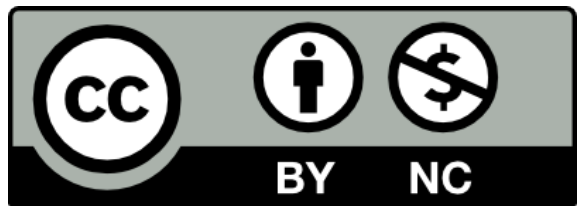
Business 4720 - Class 1

Introduction and Processes

Joerg Evermann

Faculty of Business Administration
Memorial University of Newfoundland
jevermann@mun.ca



MEMORIAL UNIVERSITY

Introduction

Terminology

- Data Analytics
- Data Science
- Business Analytics
- Machine Learning
- Artificial Intelligence
- Big Data
- Statistics

MEMORIAL UNIVERSITY

Introduction

Terminology

- Method
- Technique
- Tool

MEMORIAL UNIVERSITY

Types of Analytics

Aims and Outcomes

- *Descriptive Analytics*: Describes ”what is”. Summarizes the data, makes comparisons, identifies historical trends. Typically model-free.
- *Predictive Analytics*: Predicts ”what may be” in the future. Typically builds a model based on past data to predict future cases/events/outcomes.
- *Prescriptive Analytics*: Prescribes ”what should be done”. Builds a model that learns best actions based on past observations and actions.
- *Visual Analytics*: Uses graphs to visualize information for gaining insight. Employs human abilities to visually identify trends, make comparisons, etc.

MEMORIAL UNIVERSITY

Data and Models

Data

- In many analytics applications, 80% of time/cost/effort is spent on data management
- Data quality determines the quality of the analytics outcome

Models

- Statistical models, for example a linear regression model for prediction
- Describes the data (generating mechanism) in parameterized form, for example intercept and slope of linear model.

MEMORIAL UNIVERSITY

Learning

Supervised Learning

- Correct outcome (numerical value or target category) is given for each input/observation
- Examples: Linear regression model, generative pre-trained transformers (GPT)

Unsupervised Learning

- Unsupervised learning: No outcomes are provided
- Example: Clustering, components analysis

MEMORIAL UNIVERSITY

Analytics is not Statistics

- Both may use the same kinds of mathematical models

Explanation

- Sample or population statistics
- Model quality determined by fit between model and data
- Inferential statistics concerned with generalizing from sample to population
- Aims to identify data generating mechanism
- Not focused on individual cases/observations

Prediction

- Focus on individual cases/observations
- Model quality determined by precision/accuracy of individual predictions
- No inference to population
- Pragmatic, does not aim or claim to identify data generating mechanism

This Course

What You Will Learn:

- Introduction to Business Analytics
- Data Management (On-Premises & Cloud; SQL & NoSQL)
- Computational Foundations
- Descriptive and Visual Analytics
- Supervised and Unsupervised Machine Learning
  - Regression, Classification
  - Clustering, Component Analysis
  - Time-Series Models, incl. ARIMA and GARCH
- Predictive Analytics with Deep Neural Networks
- Prescriptive Analytics (Reinforcement Learning)
- Process Analytics (Mining and Prediction)
- ML and AI Processes
- Ethical and Legal Issues
- Management and Governance

Tools

- Open-source
- Cross-platform (Linux, Windows Mac)
- Multi-architecture (x64/amd64 ”Intel/AMD” and arm64 ”ARM/Apple”)

MEMORIAL UNIVERSITY

R

- Language and Software for Statistical Computing
- First appeared August 1993
- Open-source, multi-platform
- More than 16,000 packages available on CRAN (”Comprehensive R Archive Network”)

- Tidyverse packages (tidyr, dplyr, stringr)
- GGPlot2

`https://www.r-project.org/`

MEMORIAL UNIVERSITY

Python

- Programming Language
- First appeared February 1991
- Open-source reference implementation, multi-platform
- More than 450,000 packages available on PyPI (”Python Package Index”)

- PyCharm
- JupyterLab Desktop
- Numpy, Pandas, Plotly
- Tensorflow, Keras

`https://www.python.org/`

MEMORIAL UNIVERSITY

Virtual Machine Environment

- Virtual Machine Appliance for x64/amd64 architecture
- ”Computer within a computer”
- Includes guest operating system and all required software and data sets

---

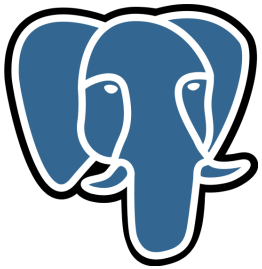
Ubuntu

- Most popular Linux distribution
- Open-source, multi-hardware
- Frequently used in software development

Username: busi4720 **Password:** busi4720

MEMORIAL UNIVERSITY

Database Management Systems



- Relational Database Management System
- Open-source, mult-hardware
- First appeared July 1996
- Active DBMS features such as triggers and stored procedures in PLSQL and other languages

---

- Graph Database (NoSQL)
- Supports property graphs
- Open-source community edition
- Cypher query language

MEMORIAL UNIVERSITY

Software Versions and Sources

R	4.1.2	www.r-project.org
dplyr	1.1.3	www.tidyverse.org
tidyr	1.3.0	www.tidyverse.org
ggplot2	3.4.4	www.tidyverse.org
Python	3.8	www.python.org
numpy	1.24.4	numpy.org
pandas	2.0.3	pandas.pydata.org
plotly	5.18.0	plotly.com
tensorflow	2.13.1	www.tensorflow.org
Postgres	16.0-1	www.postgresql.org
pgAdmin4	7.8	www.pgadmin.org
PyCharm	2023.2.3	www.jetbrains.com/pycharm/
Jupyterlab	4.0.7-1	//github.com/jupyterlab/jupyterlab-de
Neo4J	5.14.0	www.neo4j.com

MEMORIAL UNIVERSITY

Related Tools

Database Management Systems

- MongoDB (a document database)
- ArangoDB (a multimodel database)
- OrientDB (a graph database)
- Redis (a key/value database)
- Cassandra (a NoSQL database)

Infrastructure

- Hadoop (a distributed file system)
- Spark (an analytics engine on Hadoop)
- HBase (a distributed database)
- Hive (a data warehouse system)

MEMORIAL UNIVERSITY

In-Class Activity

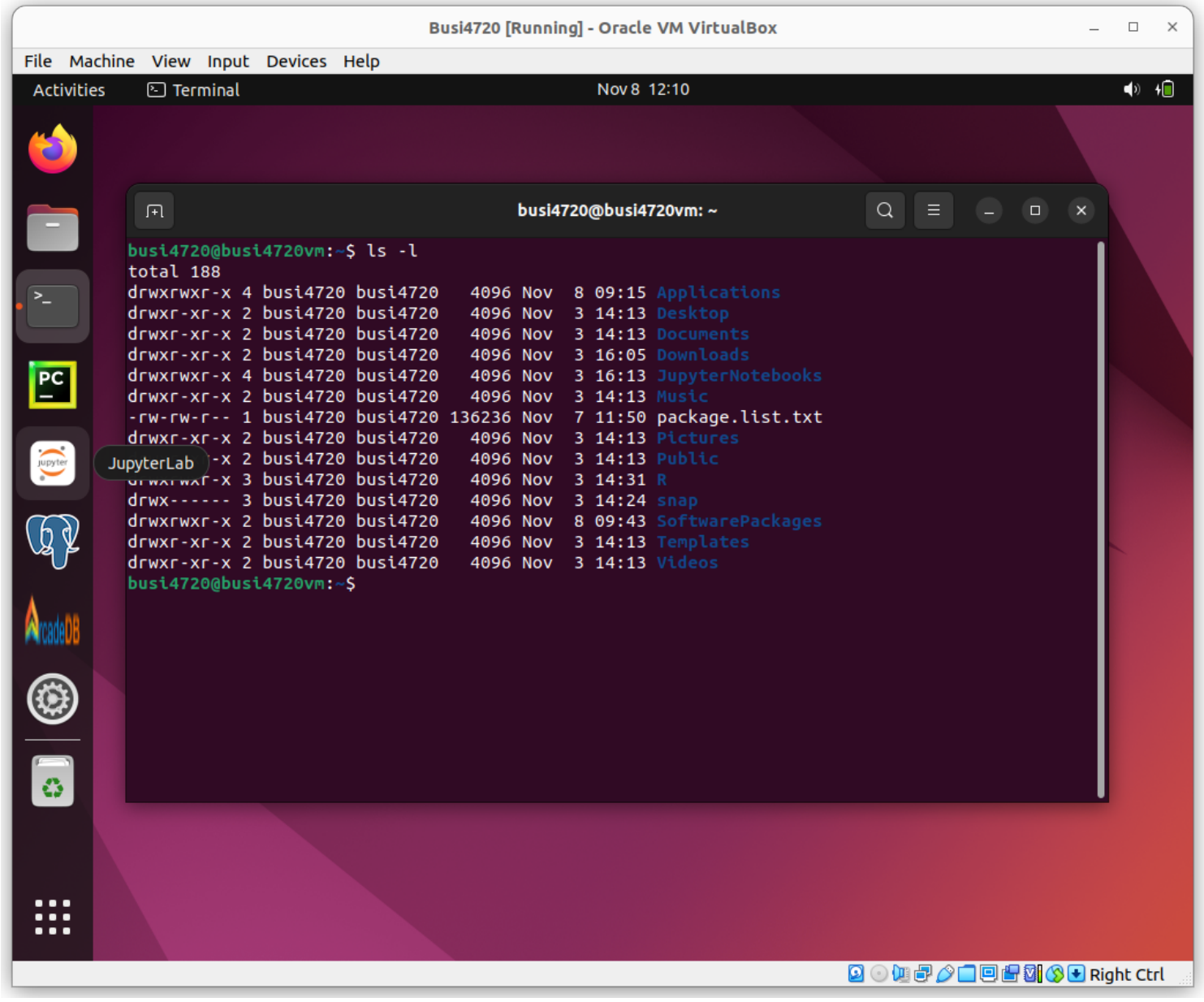
Either

1. Download and install the required software on your computer[1]:
   - R → `www.r-project.org`
   - Python → `www.python.org`
     - Ensure you install "pip", the Python package installer
   - Postgres → `www.postgresql.org`
   - PyCharm → `www.jetbrains.com/pycharm/`
   - Neo4J → `www.neo4j.com`
2. Download VirtualBox and import the virtual machine appliance from the course Brightspace site
   - VirtualBox → `www.virtualbox.org`

[1] You do not need to install the R and Python packages yet, you will do this later in the course when you need them.

MEMORIAL UNIVERSITY

Intro to Ubuntu



Ubuntu Basics

- Version of Linux created and distributed by Canonical Inc.
- Open source and multi-hardware support
- Long term support (LTS) versions released every 2 years, supported for 5 years
- Based on the Debian distribution
- User interface is called ”Gnome” or ”Gnome Shell”
- Software is installed in packages
  - Typically from the Ubuntu online repository or manually using Debian package files
  - Use the ”apt” command-line tool or the ”Synaptic” graphical tool to manage packages

MEMORIAL UNIVERSITY

Ubuntu Basics [cont'd]

- Multi-user capable
  - User files are in the directory/folder "`/home/busi4720/`"
- Users have limited permissions to files and folders (read, write, eXecute)
- Users have limited privileges
  - But some users are *administrator* users and/or have "**sudo**" privileges

MEMORIAL UNIVERSITY

The Ubuntu Command Line

In Ubuntu, you can open the Terminal using the key combination Ctrl-Alt-T, or by selecting the Terminal icon from the side bar or the application list. You can also open a terminal from the file browser.
Print the working directory:

```
busi4720@busi4720vm:~$ pwd
/home/busi4720
```

Make a folder/directory:

```
busi4720@busi4720vm:~$ mkdir someFolder
```

Change the working directory:

```
busi4720@busi4720vm:~$ cd someFolder
busi4720@busi4720vm:~/someFolder$ cd ..
busi4720@busi4720vm:~$ cd ~
```

MEMORIAL UNIVERSITY

Special directory characters

`~`	User home directory
`.`	Current directory
`..`	Upwards in the directory tree
`/`	Root of directory tree

Tips

- Autocompletion of file names with the ” tab ” key (up to ambiguity)
- Recall earlier commands with the ” up arrow ” key
- Search earlier commands with the ” Ctrl-R ” key
- Cut/copy/paste with Ctrl-Shift-X , Ctrl-Shift-C , Ctrl-Shift-V

MEMORIAL UNIVERSITY

The Ubuntu Command Line

List folder/directory contents:

```
busi4720@busi4720vm:~$ ls -l ~/Applications
total 4
drwxr-xr-x 8 busi4720 busi4720 4096 Nov  7 11:45 pycharm-community-20
```

The results show the total size in kB, and a list of entries:

- Type of entry (”d” = directory)
- Permissions for owner (”rwx”), users in the same group (”r-x”) and other users (”r-x”)
- Names of owner and groups (”busi4720”)
- Size (in bytes)
- Last modification date and time
- File or directory name

MEMORIAL UNIVERSITY

The Ubuntu Command Line

Print a string of text:

```
$ echo "To be or not to be"
To be or not to be
```

Redirect output to a file using the *redirect* symbol ”>”:

```
$ echo "To be or not to be" > someFile.txt
$ ls -l someFile.txt
-rw-rw-r-- 1 busi4720 busi4720 19 Nov  8 14:50 someFile.txt
```

Print contents of a file (”concatenate”):

```
$ cat someFile.txt
To be or not to be
```

MEMORIAL UNIVERSITY

The Ubuntu Command Line

Copy a file:

```
$ cp someFile.txt someCopy.txt
```

Move a file:

```
$ mv someCopy.txt ~/someFolder
```

Renaming is moving:

```
$ mv someFile.txt newName.txt
```

Remove (delete) a file:

```
$ rm someFolder/someFile.txt
```

MEMORIAL UNIVERSITY

The Ubuntu Command Line

Remove a directory recursively (i.e. remove all its contents first):

```
$ rm -r ~/someFolder
```

Use this very carefully!

View the command line history:

```
$ history
    1  echo "To be or not to be"
    2  echo "To be or not to be" > someFile.txt
    3  ls -l someFile.txt
    4  less someFile.txt
    5  cat someFile.txt
...
```

MEMORIAL UNIVERSITY

Remove write permissions:

```
$ chmod -w newName.txt
```

Add write permissions:

```
$ chmod +w newName.txt
```

Add execute permissions:

```
$ chmode +x newName.txt
```

The Ubuntu Command Line

Get the manual for a command:

```
$ man ls
```

Find the processes running:

```
$ ps
```

Find something in a file or input stream:

```
$ cat newName.txt | grep be
$ ls -l | grep .txt
$ history | grep .txt
```

Note: The vertical bar is called a *"pipe"*, it pipes the output of one command as input into the next one

MEMORIAL UNIVERSITY

Command Line Tutorials

- https://ubuntu.com/tutorials/command-line-for-beginners
- https://www.digitalocean.com/community/tutorials/a-linux-command-line-primer
- https://www.digitalocean.com/community/tutorial-series/getting-started-with-linux

MEMORIAL UNIVERSITY

Hands-On Exercise I

The following are a set of connected exercises to help you practice your command line skills. Do them in the order listed.

1. Navigation and Listing
   - 1.1 Open the terminal and use the `pwd` command to print the current working directory.
   - 1.2 Use `ls` to list the contents of the current directory.
   - 1.3 Create a new directory named "Exercise1" using `mkdir`.
   - 1.4 Navigate into the "Exercise1" directory using `cd`.
2. File Manipulation
   - 2.1 Create a new file named "file1.txt" inside the "Exercise1" directory using `touch`.
   - 2.2 Use `cat` to display the contents of "file1.txt".
   - 2.3 Append the text "Hello, Bash!" to "file1.txt" using `echo` and `»`.
   - 2.4 Display the updated contents of "file1.txt" using `cat`.
3. Removing and Renaming
   - 3.1 Remove "file1.txt" using the `rm` command.

MEMORIAL UNIVERSITY

Hands-On Exercise II

3.2 Create a copy of the "Exercise1" directory named "Exercise1_backup" using `cp -r`.
3.3 Remove the original "Exercise1" directory using `rm -r`.

4 Directory Manipulation
4.1 Recreate the "Exercise1" directory.
4.2 Create three subdirectories inside "Exercise1" named "Subdir1", "Subdir2", and "Subdir3" using `mkdir`.
4.3 List the contents of "Exercise1" to verify the creation of subdirectories.

5 Searching and Filtering
5.1 Create a file named "keywords.txt" inside "Exercise1" and add some random text.
5.2 Use `grep` to search for a specific word (e.g., "Bash") in "keywords.txt".
5.3 Create a new file named "filtered.txt" and use `grep` to filter lines containing the word you searched for in "keywords.txt".

6 Process Management

MEMORIAL UNIVERSITY

Hands-On Exercise III

6.1 Use `ps` to display information about the current processes running on your system.

6.2 Use `ps aux | grep bash` to filter and display information about Bash processes.

7 Cleanup

7.1 Remove the entire "Exercise1" directory and its contents using `rm -r`.

7.2 Confirm that the "Exercise1" directory no longer exists by listing the contents of the current directory.

MEMORIAL UNIVERSITY