STAA57 Group Project

What are the dominant research areas in Ontario's automotive sector, and how do institutions specialize in different fields

Minh Tran - 1006804914, Swajeet Jadhav - 1009888276

2025-03-15

I. Introduction

Research Goal

With the growing popularity of electric vehicles (EVs) and the rapid advancements in artificial intelligence (AI), particularly in the development of autonomous vehicles like Tesla, it is crucial for automotive research facilities in Ontario to focus their efforts and resources on the right areas. This will ensure a significant contribution to the academia world while assisting the government in implementing effective policies and educational institutions better educate, prepare the next generation of labors.

This report aims to examine the automotive research areas prioritizing by Ontario's research facilities and identify specialized facilities in trending fields. It will also explore which areas are attracting the most researchers and which institutions are specializing in particular research domains.

Dataset description

Data source

Our dataset was collected by researchers from teh Ministry of Economic Development, Job creation and Trade and was published on the Government of Ontario's public database in 2018.

Table 1: Data set Variables

```
## Rows: 545
## Columns: 37
## $ Institution                      <chr> "Brock University", "Carleton Univers~
## $ Researcher.Name                  <chr> "Ahmed, Syed Ejaz", "Ahmadi, Mojtaba"~
## $ Associated.Facilities            <chr> "Centre for Statistical Consulting", ~
## $ Research.Areas                   <chr> "Traffic and road injury prevention."~
## $ Research.Chairs.Grant.Funding    <chr> "", "", "", "", "Canada Research Chai~
## $ Tag.1                            <chr> "Injury Prevention", "Mechatronics", ~
## $ Tag.2                            <chr> "", "Control", "Autonomy and AI", "Ne~
## $ Tag.3                            <chr> "", "", "Sensors", "", "", "", "", ""~
## $ Tag.4                            <chr> "", "", "", "", "", "", "", "", "", "~
## $ Tag.5                            <chr> "", "", "", "", "", "", "", "", "", "~
## $ Alternative.Fuels                <chr> "", "", "", "", "", "", "", "", "", "~
## $ Autonomy.and.AI                  <chr> "", "", "x", "", "", "", "x", "", "",~
## $ Batteries.and.Fuel.Cells         <chr> "", "", "", "", "", "", "", "", "", "~
## $ Biocomposites                    <chr> "", "", "", "", "", "", "", "", "", "~
## $ Coatings.and.Corrosion           <chr> "", "", "", "", "", "", "", "", "", "~
## $ Connected.Vehicles               <chr> "", "", "", "", "", "", "", "", "", "~
## $ Control                          <chr> "", "x", "x", "", "", "", "", "", "",~
## $ Crashworthiness                  <chr> "", "", "", "", "", "", "", "", "", "~
## $ Electronics                      <chr> "", "", "", "", "", "", "", "", "", "~
## $ Forming.and.Joining              <chr> "", "", "", "", "", "", "", "", "", "~
## $ High.Strength.Steel              <chr> "", "", "", "", "", "", "", "", "", "~
## $ Hybrid.and.Electric.Vehicles     <chr> "", "", "", "", "", "", "", "", "", "~
## $ Industrial.Processes             <chr> "", "", "", "", "", "", "", "", "", "~
## $ Injury.Prevention                <chr> "x", "", "", "", "", "", "", "", "", ~
## $ Internal.Combustion.Engines      <chr> "", "", "", "", "", "", "", "", "", "~
## $ Lightweight.Metals               <chr> "", "", "", "", "", "", "", "", "", "~
## $ Mechatronics                     <chr> "", "x", "", "", "", "", "", "", "", ~
## $ Nanotechnology                   <chr> "", "", "", "", "", "", "", "", "", "~
## $ Networks.and.Security            <chr> "", "", "", "x", "", "x", "", "x", "x~
## $ Noise..Vibration.and.Harshness   <chr> "Err:507", "Err:507", "Err:507", "Err~
## $ Polymers.and.Composite.Materials <chr> "", "", "", "", "", "", "", "", "", "~
```

```
## $ Powertrain                      <chr> "", "", "", "", "", "", "", "", "", "~
## $ Sensors                         <chr> "", "", "x", "", "x", "", "x", "", ""~
## $ Software                        <chr> "", "", "", "x", "", "", "", "", "", ~
## $ Stress.and.Fracture             <chr> "", "", "", "", "", "", "", "", "", "~
## $ Transportation.and.Charging     <chr> "", "", "", "", "", "", "", "", "", "~
## $ Vehicle.Design                  <chr> "", "", "", "", "", "", "", "", "", "~
```

Based on a summary of the data set, the `Noise..Viration.and.Harshness` variable was corrupted during the process of uploading or collecting the data and therefore will not be used in our report

```
# Remove variable with errors
researchers = researchers %>% select(!Noise..Vibration.and.Harshness )
```

Some of the main categorical variables that can help categorizing the data include

- **Researcher.Name**
- **Institution**: Name of university or research center that the researcher works at
- **Research.Areas**: Researcher general field of research
- **Tag**: Researcher's specialized field of research

In addition, the data set also contains research fields as variable (e.g. Alternative Fuels, Autonomy and AI, Vehicle Design) to indicate whether or not a researcher work fall into that field of research

II. Data Overview Analysis

Descriptive Statistic

- Table 2: Summary statistics of research areas (count, frequency)
  - Identify the most common research fields

```
research_count = researchers %>% group_by(Tag.1) %>%
  summarise(Count = n()) %>%
  mutate(Frequency = Count / sum(Count)) %>%
  arrange(desc(Count))
top_research_field = research_count$Tag.1[1]
kable(research_count, caption = "Summary statistics of research areas")
```

Table 1: Summary statistics of research areas

Tag.1	Count	Frequency
Networks and security	67	0.1229358
Autonomy and AI	45	0.0825688
Transportation and charging	38	0.0697248
Industrial processes	34	0.0623853
Software	32	0.0587156
Lightweight metals	28	0.0513761
Polymers and composite materials	28	0.0513761
Batteries and fuel cells	26	0.0477064
Connected vehicles	23	0.0422018
Forming and joining	21	0.0385321
Vehicle design	21	0.0385321
Alternative fuels	18	0.0330275
Biocomposites	18	0.0330275
Electronics	16	0.0293578

Tag.1	Count	Frequency
Nanotechnology	16	0.0293578
Control	14	0.0256881
Sensors	14	0.0256881
Coatings and corrosion	13	0.0238532
Internal combustion engines	13	0.0238532
Injury Prevention	12	0.0220183
Hybrid and electric vehicles	9	0.0165138
Mechatronics	7	0.0128440
Noise, vibration and harshness	7	0.0128440
Powertrain	6	0.0110092
Stress and fracture	6	0.0110092
High strength steel	5	0.0091743
Crashworthiness	4	0.0073394
Other	4	0.0073394

- Table 3: Top 5 institutions with the most researchers in the top research field
  - Identify the most dominant institution in a field

```
top_institutions = researchers %>%
  filter(Tag.1 == top_research_field) %>%
  count(Institution, sort = TRUE)%>%
  rename(Reseacher_count = n)%>% head(5)
kable(top_institutions, caption = paste("Top 5 Institutions in", top_research_field))
```

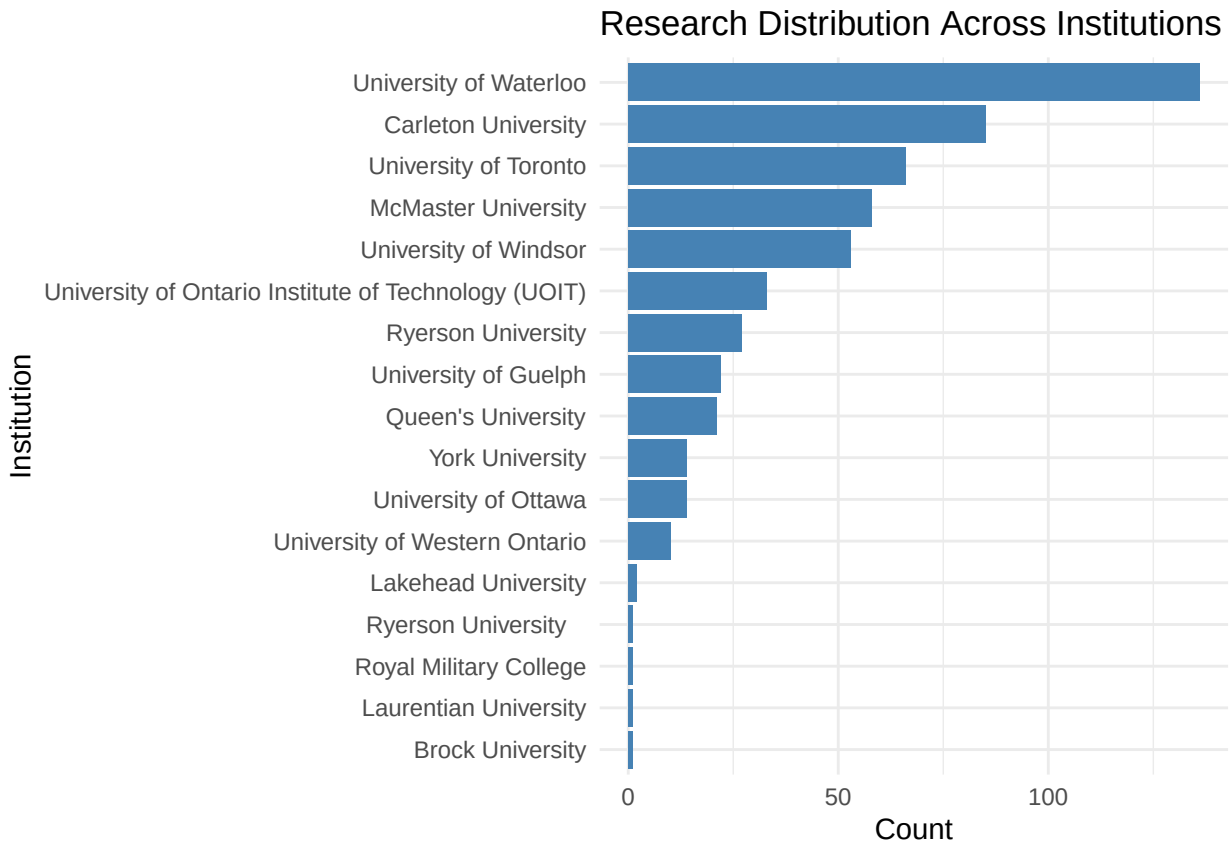
Table 2: Top 5 Institutions in Networks and security

Institution	Reseacher_count
Carleton University	33
University of Waterloo	15
Ryerson University	6
University of Ontario Institute of Technology (UOIT)	3
University of Ottawa	3

Graph & Visualizations

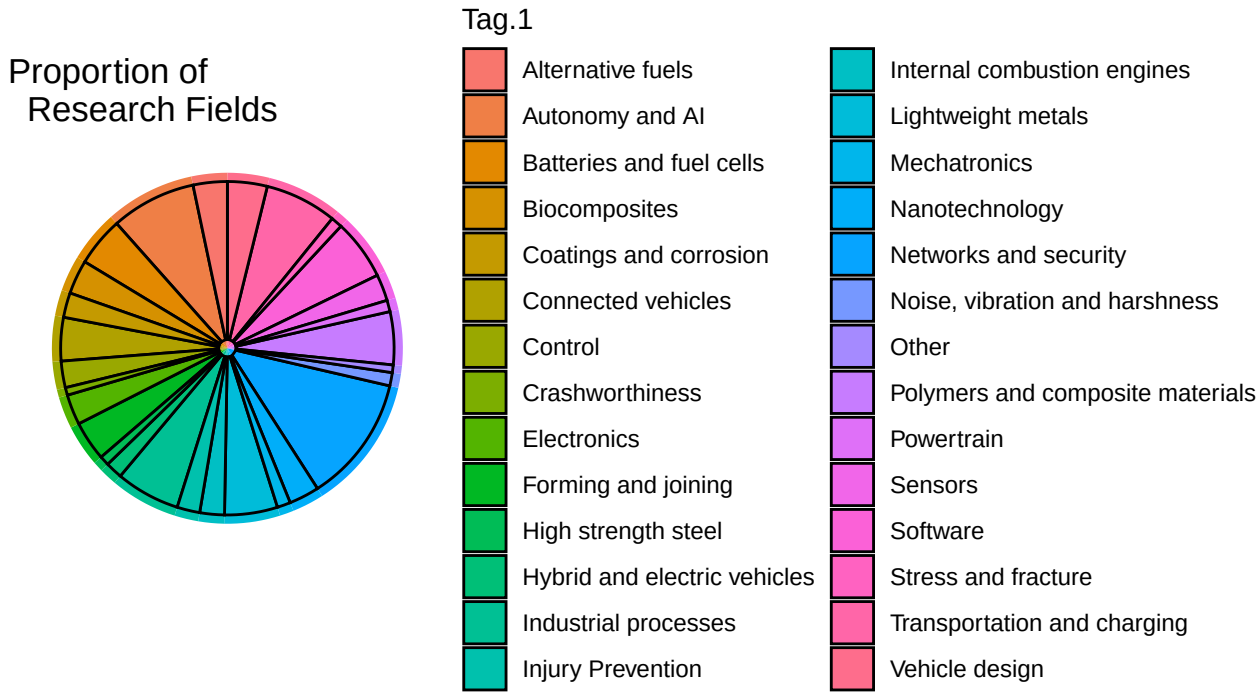
- Bar chart: Research distribution across institution

```
ggplot(researchers %>% count(Institution, sort = TRUE), aes(x = reorder(Institution, n), y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Research Distribution Across Institutions",
       x = "Institution", y = "Count")
```



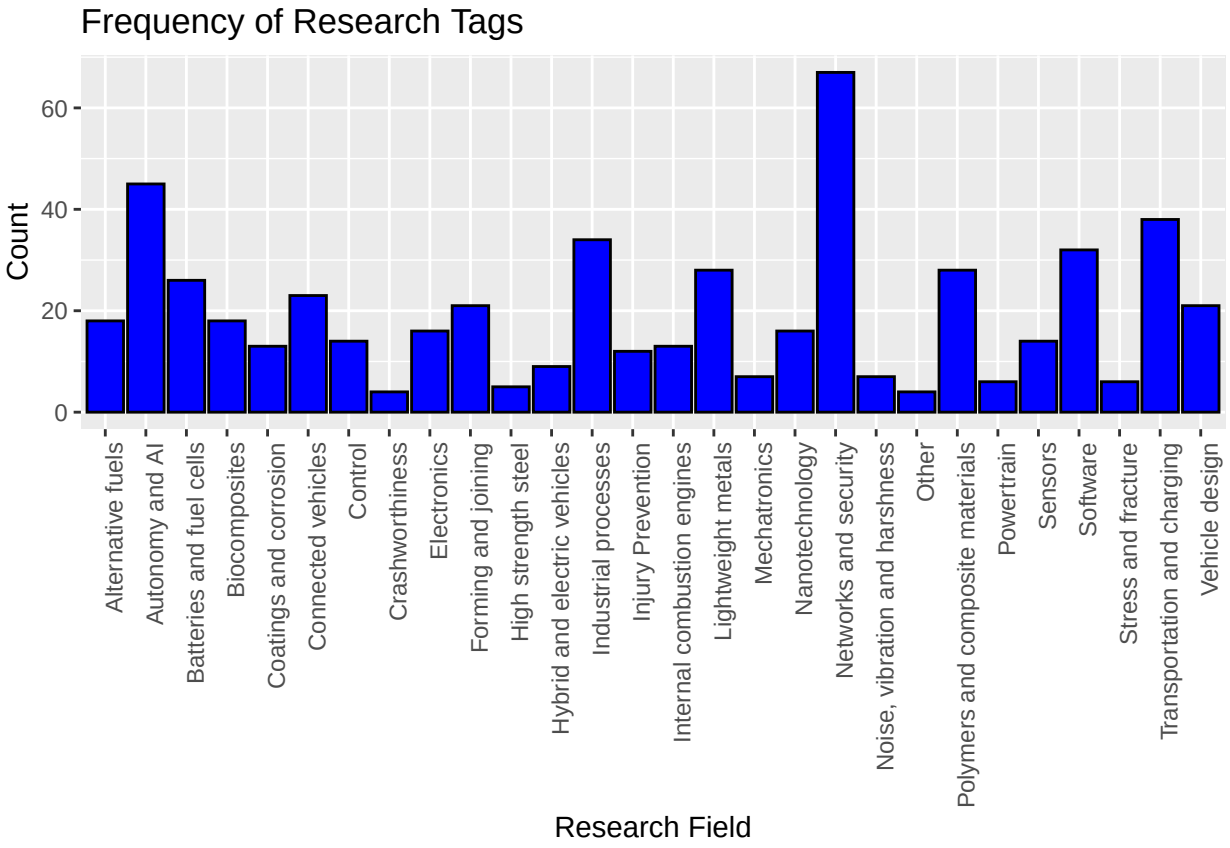
- Pie chart: Proportion of different research fields

```
ggplot(research_count, aes(x = "", y = Count, fill = Tag.1)) +
  geom_bar(stat = "identity", width = 1) +  geom_col(color = "black") +
  coord_polar("y", start = 0) +
  theme_void() +
  labs(title = "Proportion of
  Research Fields")
```



- History, frequency of research tag

```
ggplot(researchers, aes(x = Tag.1)) +
  geom_bar(fill = "blue", color = "black") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Frequency of Research Tags", x = "Research Field", y = "Count")
```



III. Statistical Analysis

We could pick an area we think that will be most trending (AI for example) and test to see if that is true

- Confidence interval for the average number of researcher
- Hypothesis Testing
- Bootstrapping

IV. Predictive modeling (Regression)

Build a regression model to predict number of researcher in a field, institutions

Build a regression model to see if an institution is likely to have to a grant funding

Using cross-validation to validate the 2 models built above

V. Summary

Appendix