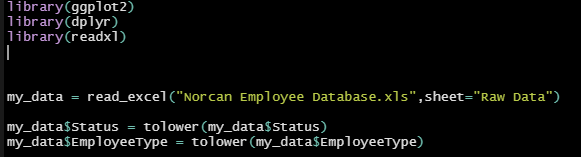
Case Study 3

Elin Huliiev

For this assignment I will use r studio to produce all the visualizations. Besides, r provides us with the great package to do dashboards, so let’s begin!

Firstly, we will analyze Norcan Employee Database data set.

Let’s observe the distribution of Occupation in general.



Loading data set and doing some formatting (converting to lower case)

Now, let’s count frequency for each Occupation.



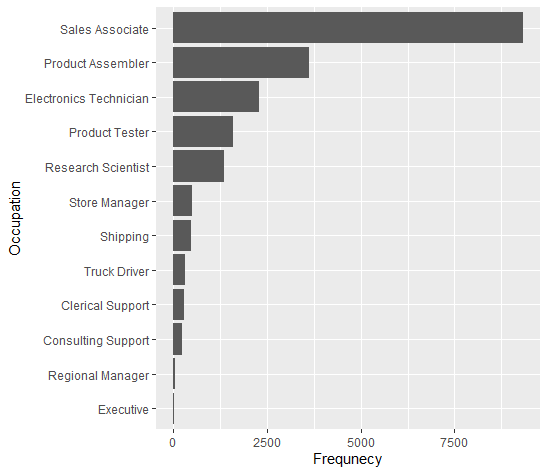
Output:



Let’s plot this out:

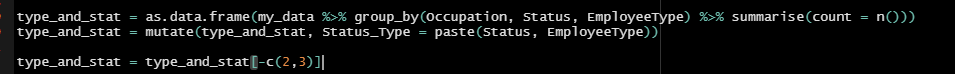


Output



Okay, so it seems that Sales Associate is the most popular job in our dataset. The least one is Executive, which is obvious, because there cannot be many executive managers.

Now, let’s group our data by employee status and type.

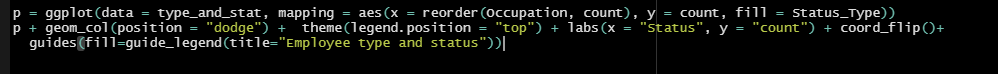


Output:

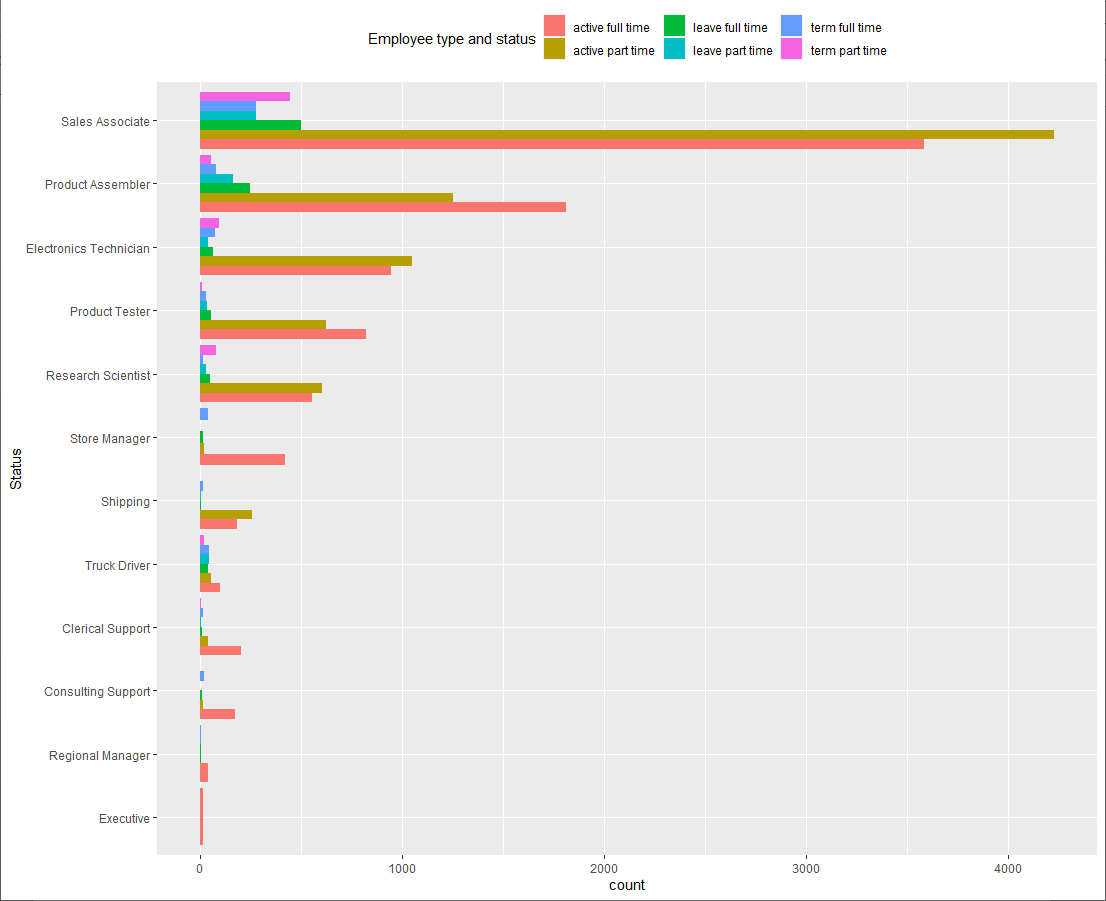


Note, that I kind of “merged” together status and type. This was done so that we could easily plot this.

Like this:



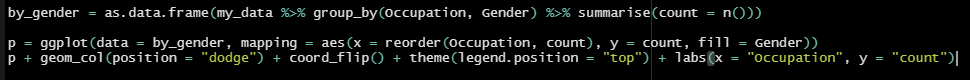
Result:



So, for our most frequent job – Sales Associate, we have the following: most of the employees within this category are active status and they are working part time. Also, there are lots of employees who work full time and are active, for this category. So, this is pretty “busy” category – a lot of people working as full/part time employees. Also, shipping category is interesting, because we do not have any employees who left this Occupation. Well, there is someone with “leave part time” status, but overall almost none of our employees has left this position, in regards to this data set.

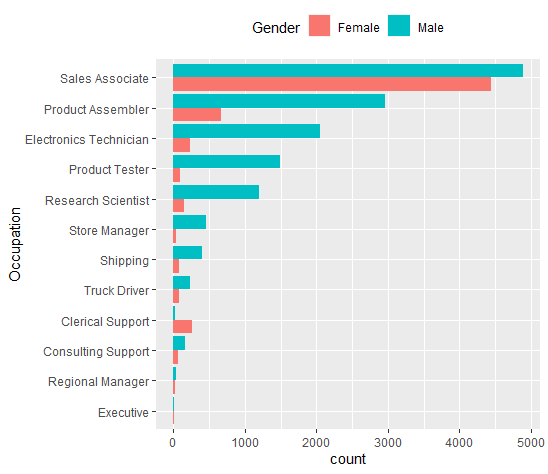
Moving on, let’s see Occupation by gender.

So,



Grouping and plotting.

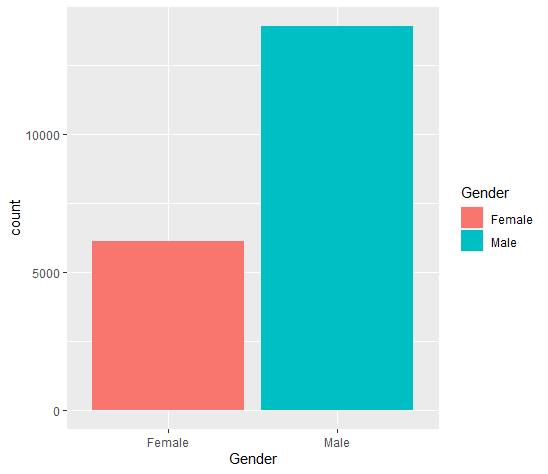
Result:



Males are dominant in almost all Occupation. However, there are two thing that might draw our attention. Firstly, for Sales Associate, amount of males (4800) and females (3900) is more like equal, at least in comparison to the other occupations. Meanwhile, the only category where females are dominant is Clerical Support.

But let’s see, if this distribution is biased or not.

Observe the distribution of genders:

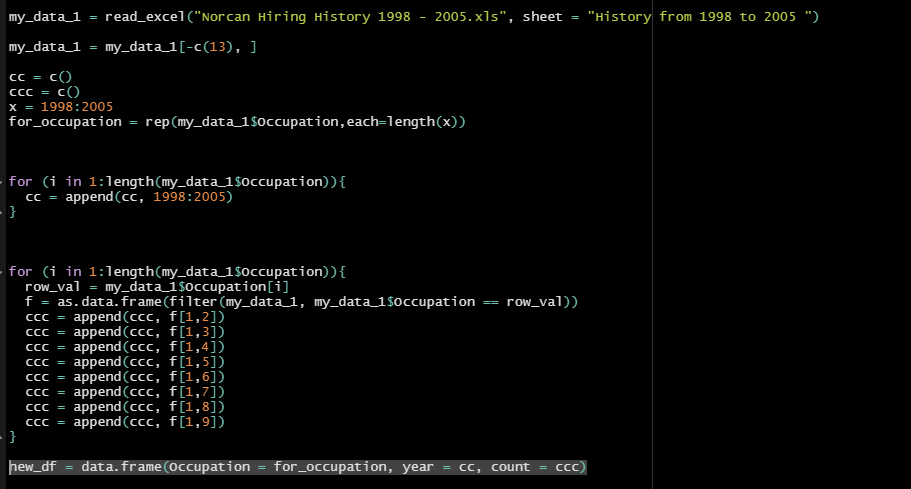


Indeed, we have much more males than females. This may lead to the wrong estimation of our distributions, like: “it seems that males are doing better in Store Manger Occupation, that’s why we have more males for that occupation”. It might be wrong, because we have over 15k of males and 6-7k of females.

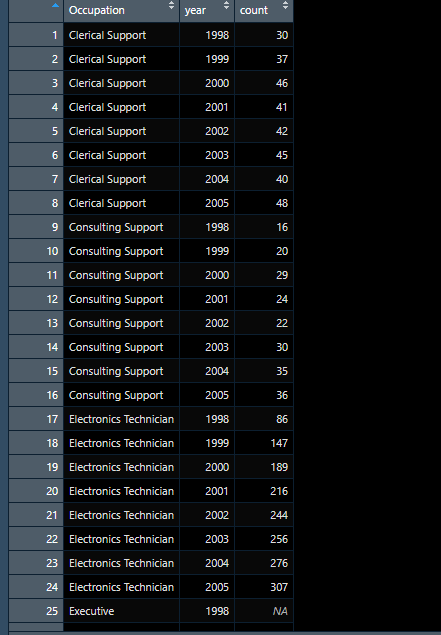
Let’s move on to the Norcan Hiring History 1998 – 2005 data set.

Firstly, let’s see occupation history for each year.

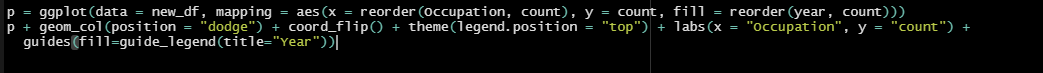
We will have to create another data frame to do this.



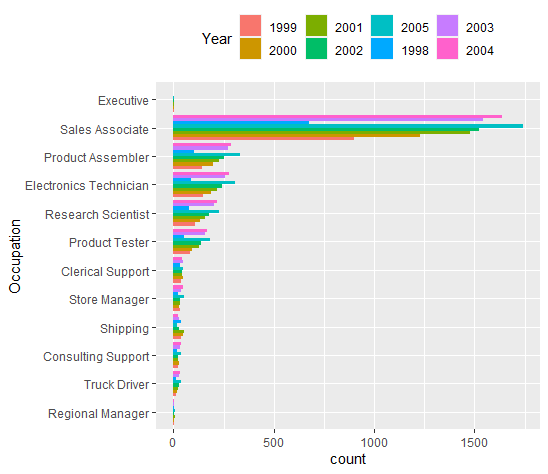
Result:



Fine, now let’s plot.



And the result:



Okay, so for Sales Associates most of our employees were hired in 2005 and 2004. The least amount was hired in 1998. Overall, we see the trend there: most employees were hired within 2002-2005 and small amount was hired in 1998-2001.

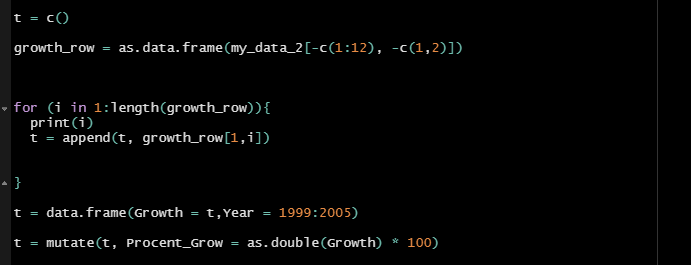
Now, what about the annual growth here? Well, It was already calculated for us here.



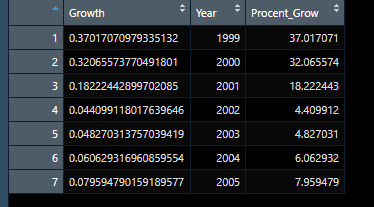
Basically, grand total field was just the sum of hired employees from all occupations. And then we calculated the difference in percent’s of these sums between current and previous years, except from 1998 which was the first.

We see that it has been decreasing from 1998. We can plot the trend as line.

Another new data frame for this.



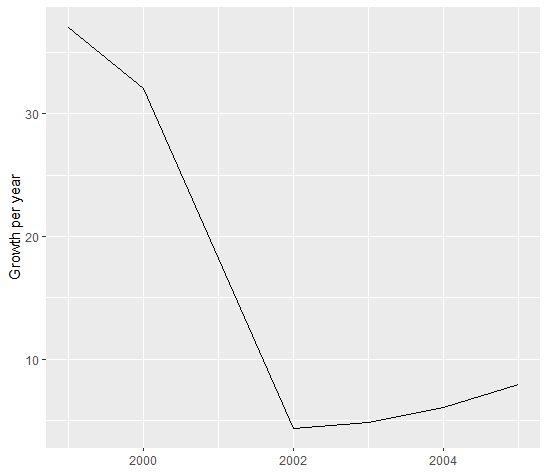
Result:



Plot:



Result:

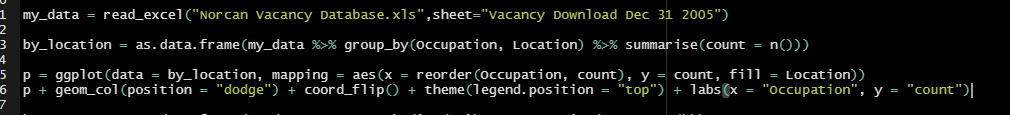


The trend is all about decreasing. This is because from 1998 to 2001 amount of hired was growing fast, however, up to 2005 it became – around 2700 on average.

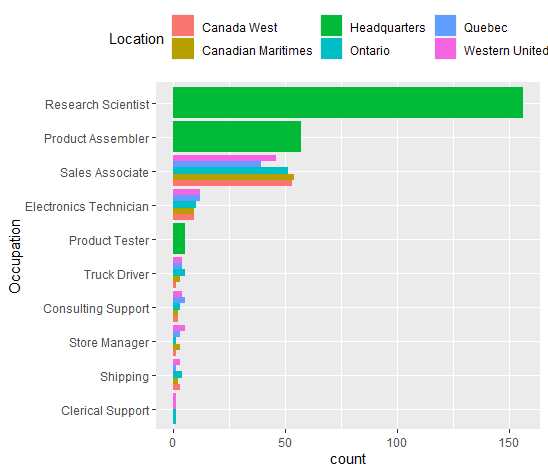
Finally, Norcan Vacancy Database data set.

We will have to group our data by location to see where most vacancies are located at.

So,

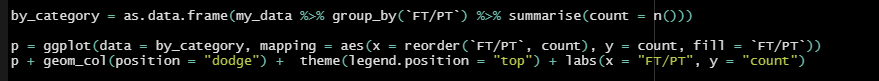


And the result:

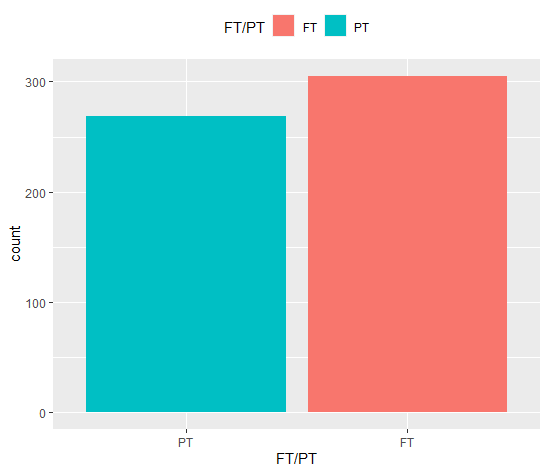


So, we see that we have Sales Associates spreaded all over the places! On the other hand, positons like Research Scientists, Product Assembler and Product Tester are located only at Headquarters location. All other vacancies are located (offered) in different places.

Finally, let’s plot the vacancies only this time grouped by FT/PT column.



And the plot:



Vacancies with FT/PT equal to FT are dominant here.

Recommendations:

Firstly, it would be nice to hire more females, otherwise it would be hard to draw some conclusion while grouping by genders. Also, there should be a reason, why hiring process stopped being so active in 2002, probably lack of spaces for vacancies. Meanwhile, we could offer some vacancies (like Research Scientist) at more locations to be more flexible.

**Thank you!**