Using R for Advanced Analytics in the Department of National Defence

Yvan Gauthier[1,2]

[1] Centre for Operational Research & Analysis, ADM(DRDC)
[2] Directorate of Data Science, ADM(DIA)
Department of National Defence

DRDC | RDDC

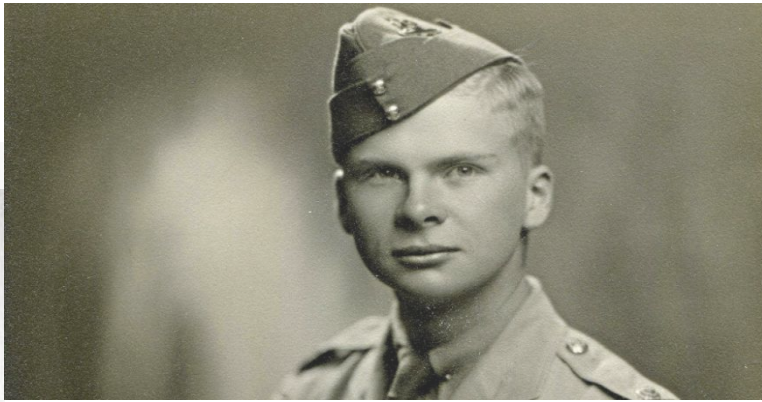
Presentation to the R Conference for the Government & Public Sector

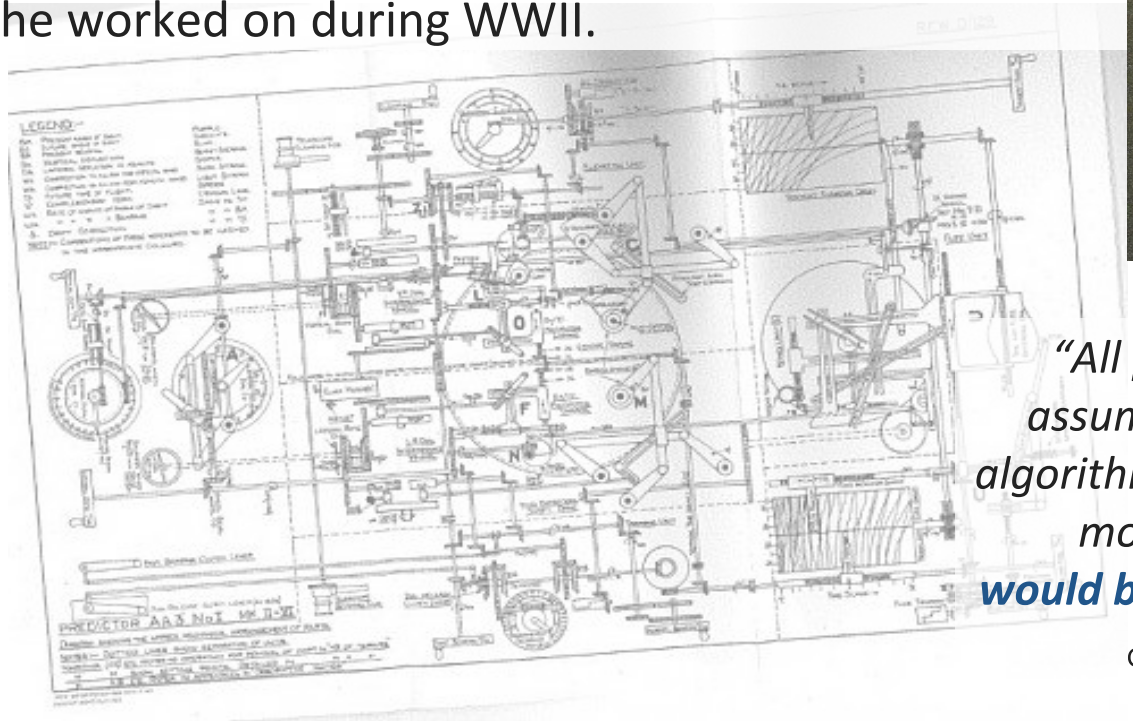
04 December 2020

Canada

DND's Long Tradition of Data Analytics…

Dr. George Lindsey's recollections on **predictors** (automated anti-aircraft fire-control systems) that he worked on during WWII.





"All predictors had to be provided with an assumption (which would now be called an algorithm) regarding the future motion of the moving target [...] ***Today such problems would be classified as artificial intelligence."***

G. Lindsey, Canadian Military History, Volume 4, Issue 2, 2012

Operations Research

STATISTICAL DATA USEFUL FOR THE OPERATION OF A BASEBALL TEAM

G. R. Lindsey

Operational Research Branch, Air Defence Command, Royal Canadian Air Force, St. Hubert, Quebec

(Received August 10, 1958)

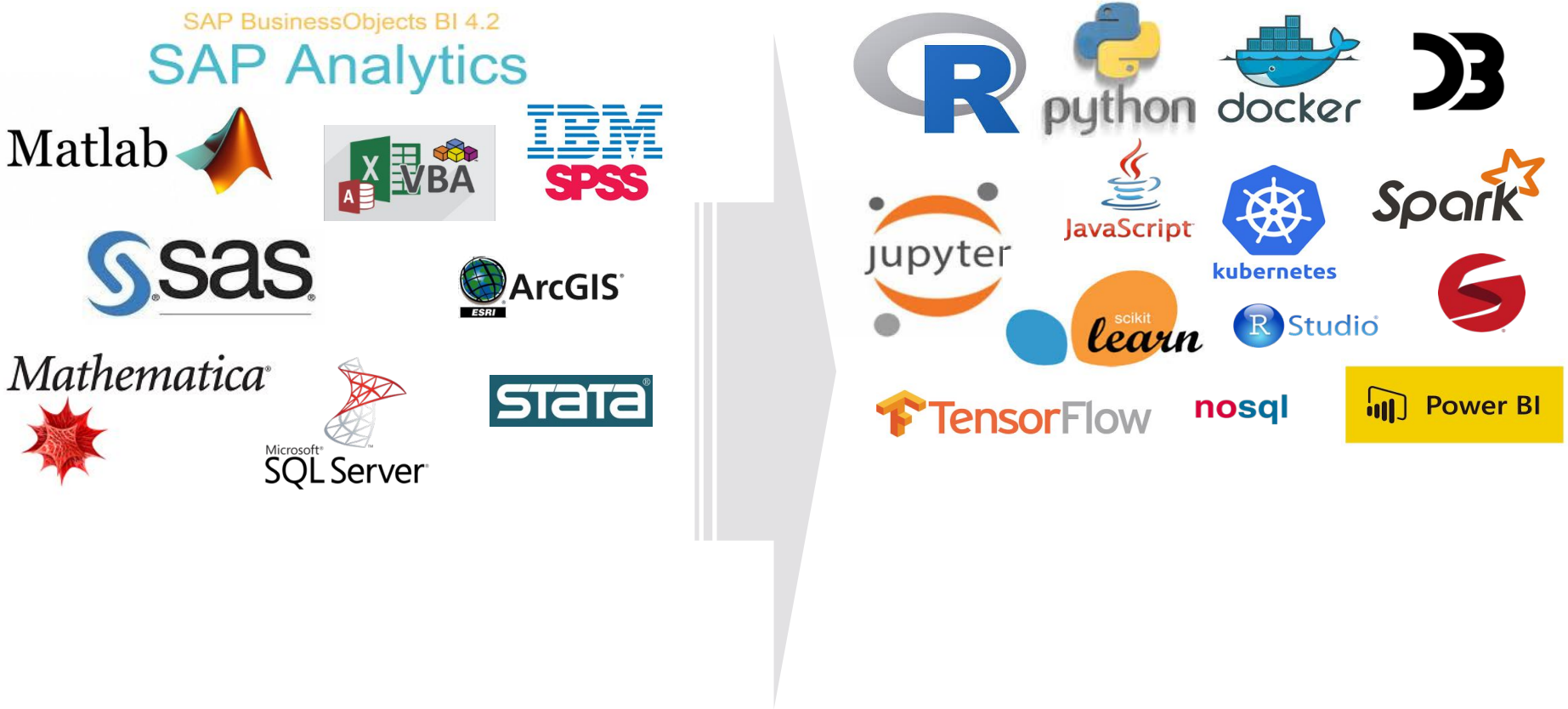
Baseball is a game well suited for operations analysis, and is already well provided with statistical records of past performances of individuals. For the making of operational decisions before and during a game, the manager must use the existing records of past performances with caution, but he could derive considerable assistance from types of statistics not

The same old thing?

Data Science Team

- Team of data scientists embedded with Chief Data Officer organisation
- Currently focuses on defence management problems involving:
  - advanced methods, including ML
  - large sets of enterprise data
  - model deployment into IT landscape for recurring decision support

“Skate where the puck is going to be, not where it has been”


SAP BusinessObjects BI 4.2
SAP Analytics
Matlab
VBA
IBM
SPSS
sas
ArcGIS
ESRI
Mathematica
Microsoft
SQL Server
STATA
python
docker
jupyter
JavaScript
kubernetes
Spark
scikit
learn
R Studio
TensorFlow
nosql
Power BI

SAP BusinessObjects BI 4.2

SAP Analytics

Matlab

VBA

IBM SPSS

sas

ArcGIS

Mathematica

Microsoft SQL Server

STATA

R

python

docker

D3

jupyter

JavaScript

kubernetes

Spark

scikit learn

R Studio

TensorFlow

nosql

Power BI

ON-PREMISE

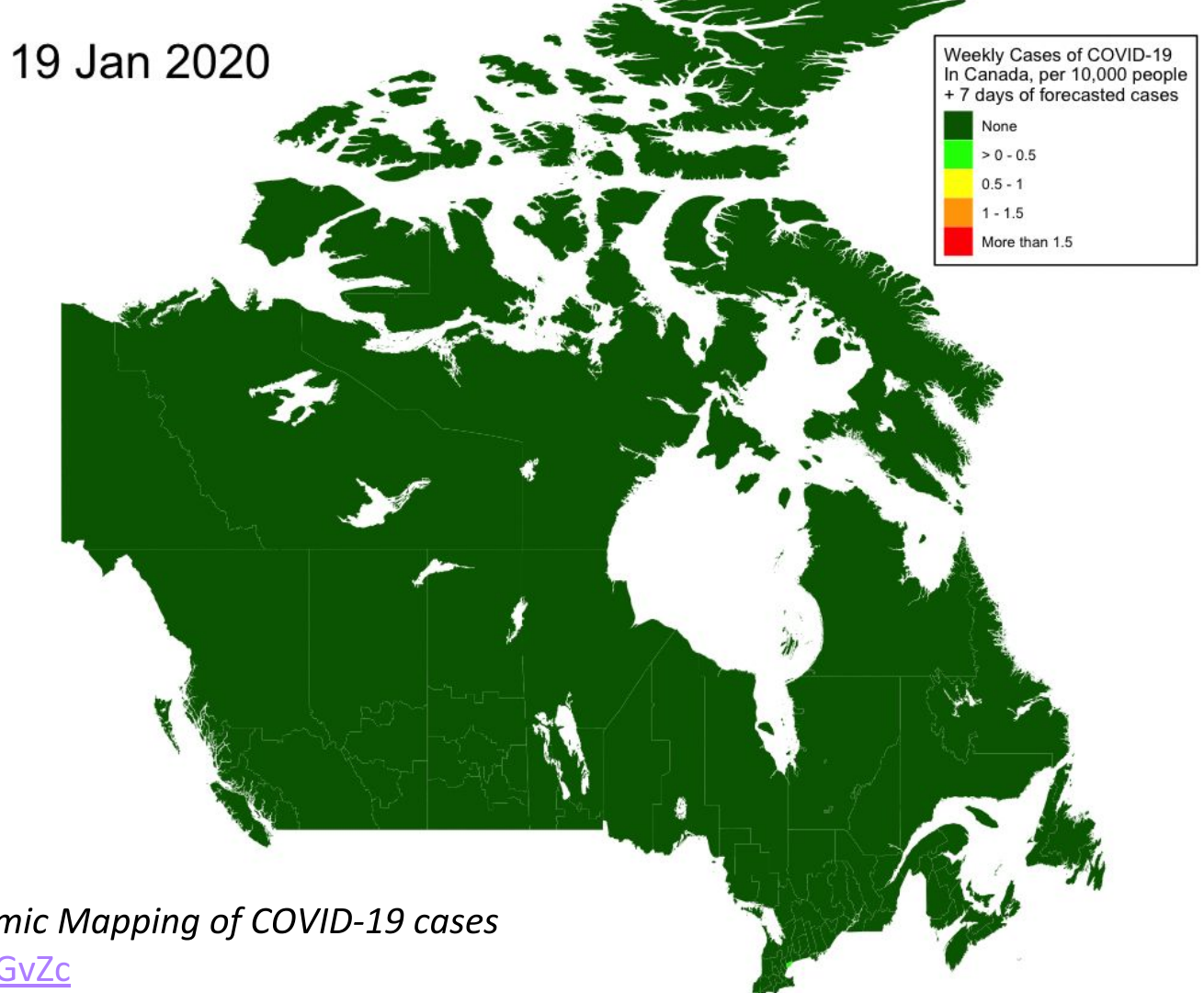
HYBRID

CLOUD

Some examples of recent projects...

COVID-19 and Operation LASER

COVID-19 and Operation LASER

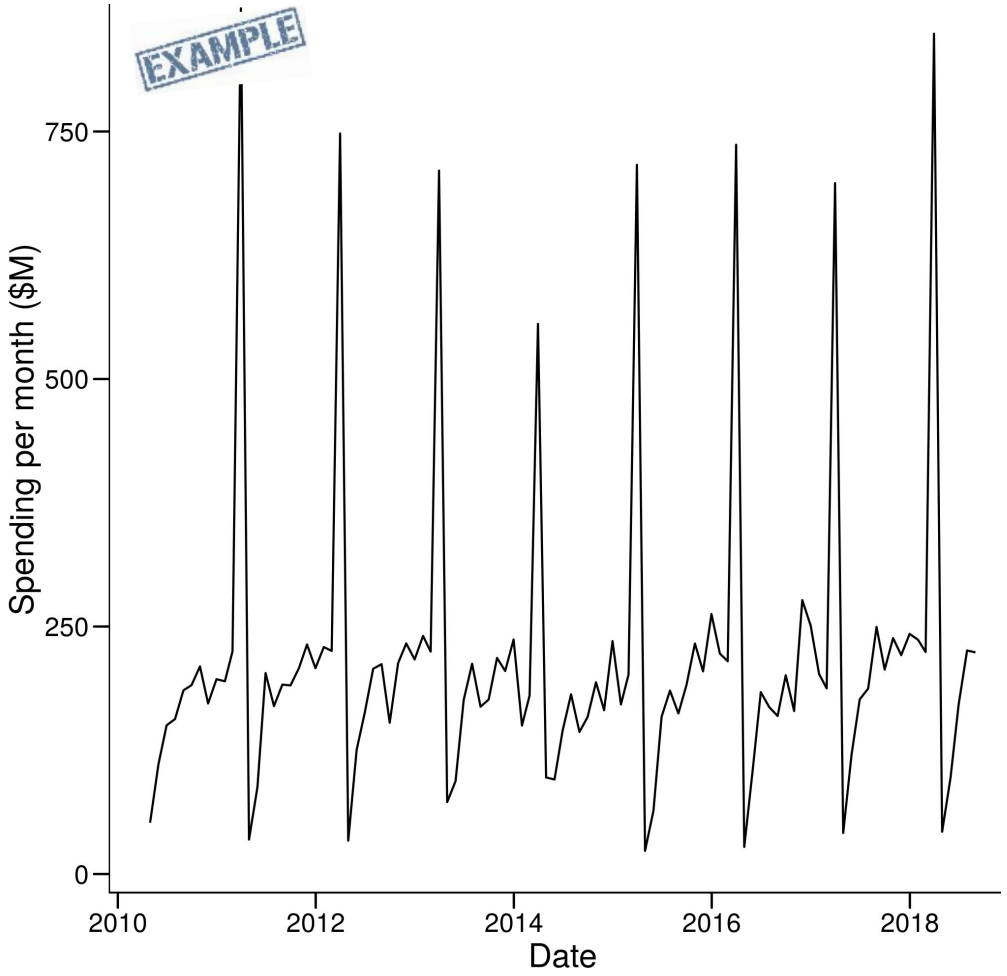


Source: T. Beech et al., *Dynamic Mapping of COVID-19 cases in Canada*, https://bit.ly/39FGvZc

Source: T. Beech et al., *Predictive Mapping of new COVID-19 cases in Canada*, https://bit.ly/394UuHy

“March Madness”

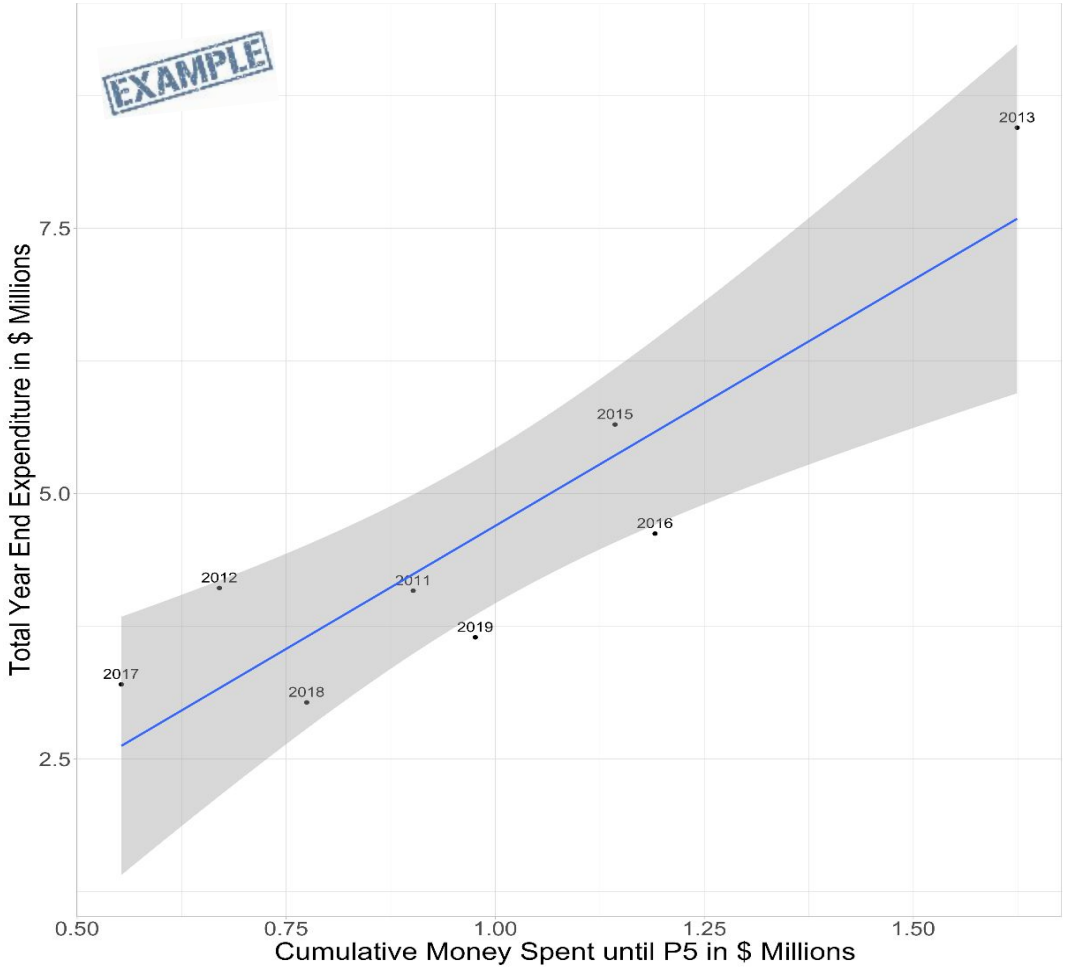
- Financial expenditures are seasonal, with a notable surge at the end of every fiscal year (in March in the Government of Canada)
- Predicting total expenditures early in fiscal year can help reallocate funds as needed.



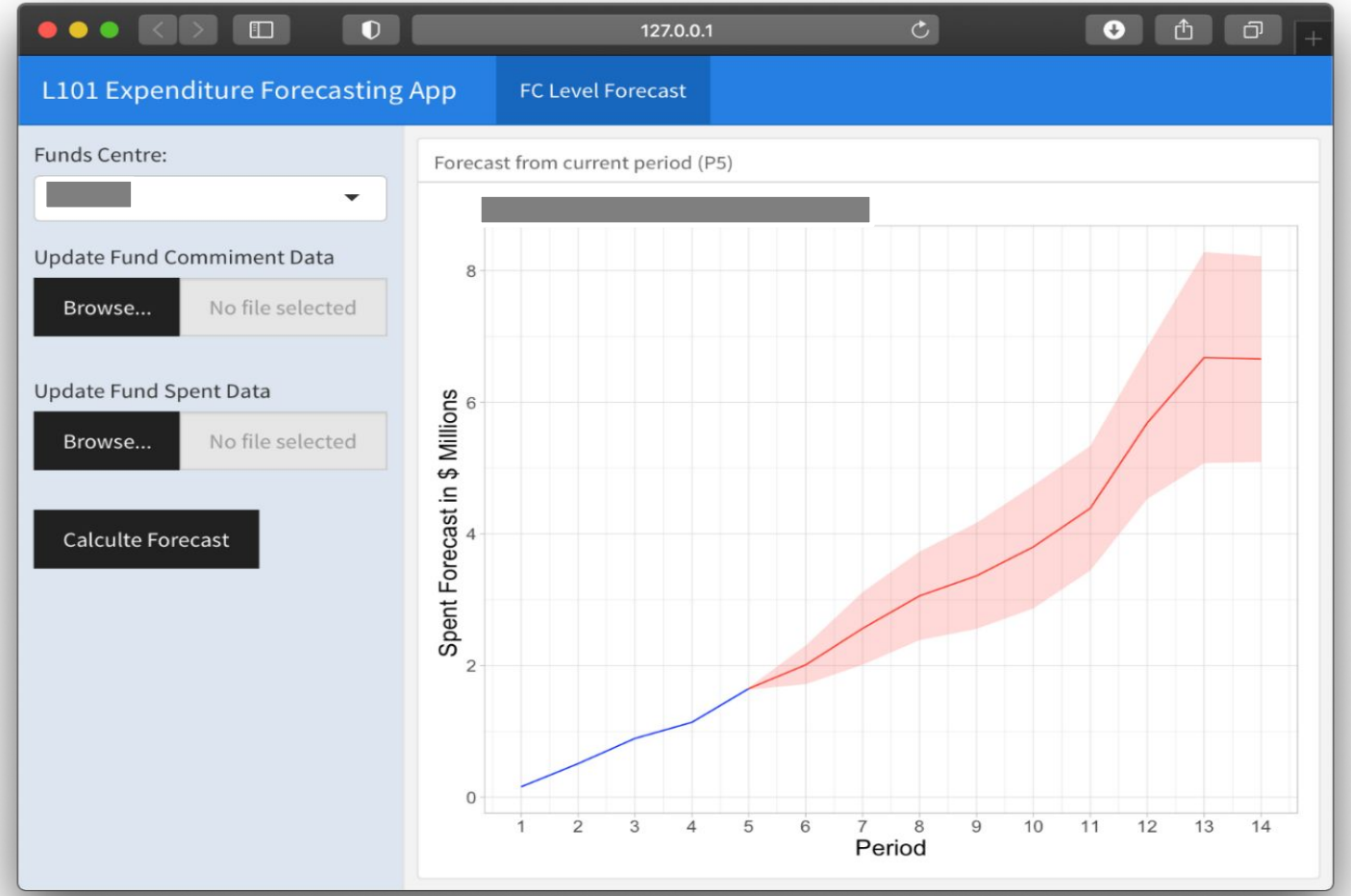
Source: Levesque (2018)

We can leverage historical spending patterns to predict expenditures at year end by doing regressions on different variables, e.g.,

- Money spent after five months (P5)
- Money committed after P5
- etc.



Source: Gunasekara and Beech (2020)



Source: Gunasekara and Beech (2020)

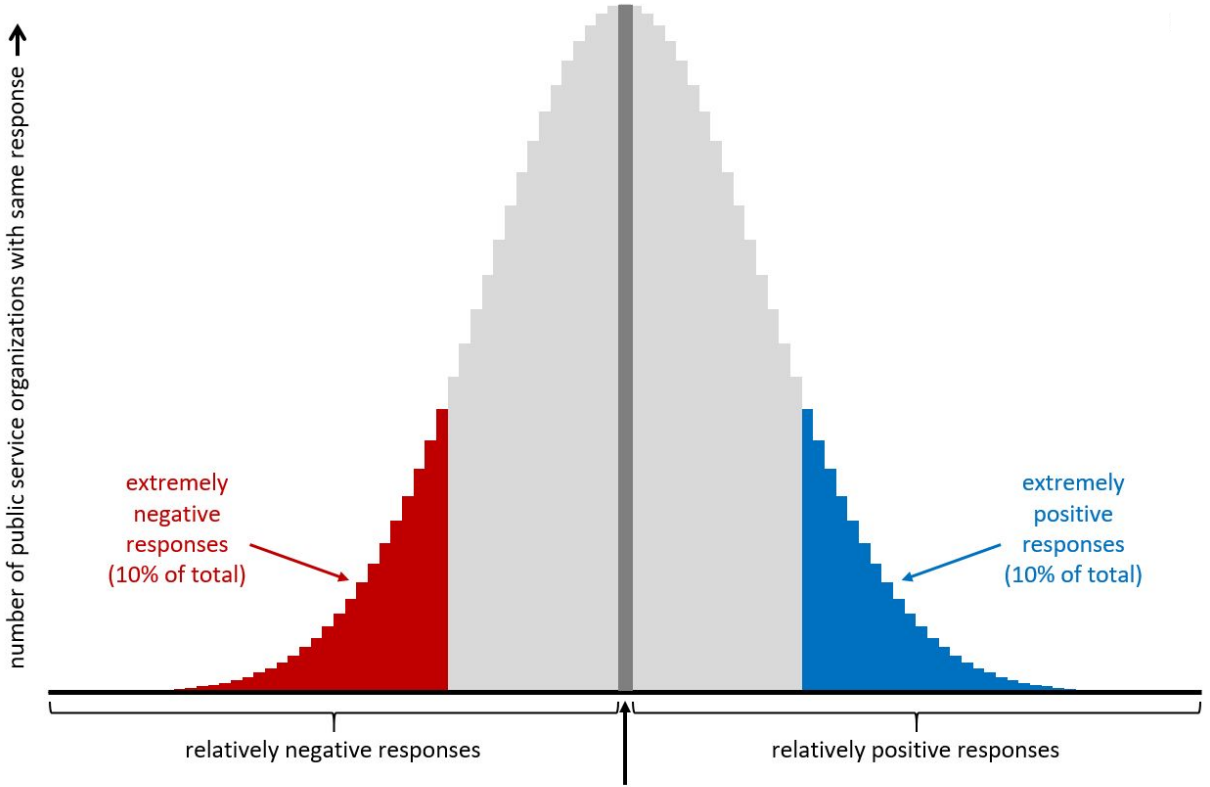
Public Service Employee Survey

- Conducted annually to obtain the views of all employees of the federal Public Service about their workplaces.
- Over 180 questions organized in terms of ***6 themes*** and 18 *sub-themes*

- ***Employee engagement***
  - *Employee engagement*
- ***Leadership***
  - *Immediate supervisor*
  - *Senior management*
- ***Workforce***
  - *Performance management*
  - *Job fit & development*
  - *Empowerment*
  - *Work-life balance & workload*
  - *Mobility & retention*
- ***Workplace***
  - *Organizational goals*
  - *Organizational performance*
  - *Respectful workplace*
  - *Ethical workplace*
  - *Harassment*
  - *Discrimination*
- ***Workplace well-being***
  - *A psychologically healthy workplace*
  - *Work-related stress*
- ***Compensation***
  - *Pay or other compensation issues*
  - *Support to resolve pay or other compensation issues*

Public Service Employee Survey
Peer comparisons

- New dashboards help to rapidly identify where an organisation is **better** (or **worse)** than 90% of same-level organisations in the Public Service
- Subsequent questions:
  - *What **causes** some exceptional results?*
  - *Can we generate **actionable insights** using more advanced techniques?*



Source: Dooley and Gauthier (2018)

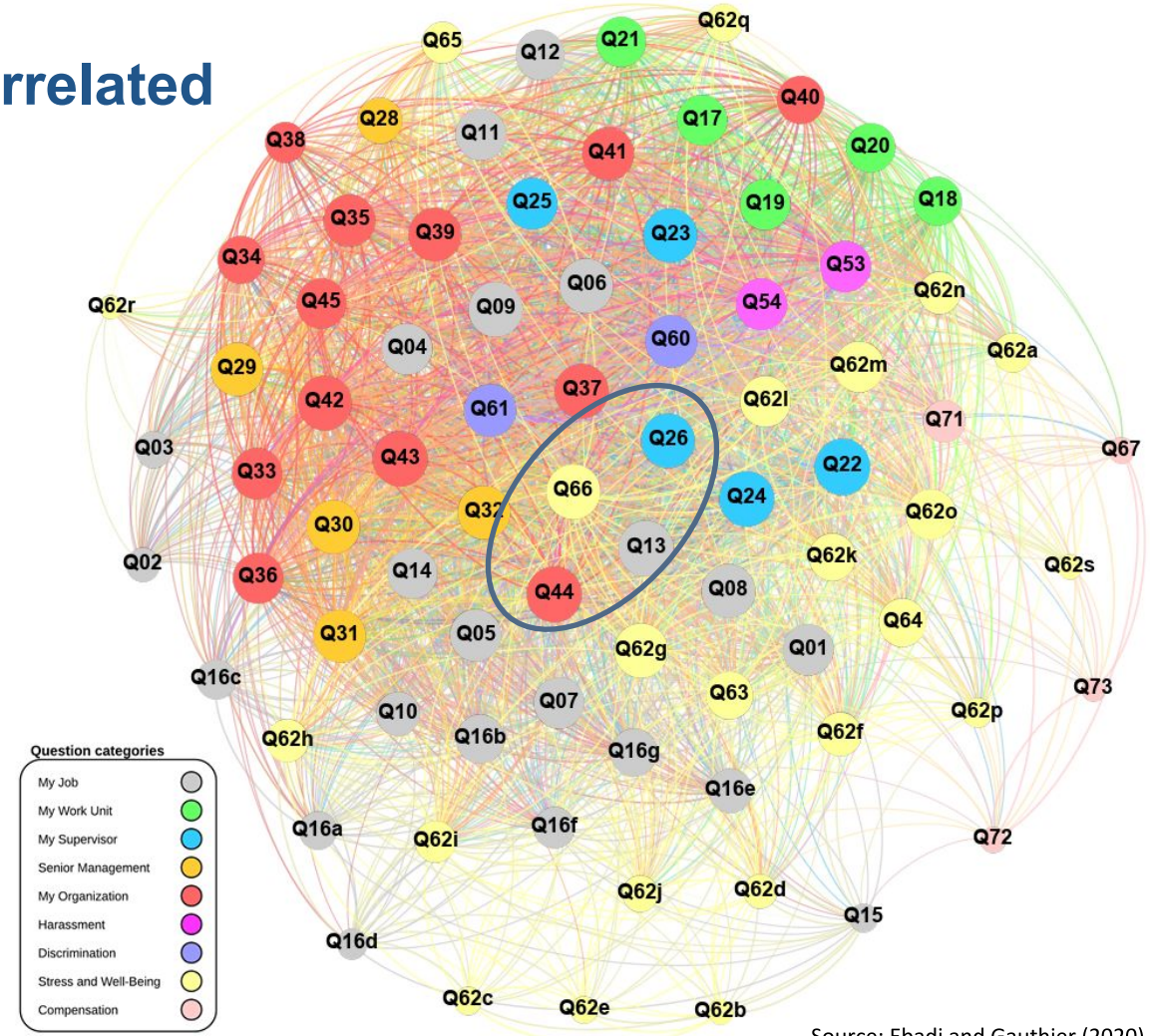
Survey responses are correlated

Responses that are most highly correlated with others are seen in the **center** of this graph

EXAMPLE

Some central questions…

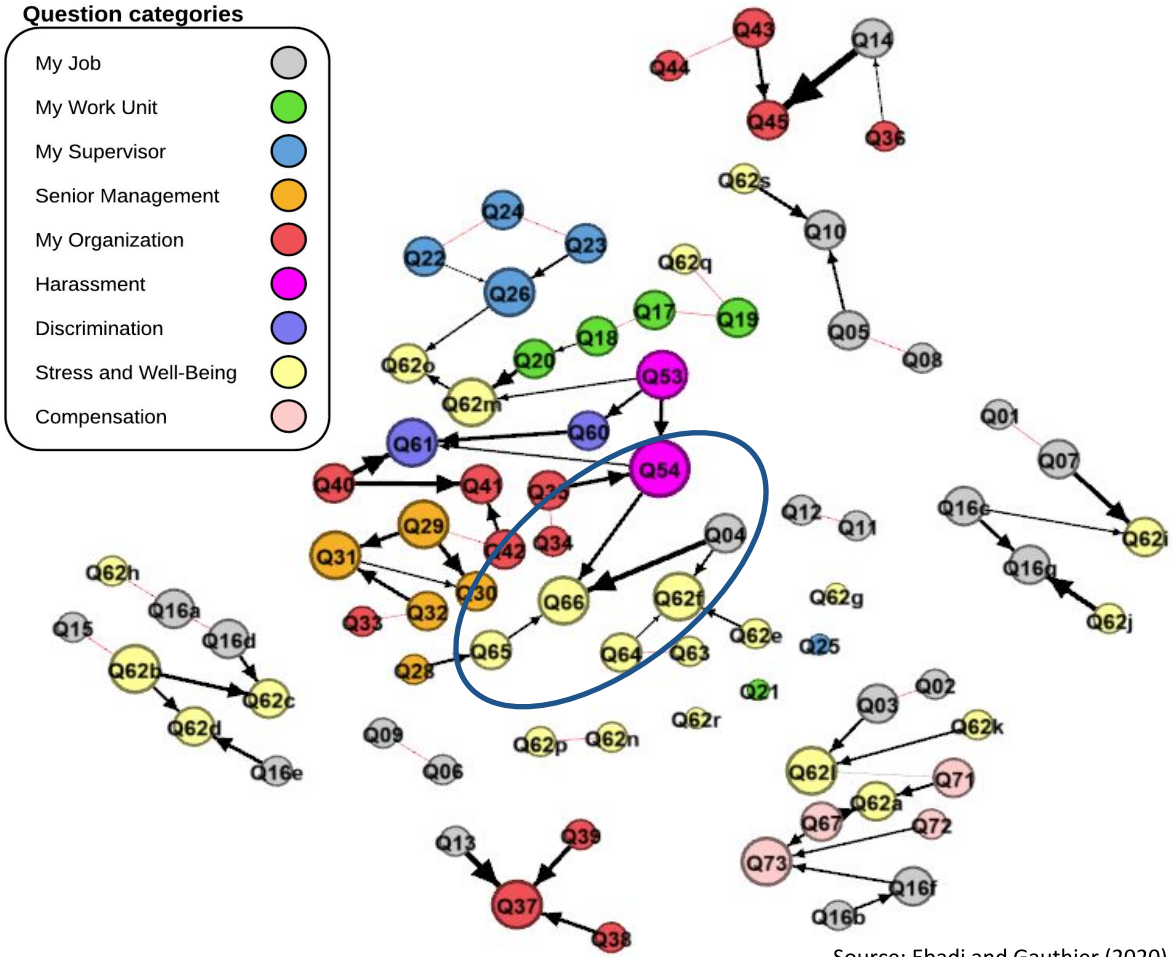
Q44 – I am satisfied with my department
Q66 – My workplace is psychologically healthy
Q26 – I am satisfied with the quality of supervision I receive



Source: Ebadi and Gauthier (2020)

Causal relationships can be inferred

We can **estimate** which responses influence other responses, and by how much, using causal graphs



EXAMPLE

Responses to

Q66 – My workplace is **psychologically healthy**

are influenced by responses to:

Q65 – My department does a good job of raising **awareness** of **mental health** in the workplace

Q04 – I have support at work to **balance** my work and personal life

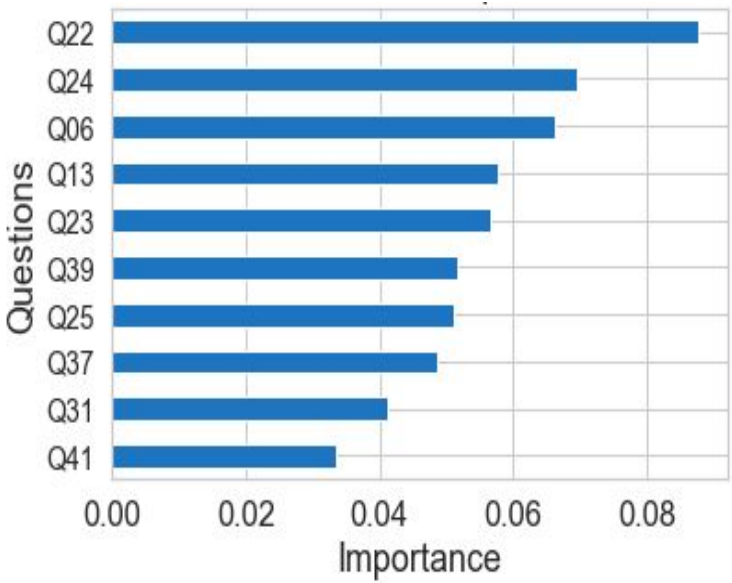
Q54 – My department works hard to create a workplace that prevents **harassment**

Source: Ebadi and Gauthier (2020)

Another example

Q26 (I am satisfied with the quality of supervision I receive)

Top 10 predictors of Q26 responses are the responses to...



Q22. I receive useful feedback from my supervisor on my job performance.
Q24. My Supervisor keeps me informed about issues affecting my work.
Q06. I receive meaningful recognition for work well done.
Q13. I have support at work to provide a high level of service.
Q23. I can count on my supervisor to keep his/her promise.
Q39. I feel I would be supported by my department if I proposed a new idea.
Q25. My supervisor seems to care about me as a person.
Q37. My department does a good job of supporting employee career development.
Q31. Senior management in my department makes effective and timely decisions.
Q41. I think my department respects individual differences.

Source: Ebadi and Gauthier (2020)

Other examples of recent projects done using R

(and associated techniques)

- Personnel enrolment & selection modelling

  (time series modelling, machine learning)
- Procurement cost forecasting

  (time series modelling, Bayesian modelling using STAN)
- Predictive maintenance - equipment failure forecasting

  (unsupervised and supervised learning)
- Supply chain forecasting

  (supervised machine learning)


Thank you!
@ygauthie