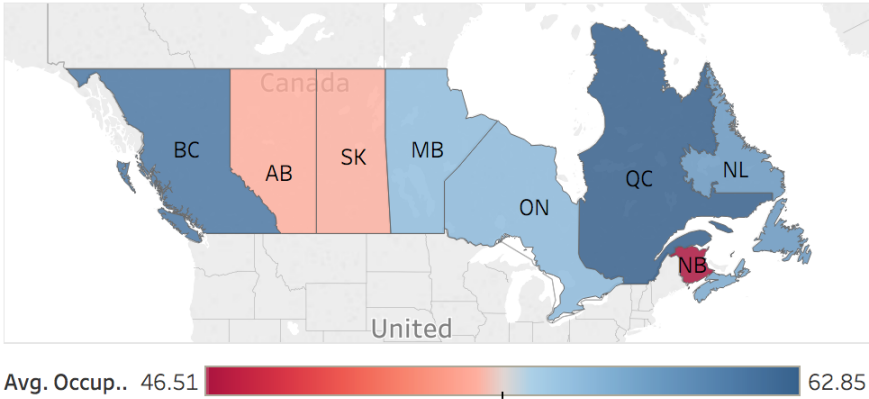
Using Tableau to present
lay-friendly information



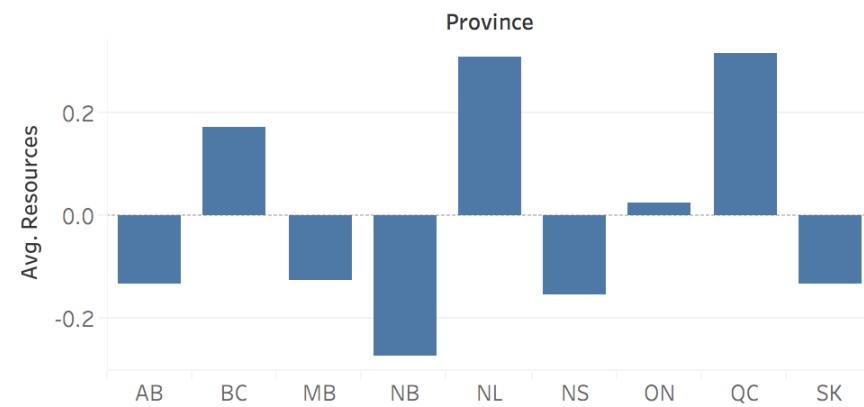


2015 PISA Data for Canada:
BC students perform best, Newbrunswick (NB) students worst.
NB with lowest parent occupation status and home resources.

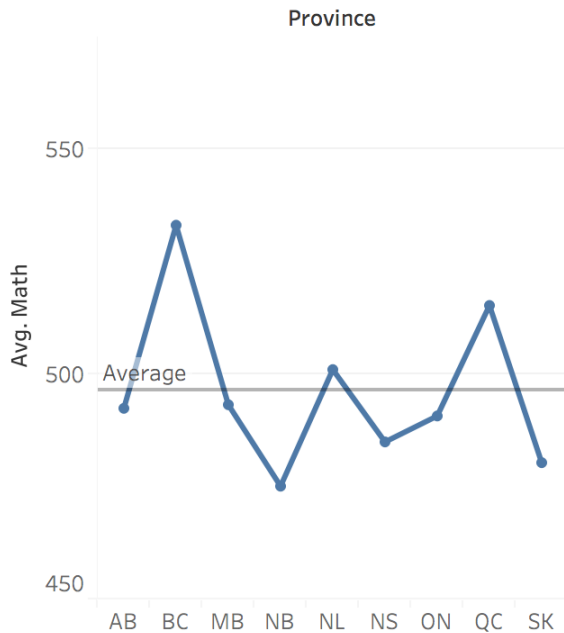
Parent Occupation Status



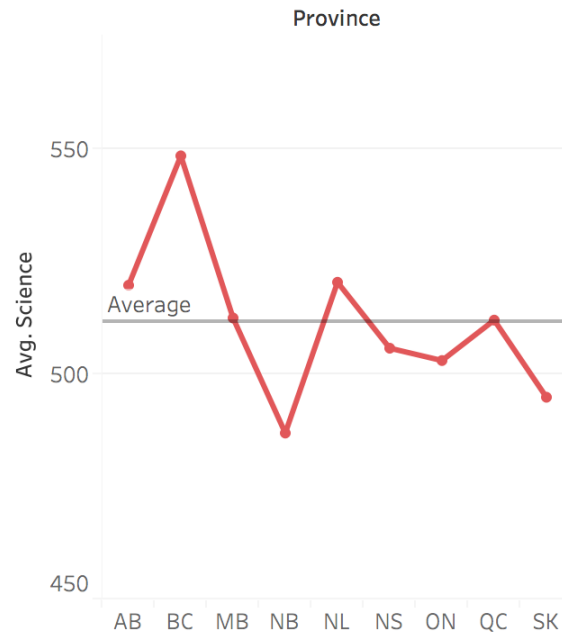
Home Educational Resources



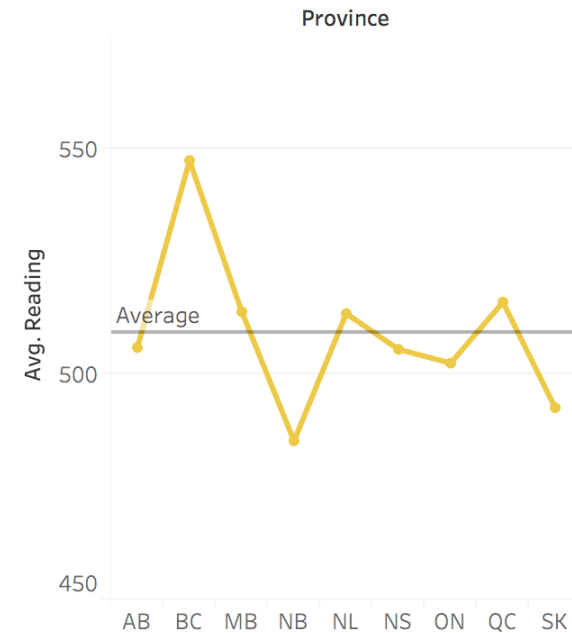
Math Performance



Science Performance



Reading Performance



Improvements & Conclusion

Improvements

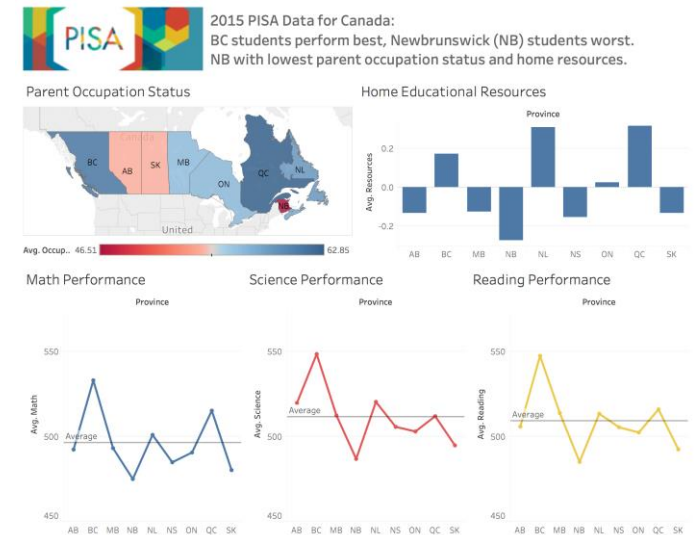
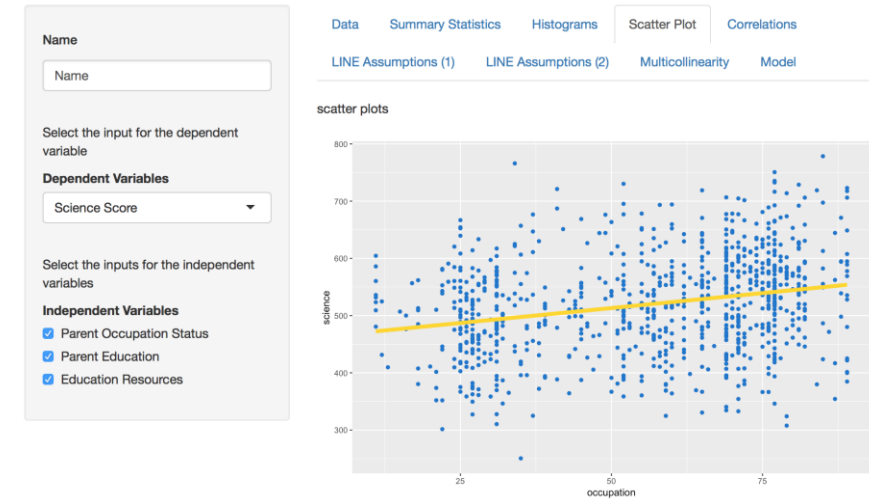
Shiny R

- Random sampling
- Adding interaction terms in the model
- Building the tool to use for stepwise model selection

Tableau

Depending on the audience, can make it into a more attractive infographic

Multiple Regression Analysis



Conclusion

- Tool we developed in Shiny R provides a summary of the important information from regression analysis
- Can be useful for researchers doing regression, especially those less familiar with it
- Tableau is a quick and easy way to explore data and develop lay-friendly visualizations

