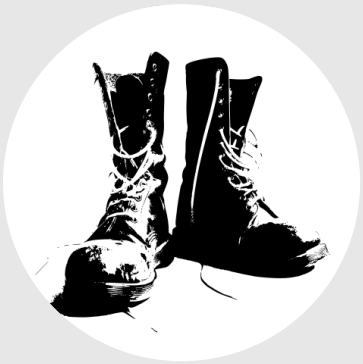


Lab 6/CIS*2250

Project Team Kickoff

courselink.uoguelph.ca

A. Hamilton-Wright &
K. Raymond

Overview

Learning objectives: ○ Extending our pair programming skills to the project team ◑ Dealing with large data sets ● Extracting and examining fields

Skills

coordination + communication (3/6)

organization + planning (3/6)

teamwork (3/6)

programming + tools (5/6)

strategy (3/6)

visualization (0/6)

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Main Focus).]

Image description

A pair of black books. Image source publicdomainvectors.org CC0 1.0

Team/Pair Organization

In this lab, we will work in our project teams. Get organized with your team members, and determine how best to perform the lab work.

The focus of this lab is answering the question "how does average employment income change over the last years for people with a computer science degree, when examined two and five years after graduation?"

We can answer this question (and more!) based on the data in `educationON.csv`.

This data is a subset of a larger data set available from Statistics Canada and is available online at the website:
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710015701

This larger data set is entitled "Characteristics and median employment income of longitudinal cohorts of postsecondary graduates two and five years after graduation, by educational qualification and field of study (STEM and BHASE (non-STEM) groupings), 2010 to 2012 cohorts." and allows answers to questions regarding education and employment.

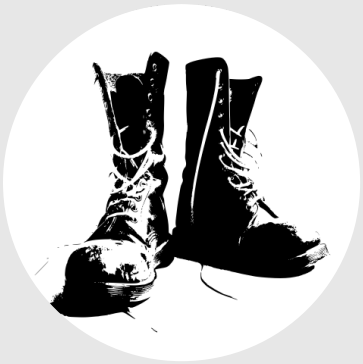
The `educationON.csv` file is a subset of the larger file, and contains information regarding

- earnings reported (in 2017 constant dollars) for
- Canadian and international students
- who are residents of Ontario,
- who are university graduates reporting employment income
- and have a degree in computer or information sciences.

When you look at the data file you will see that it has columns named:

- "REF_DATE",
- "GEO",
- "DGUID",
- "Educational qualification",
- "Field of study",
- "Gender",
- "Age group",
- "Status of student in Canada",
- "Characteristics after graduation",
- "Graduate statistics",
- "UOM",
- "UOM_ID",
- "SCALAR_FACTOR",
- "SCALAR_ID",
- "VECTOR",
- "COORDINATE",
- "VALUE",
- "STATUS",
- "SYMBOL",
- "TERMINATED",
- "DECIMALS"

What do these columns contain? How many of them can you recognize?



Lab 6/CIS*2250

Project Team Kickoff

courselink.uoguelph.ca

A. Hamilton-Wright &
K. Raymond

Overview

Learning objectives: ○ Extending our pair programming skills to the project team ◑ Dealing with large data sets ● Extracting and examining fields

Skills

coordination + communication (3/6)

organization + planning (3/6)

teamwork (3/6)

programming + tools (5/6)

strategy (3/6)

visualization (0/6)

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Main Focus).]

Image description

A pair of black books. Image source publicdomainvectors.org CC0 1.0

Task 1 Description: Select Data To Answer The Question

Our obejctive is to produce a file that has three columns that describe

1. the year of the data
2. how many years after graduation the graduate is reporting (2 or 5)
3. the income (you will find this in a column called "`VALUE`"

Your file should describe earnings for all people holding an undergraduate degree. The `educationON.csv` file contains all of this information, and more (for instance, there is information for people with graduate degrees as well).

Your task is to determine how to obtain the correct data by writing and running a perl program.

Using any of your past perl programs as a starting point, write a program that will extract the data to answer this question.

Put your program in a file called `undergradCSincome.pl`. Your program should take a single argument: the name of the data file.

Print out the results to the screen in the format shown here below. If everything has worked correctly, you should see exactly this output:

```
Year,YearsAfterGraduation,Dollars
2010,"Median employment income two years after graduation",55600
2010,"Median employment income five years after graduation",72800
2011,"Median employment income two years after graduation",58800
2011,"Median employment income five years after graduation",72800
2012,"Median employment income two years after graduation",58900
2012,"Median employment income five years after graduation",75800
```

A file containing this data is available on CourseLink for your reference.

Task 2 Description: Produce a plot based on your data

Collect the output of your program into a data file, as we did in the last lab:

```
$ perl undergradCSincome.pl educationON.csv > income.csv
```

Write a program to plot your data as a line graph, based on the `createNameRankPlot.pl` file we used in Lab 4. Call your new program `createUndergradCSincomePlot.pl`.

Running the following command should produce the plot on the next page:

```
$ perl createUndergradCSincomePlot.pl income.csv income.pdf
```

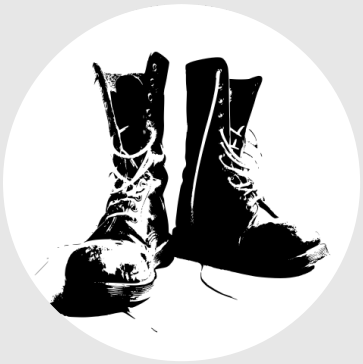
Upload to Courselink

Upload both of your new perl programs to CourseLink:

- `undergradCSincome.pl`
- `createUndergradCSincomePlot.pl`

Project Kickoff

Once you have completed these tasks, be sure everyone in your group has read the "Project Overview" document in CourseLink, and come up with a strategy to reach the first milestone described in this document for the next lab.



Lab 6/CIS*2250

Project Team Kickoff

courselink.uoguelph.ca

A. Hamilton-Wright &
K. Raymond

Overview

Learning objectives: ○ Extending our pair programming skills to the project team ◑ Dealing with large data sets ● Extracting and examining fields

Skills

coordination + communication (3/6)

organization + planning (3/6)

teamwork (3/6)

programming + tools (5/6)

strategy (3/6)

visualization (0/6)

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Main Focus).]

Image description

A pair of black books. Image source publicdomainvectors.org CC0 1.0