Week 7 Presentation

Tingwen Hua, Bowen Yang,
Song(Alice) Zhang, Yilin Sun

Overview - The Analysis of Rural Business Performance

Objectives:

1. Exploratory analysis of the Rural Canada Business Profiles (RCBP) database.

2. Interactive dashboard**s** (**RCBP and Overview**).

3. Cross-analysis of RCBP with other public Statistics Canada data (open ended)

Research questions:

1. How does the Canadian rural business perform according to different dimensions/variables?

2. How can the RCBP data be connected with other StatCan data to produce valuable analysis?

Client Meeting

Wednesday June 8th 9:00am-10:00am

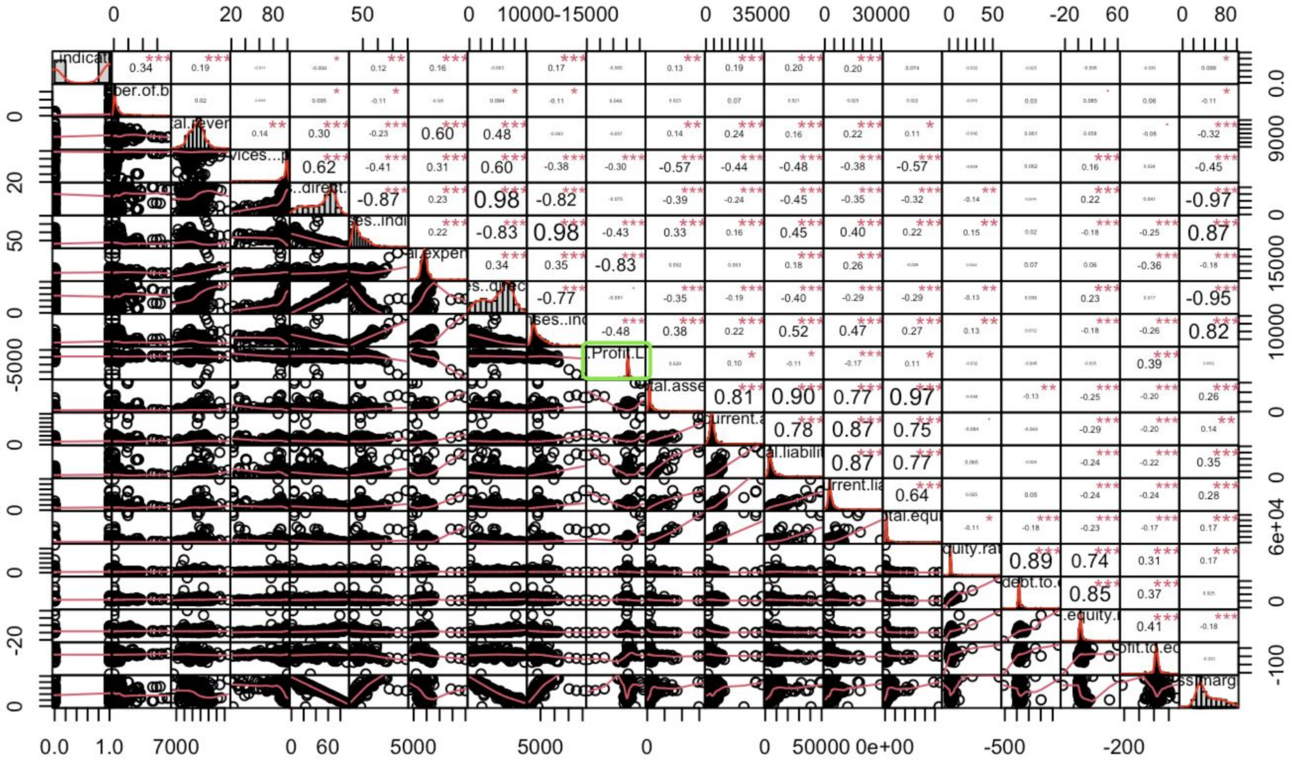
Presentation of Overview Dashboard

Discuss Research Questions and Modeling of RCBP and cross-analysis

Progress

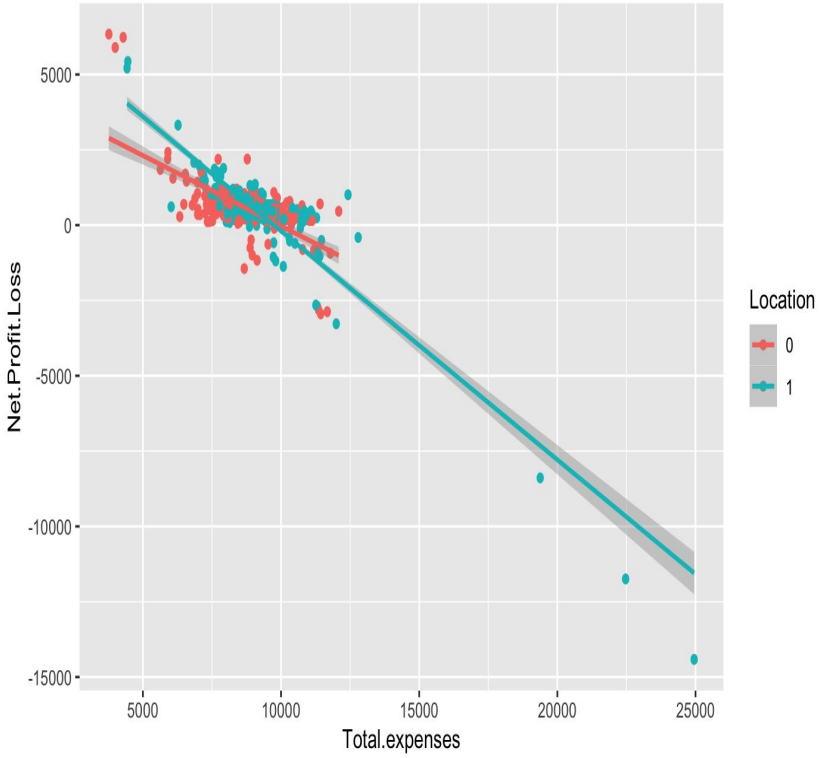
1. RCBP Modeling (Research Question 1)
    - Medium Business Modeling — Song (Alice)
    - Small Business Modeling — Yilin
2. Cross Analysis with unemployment rate dataset (Research Question 2)
    - Data wrangling and plotting — Bowen
    - Modeling — Tingwen
3. Start to write final paper

RCBP Modeling: Medium Business



RCBP Modeling: Medium Business

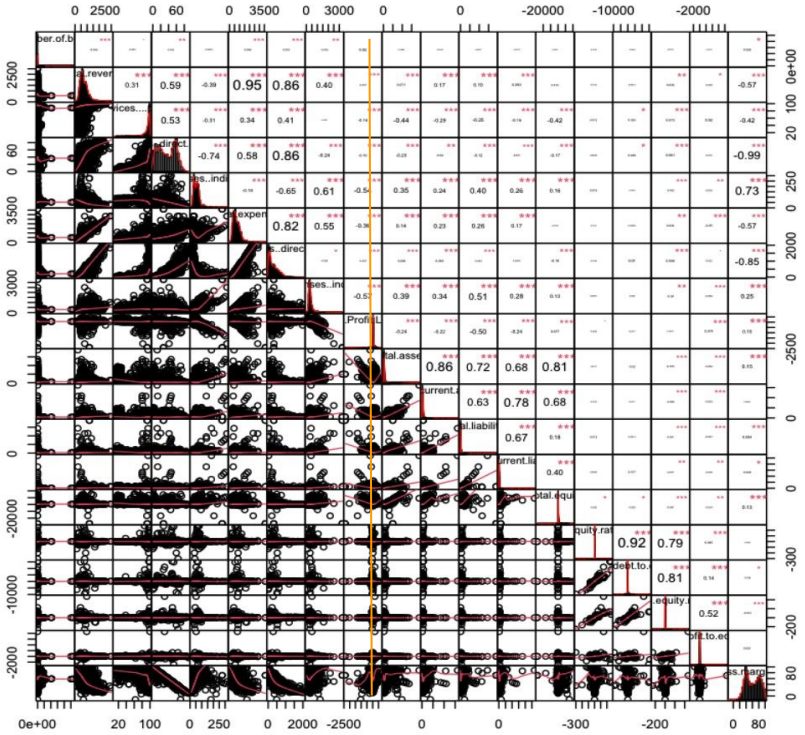
```
Call:
lm(formula = Net.Profit.Loss ~ Total.expenses * Location.indicator...code,
    data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2867.4  -327.7    37.1   350.5  3588.3

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                             4651.4470   321.5956  14.464  < 2e-16 ***
Total.expenses                            -0.4679     0.0367 -12.748  < 2e-16 ***
Location.indicator...code               2748.8947   391.6406   7.019 8.19e-12 ***
Total.expenses:Location.indicator...code   -0.2915     0.0438  -6.655 8.17e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 692 on 454 degrees of freedom
Multiple R-squared:  0.722,     Adjusted R-squared:  0.7202
F-statistic: 393.1 on 3 and 454 DF,  p-value: < 2.2e-16
```



RCBP Modeling: Medium Business

```
Call:
lm(formula = Net.Profit.Loss ~ ., data = mydata5)

Residuals:
    Min      1Q  Median      3Q     Max
-671.69  -35.44    5.26   36.78  515.03

Coefficients:
                                                    Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                        1.012e+04  1.229e+02   82.386  < 2e-16
Sales.of.goods.and.services...percent.of.total.revenue.  2.876e+00  7.676e-01    3.746 0.000203
Cost.of.sales..direct.expenses.....               -1.050e+02  1.768e+00  -59.387  < 2e-16
Operating.expenses..indirect.expenses.....        -1.017e+02  5.526e-01 -184.059  < 2e-16
Total.assets                                      -3.072e-03  6.663e-04   -4.611 5.23e-06
Total.current.assets                               2.303e-02  4.546e-03    5.065 5.98e-07
Total.current.liabilities                         -1.703e-02  5.392e-03   -3.157 0.001700
Revenue.to.equity.ratio..times.                    8.487e+00  1.338e+00    6.342 5.54e-10
Net.profit.to.equity....                          -2.539e+00  2.469e-01  -10.284  < 2e-16
Gross.margin....                                  -1.921e+00  1.604e+00   -1.198 0.231536

(Intercept)                                       ***
Sales.of.goods.and.services...percent.of.total.revenue. ***
Cost.of.sales..direct.expenses.....               ***
Operating.expenses..indirect.expenses.....        ***
Total.assets                                      ***
Total.current.assets                              ***
Total.current.liabilities                         **
Revenue.to.equity.ratio..times.                   ***
Net.profit.to.equity....                          ***
Gross.margin....
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 107.8 on 448 degrees of freedom
Multiple R-squared:  0.9933,     Adjusted R-squared:  0.9932
F-statistic:  7427 on 9 and 448 DF,  p-value: < 2.2e-16
```

RCBP Modeling: Small Business



RCBP Modeling: Small Business

```
Call:
lm(formula = Net.Profit.Loss ~ Cost.of.sales..direct.expenses..... +
    Operating.expenses..indirect.expenses..... + Location.indicator...code,
    data = sb1)

Residuals:
     Min       1Q   Median       3Q      Max
-1132.27   -11.88    -0.21    13.83   462.20

Coefficients:
                                        Estimate Std. Error t value
(Intercept)                            583.23391    3.35586  173.80
Cost.of.sales..direct.expenses.....     -5.48285    0.03595 -152.50
Operating.expenses..indirect.expenses.....  -6.21042    0.03310 -187.61
Location.indicator...code                4.03545    1.09969    3.67
                                       Pr(>|t|)
(Intercept)                             < 2e-16 ***
Cost.of.sales..direct.expenses.....     < 2e-16 ***
Operating.expenses..indirect.expenses.....  < 2e-16 ***
Location.indicator...code              0.000245 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.34 on 6268 degrees of freedom
  (4717 observations deleted due to missingness)
Multiple R-squared:  0.8507,    Adjusted R-squared:  0.8506
F-statistic: 1.191e+04 on 3 and 6268 DF,  p-value: < 2.2e-16
```

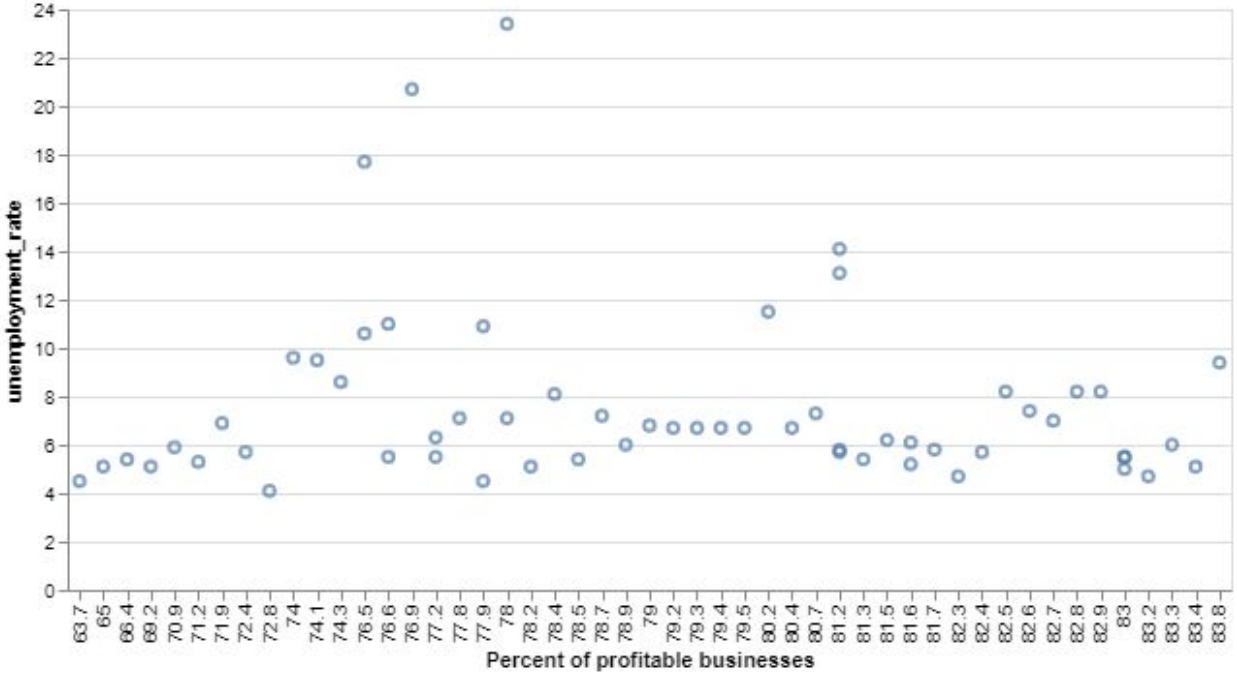
RCBP Modeling: Small Business

```
Call:
lm(formula = Net.Profit.Loss ~ ., data = my_data9)

Residuals:
     Min       1Q  Median      3Q     Max
-629.27  -11.90   -0.72   12.28  458.88

Coefficients:
                                        Estimate Std. Error t value
(Intercept)                             5.651e+02  2.648e+00  213.39
Cost.of.sales..direct.expenses.....    -5.334e+00  3.124e-02 -170.72
Operating.expenses..indirect.expenses..... -5.868e+00  3.102e-02 -189.13
Total.current.assets                    2.228e-02  1.362e-03   16.36
Total.liabilities                      -2.759e-02  5.548e-04  -49.73
Total.current.liabilities               3.042e-02  1.363e-03   22.32
                                       Pr(>|t|)
(Intercept)                              <2e-16 ***
Cost.of.sales..direct.expenses.....      <2e-16 ***
Operating.expenses..indirect.expenses.....   <2e-16 ***
Total.current.assets                     <2e-16 ***
Total.liabilities                        <2e-16 ***
Total.current.liabilities                <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.25 on 6264 degrees of freedom
  (4719 observations deleted due to missingness)
Multiple R-squared:  0.8956,    Adjusted R-squared:  0.8955
F-statistic: 1.074e+04 on 5 and 6264 DF,  p-value: < 2.2e-16
```

Cross-Analysis

Scatter plot and outliers

Unemployment rate higher than 10.5% came from the rural area of "Newfoundland and Labrador", "Prince Edward Island" and "New Brunswick" over 3 years.



Cross-Analysis

Outlier investigation

"Newfoundland and Labrador" and "New Brunswick" had highest population / number of businesses rate. This result partially implied the lack of job opportunities.

	Geography	Reference year	population	Total number of businesses	rate
0	Newfoundland and Labrador	2017	240942	9082	26.529619
1	Newfoundland and Labrador	2018	240942	9305	25.893821
2	Newfoundland and Labrador	2019	240942	9356	25.752672
9	New Brunswick	2017	285560	11163	25.580937
10	New Brunswick	2018	285560	11310	25.248453
11	New Brunswick	2019	285560	11327	25.210559
18	Manitoba	2017	376441	15648	24.056812
20	Manitoba	2019	376441	15779	23.857089
19	Manitoba	2018	376441	15871	23.718795
15	Ontario	2017	1424102	64260	22.161562

Cross-Analysis

Small rural business (remove outliers)

Unemployment rate higher than 10.5% came from the rural area of "Newfoundland and Labrador", "Prince Edward Island " and "New Brunswick" over 3 years.

```r
# load data
new_rural_small_rm <- read.csv('sb_merged_unemployment_rate_rm_rural_new.csv')

# linear regression model

unemployment_model2 <- lm(unemployment_rate ~ Percent.of.profitable.businesses*Total.number.of.businesses + Total.equity
+ Debt.to.equity.ratio..times., data= new_rural_small_rm)
summary(unemployment_model2)
```

result

```
Call:
lm(formula = unemployment_rate ~ Percent.of.profitable.businesses *
    Total.number.of.businesses + Total.equity + Debt.to.equity.ratio..times.,
    data = new_rural_small_rm)

Residuals:
     Min       1Q   Median       3Q      Max
-1.38642 -0.27186  0.05894  0.33291  1.13014

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                           -1.935e+01  1.926e+01  -1.005  0.33096
Percent.of.profitable.businesses                       3.978e-01  2.261e-01   1.759  0.09887 .
Total.number.of.businesses                             7.548e-04  2.619e-04   2.882  0.01141 *
Total.equity                                          -2.138e-03  4.807e-03  -0.445  0.66282
Debt.to.equity.ratio..times.                          -3.045e+00  1.975e+00  -1.542  0.14389
Percent.of.profitable.businesses:Total.number.of.businesses -9.962e-06  3.330e-06  -2.992  0.00912 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7392 on 15 degrees of freedom
Multiple R-squared:  0.7651,    Adjusted R-squared:  0.6868
F-statistic:  9.77 on 5 and 15 DF,  p-value: 0.0002629
```

Small Urban Businesses

```
Call:
lm(formula = unemployment_rate ~ Percent.of.profitable.businesses *
    Total.number.of.businesses + Total.equity + Debt.to.equity.ratio..times.,
    data = new_urban_small)

Residuals:
     Min       1Q   Median       3Q      Max
-1.07083 -0.47806 -0.09972  0.26081  1.78974

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                            3.155e+01  5.644e+00   5.590 9.41e-06 ***
Percent.of.profitable.businesses                      -2.731e-01  6.768e-02  -4.036 0.000481 ***
Total.number.of.businesses                             3.844e-05  4.521e-05   0.850 0.403529
Total.equity                                          -4.443e-03  2.899e-03  -1.533 0.138436
Debt.to.equity.ratio..times.                          -2.579e-01  1.134e+00  -0.227 0.822020
Percent.of.profitable.businesses:Total.number.of.businesses -4.918e-07  5.545e-07  -0.887 0.383910
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7442 on 24 degrees of freedom
Multiple R-squared:  0.7129,    Adjusted R-squared:  0.6531
F-statistic: 11.92 on 5 and 24 DF,  p-value: 7.171e-06
```

Final Paper

1. Executive summary
2. Introduction
3. Background and related work
4. Data
    - Data description
    - Data manipulation
5. Tools, Methodology, Techniques
    - RCBP Analysis
    - Cross Analysis
6. Analysis and Interpretation
    - Investigate several dimensions in RCBP
    - Cross Analysis -- Unemployment Model
7. Conclusion

Compare to original plan

6 (08 – 14 June)	Open-ended Question	Explore other related public StatCan data, do the cross-analysis with RCBP

Explore other related public StatCan data ✔

Do the cross-analysis with RCBP ✔

We are on the track of our original timeline in order to finish the project tasks.

Team Minutes

Meeting: Team + Clients 11 hr

1. 6/07 Team: Adjust cross analysis part plan 1 hr
2. 6/08 Client: Overview Dashboard presentation 1 hr
3. 6/08 Team: Distribute modeling and cross-analysis work 1 hr
4. 6/09 Team: Discuss the results of modeling and cross-analysis work 3 hr
5. 6/10 Team: Discuss final paper structure and review requirements 3 hr
6. 6/13 Team: Discuss how to modify the final paper 2hr

Individual Logs

Tingwen: 25 hr + 11 hr (team)

1. Final Report Written: 10 hr
2. Modelling: 5 hr
3. Cross analysis written: 5 hr
4. Presentation slides and preparation: 2 hr
5. Model adjustment: 3 hr

Individual Logs

Bowen Yang: 43 hr + 11 hr (team)

1. Prepare presentation: 2.5 hr
2. RCBP dashboard Polish: 4 hr
3. Cross analysis: 18.5 hr
4. Write final report: 18 hr

Individual Logs

Song(Alice) Zhang: 37.5 hr+11 hr (team)

1. RCBP Modelling: 15hr
2. Final Report: 19.5hr
3. Update weekly slides: 2hr
4. Take minutes and personal logs: 1hr

Individual Logs

Yilin Sun: 26 hr + 11 hr (team minutes)

1. RCBP Small Business Modeling: 10 hr
2. Final Paper Draft: 12 hr
3. Make presentation slides and practice presentation: 3 hr
4. Update team and personal logs: 1 hr

Next Week Plan

Finish Final Paper

Finish Final Presentation Slides

Modify dashboards, analytical paper based on client suggestions