

BC Stats Proposal

MDS Capstone Project - Quantifying the Responses to Open-Ended Survey Questions

Ayla Pearson | Fan Nie | Aaron Quinton

April 26, 2019

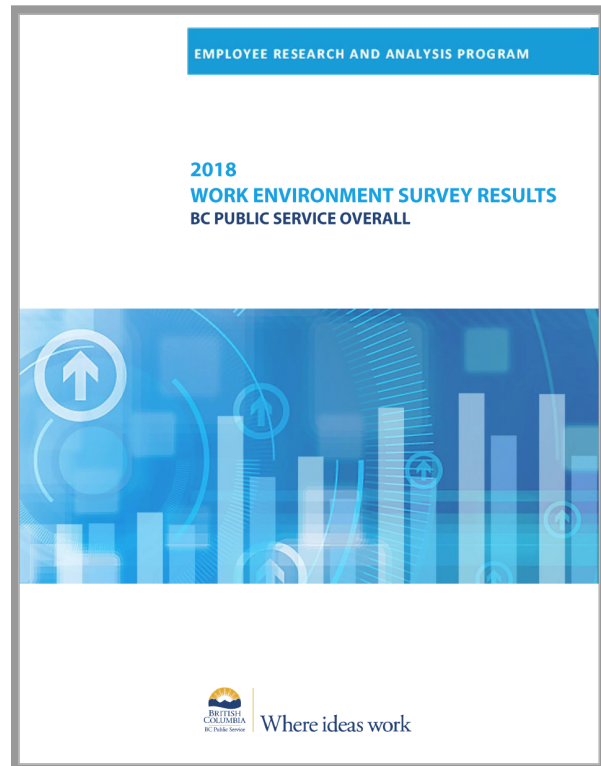
Introduction

Background

- Work Environment Survey (WES)

Primary goals:

- Understand employee experiences
- Celebrate successes
- Identify improvements



Introduction

Data

There are over 15,000 respondents across 26 ministries and over 8 survey cycles starting in 2007. All of the Survey data can be categorized in two:

Quantitative data (Likert Scale questions)

Qualitative data (comments/open ended responses)

Introduction

Quantitative Data

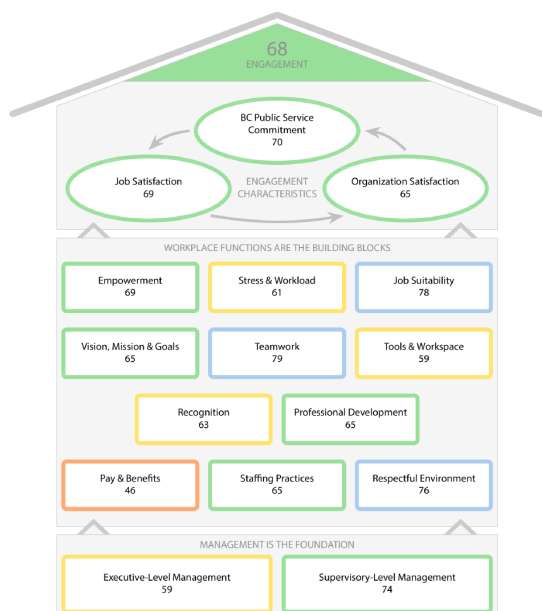
Approximatley 80 quantitative questions all related to **day-to-day work, management, performance, pay & benefits, etc.**

Examples:

- A healthy atmosphere exists in my work unit: [1-5]
- The quality of training and development I have received is satisfactory: [1-5]

Introduction

Quantitative Data



- aggregated multiple choice answers by ministry
- aggregated multiple choice answers overall
- engagement model highlighting key drivers

The current quantitative analysis report includes:

Introduction

Qualitative Data

What one thing would you like your organization to focus on to improve your work environment?

Example Themes:

- Compensation and Benefits
- Stress and Workload
- Supervisors
- etc.

Introduction

Qualitative Data

Example Comment:

Theme: Tools, Equipment & Physical Environment

Sub-theme: Provide better computer-based hardware

Introduction

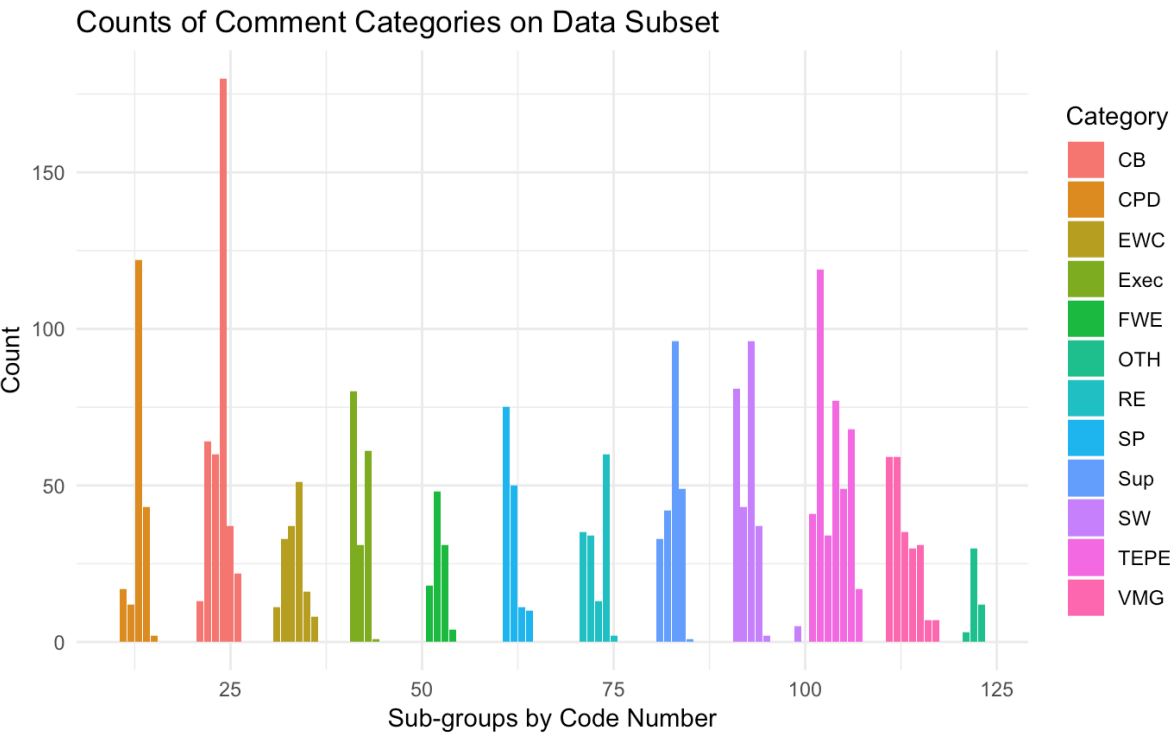
Qualitative Data

The current qualitative analysis report includes:

- manual coding of comments into 12 themes and ~60 sub-themes
- ranking areas of improvement overall and by ministry

Introduction

Qualitative Data



Research Question

1. **Coding Themes** - What is the best method to code the themes and sub-themes to the qualitative responses?
2. **Linking Quantitative to Qualitative** - How well does the sentiment of the qualitative responses agree with the quantitative responses?
3. **Trends** - What trends in key engagement drivers and multiple choice questions exist over time and across departments from the 2008 to 2018 quantitative survey data?

Data Science Techniques

1. Coding Themes

Approach:

- Themes are previously specified and labeled (Supervised Learning)
- Base Model: Bag of words and SVM Classification
- Complex Model: Multiclass and Multilabel ensemble classification models

Challenges:

- 60+ levels in Classification
- Multiple levels can apply to one comment

Deliverables:

- Trained Model for future labelling

Data Science Techniques

2. Linking Quantitative to Qualitative

Approach:

- Sentiment Analysis and Emotion Recognition Ref.

(https://en.wikipedia.org/wiki/Emotion_recognition)

Inter-rater reliability

Metrics:

- Percent agreement calculation
- Cohen's kappa
- Krippendorff's alpha

Challenges:

- Comments may lack sentiment
- Or are primarily negative

Deliverables:

- Report
- Dashboard

Data Science Techniques

3. Trends Across Ministries and Overtime

Approach:

- Descriptive statistics and visualization
- Base Model: Moving Average
- Complex Model: Regression analysis

Challenges:

- Messy data and inconsistent survey questions over time

Deliverables:

- Dashboard

Timeline

