

Labor Market Analysis part 1

y_bib

Labor Market Canada

Labor market survey results provided by Statistics Canada contain information on employment status, type of employment, gender, marital status, and whether children are present at home. For employed individuals, data on wages, hours worked, overtime, the type and field of occupation are included and etc. For part-time workers, the survey additionally explores the reasons for part-time employment. For unemployed individuals, the survey provides insights into the reasons for not working and related factors.

Goals of the Analysis:

- For employed people: provide a description and test the statistical significance of wage differences based on gender, immigrant status, age, and other relevant variables.
- Study which factors are most crucial in determining wage levels: education, place of living, occupation type and field, gender, age and etc.
- Examine differences between women and men, immigrants and non-immigrants in terms of overtime work, part-time employment, and unemployment rates.

Summary of results

Descriptive Statistics

Wages depending on gender

Mean and median of hourly rate vs gender

Male	Female
37.87467	33.13101

Male	Female
37.87467	33.13101

Mean and median wages vs gender

Male	Female
14755.46	11590.49

Male	Female
13078.5	10363.5

Both mean and median show that male workers earn more compared to female workers.

So let us see if female workers actually work less hours compared to male workers.

Means for extra hours depending on gender

Mean of extra hours by gender

gender	Mean of extra hours
Male	16.59711
Female	10.63694

Full-time and part-time workers for each gender

Full-time and part-time workers by gender

gender	Full-time	Part-time
Male	29617	4762
Female	24473	8152

Means of extra hours depending on gender

Mean of extra hours by gender

gender	Mean of extra hours
Male	16.59711
Female	10.63694

Reasons for part-time work by gender

Full-time and part-time workers by gender

gender	Full-time	Part-time
Male	29617	4762
Female	24473	8152

Reasons for part-time work by gender

Male	Female	Reason
344	519	Other reasons
262	416	Own illness or disability
98	905	Caring for children
146	565	Other personal or family responsibilities
1664	2349	Going to school
1370	2223	Personal preference
363	506	Business cond. or could not find full-time work, looked for it in last month
515	669	Business cond. or could not find full-time work, did not look for it in last month

Unemployment

Unemployment vs gender (non-weighted and weighted)

Quantities of male and female workers

Status	Male	Female
Employed, at work	31654	28718
Employed, absent from work	2725	3907
Unemployed	2842	2022
Not in labour force	19215	24507

Nonweighted unemployment rates by gender

Status	Male	Female
Employed, at work	0.56088312	0.48547858
Employed, absent from work	0.04828478	0.06604794
Unemployed	0.05035793	0.03418197
Not in labour force	0.34047417	0.41429151

Nonweighted unemployment rates by gender

Status	Male	Female
Employed, at work	0.85043389	0.82887407
Employed, absent from work	0.07321136	0.11276590
Unemployed	0.07635475	0.05836003

Weighted unemployment rate by gender

gender	Male	Female
est_est	0.07850245	0.06249231
est_var	3.526834e-06	3.226326e-06
est_sd	0.001877987	0.001796198
est_cv	2.392265	2.874270
est_lb	0.07482167	0.05897183
est_ub	0.08218324	0.06601279
est_lbq	0.07488913	0.05902658
est_ubq	0.08214689	0.06612621
est_lbq2	0.07485667	0.05899389
est_ubq2	0.08218518	0.06613314

Unemployment vs immigrant status (non-weighted and weighted)

Quantities of immigrant and non-immigrant workers

status	Immig,not more 10y	Immig,more 10y	Non-immigrant
Employed, at work	5118	9339	45915
Employed, absent from work	489	968	5175
Unemployed	583	668	3613
Not in labour force	1836	7272	34614

Nonweighted unemployment rates by immigrant status

status	Immig,not more 10y	Immig,more 10y	Non-immigrant
Employed, at work	0.63767755	0.51181016	0.51406787
Employed, absent from work	0.06092699	0.05304982	0.05793970
Unemployed	0.07263892	0.03660876	0.04045143
Not in labour force	0.22875654	0.39853127	0.38754101

Unemployment Rate by Immigrant Status

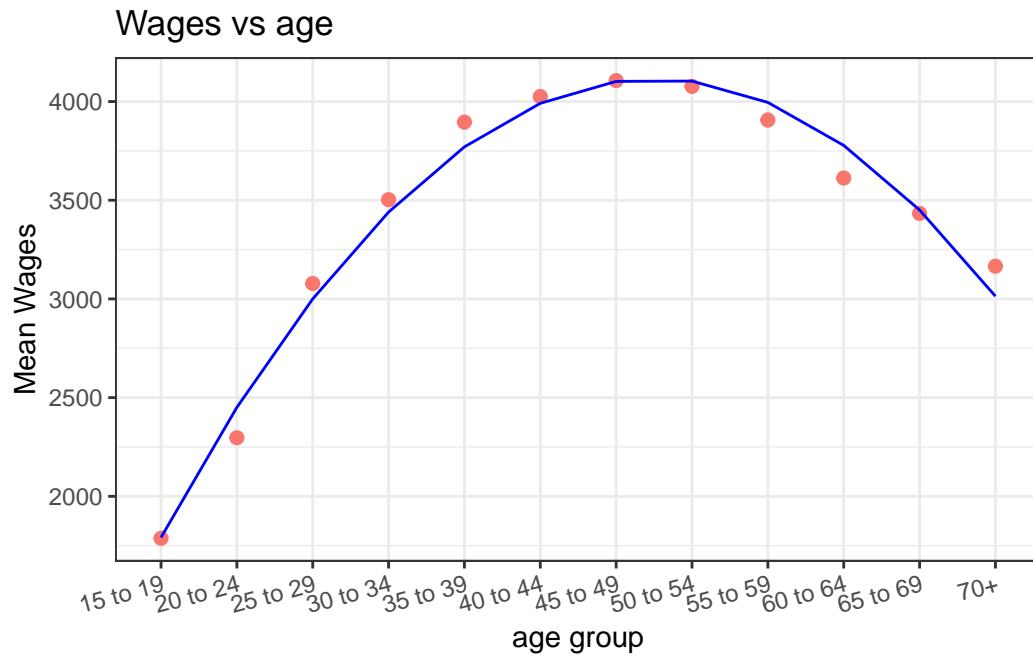
immig	Immig,not more 10y	Immig,more 10y	Non-immigrant
est_est	0.09772392	0.06771431	0.06794667
est_var	2.812005e-05	1.070288e-05	2.206648e-06
est_sd	0.005302834	0.003271526	0.001485479
est_cv	5.426342	4.831366	2.186243
est_lb	0.08733055	0.06130223	0.06503518
est_ub	0.10811728	0.07412638	0.07085815
est_lbq	0.08735608	0.06126026	0.06502754
est_ubq	0.10750023	0.07381777	0.07089137
est_lbq2	0.08729490	0.06122019	0.06502684
est_ubq2	0.10756477	0.07384755	0.07089470

Wages and Demographics

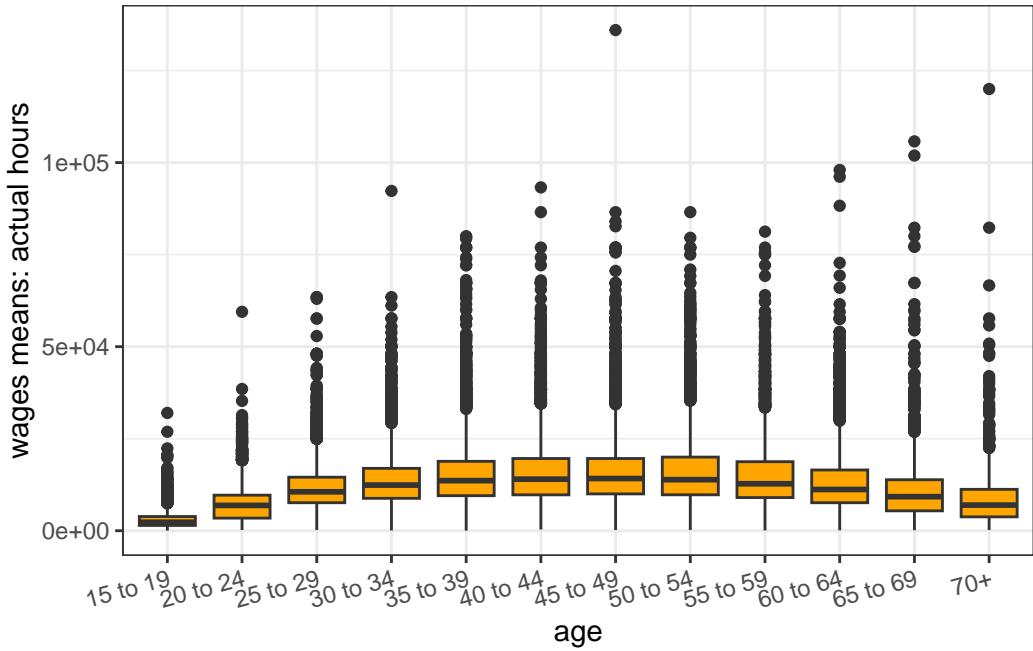
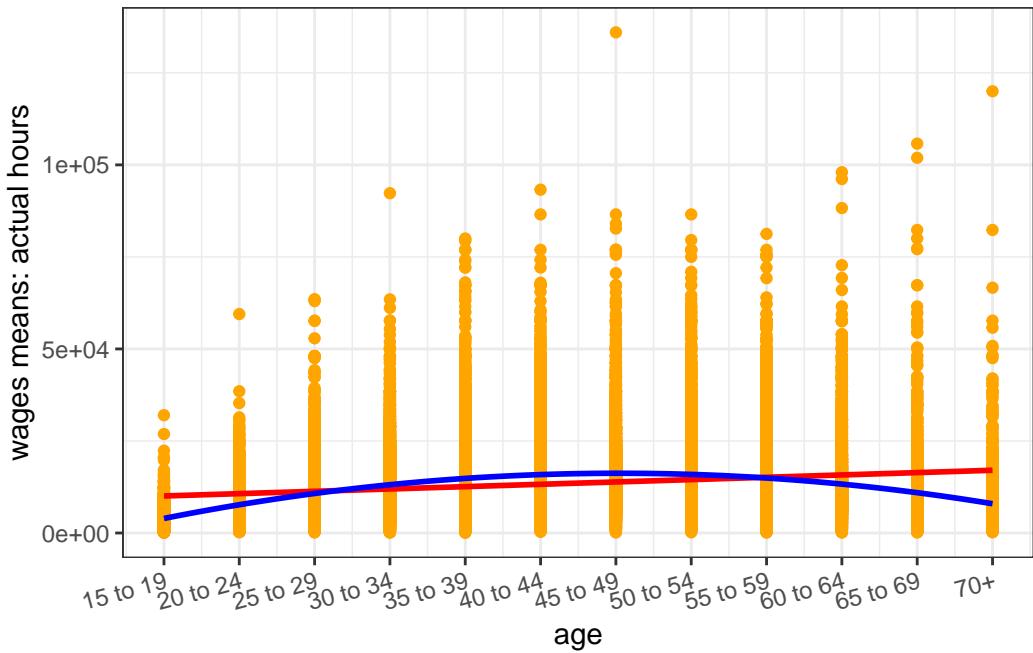
Family matter affects the hours and wages of female workers. Let us see how wages depend on age and gender.

Wages vs age (using linear regression and poly approximation)

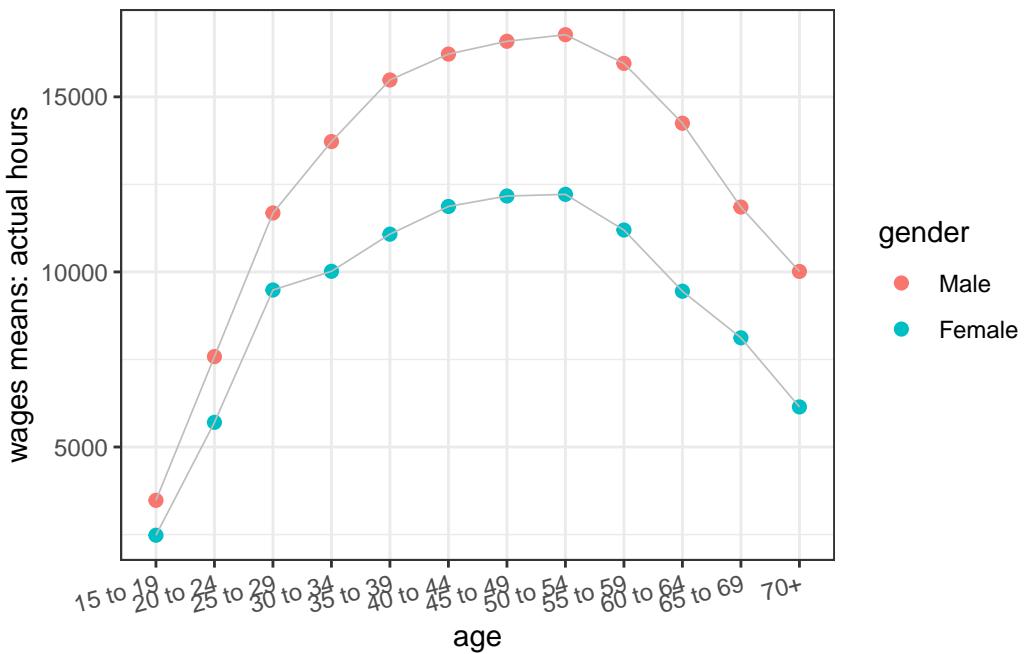
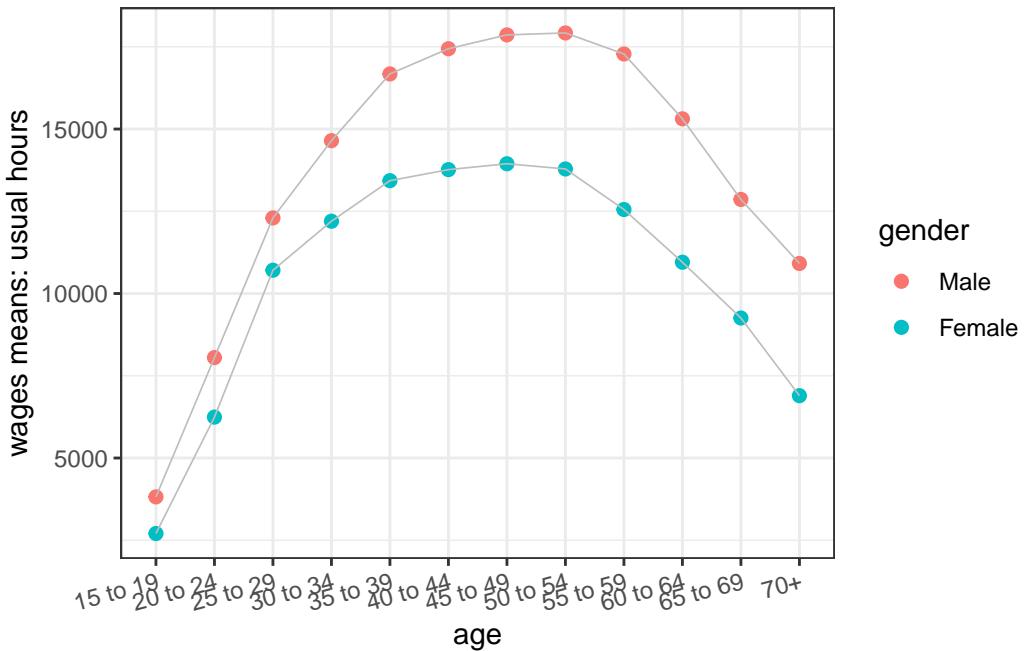
First we check how wages depend on age of workers of both genders.



Wages vs age (using linear and quadratic polynomial regressions)



Wages vs age for males and females



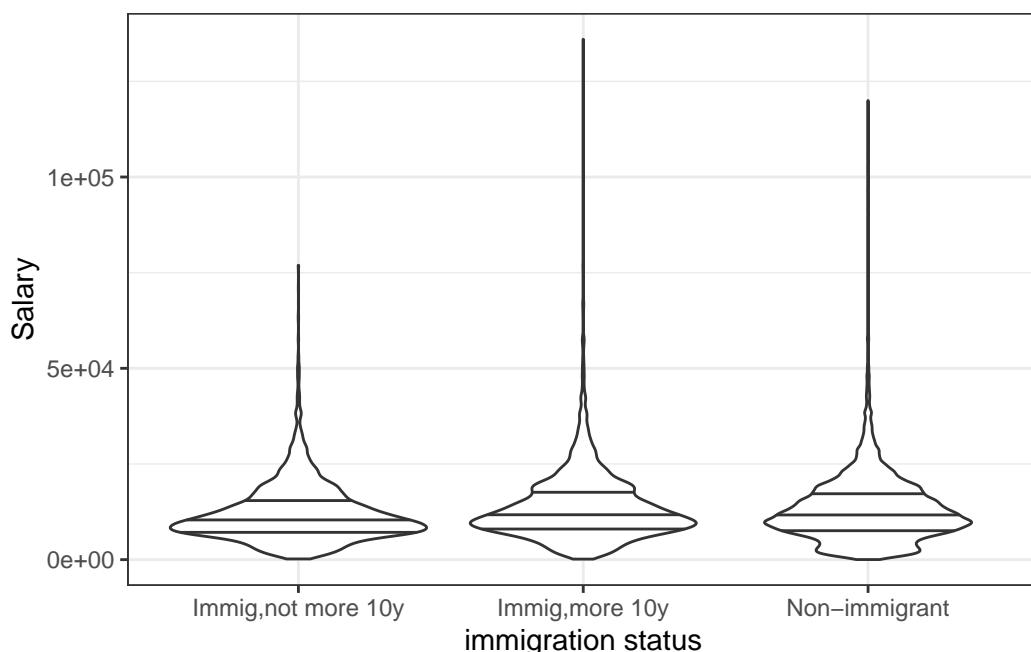
Wages depend on immigrant status

Immig,not more 10y	Immig,more 10y	Non-immigrant
12091.46	13707.10	13189.98

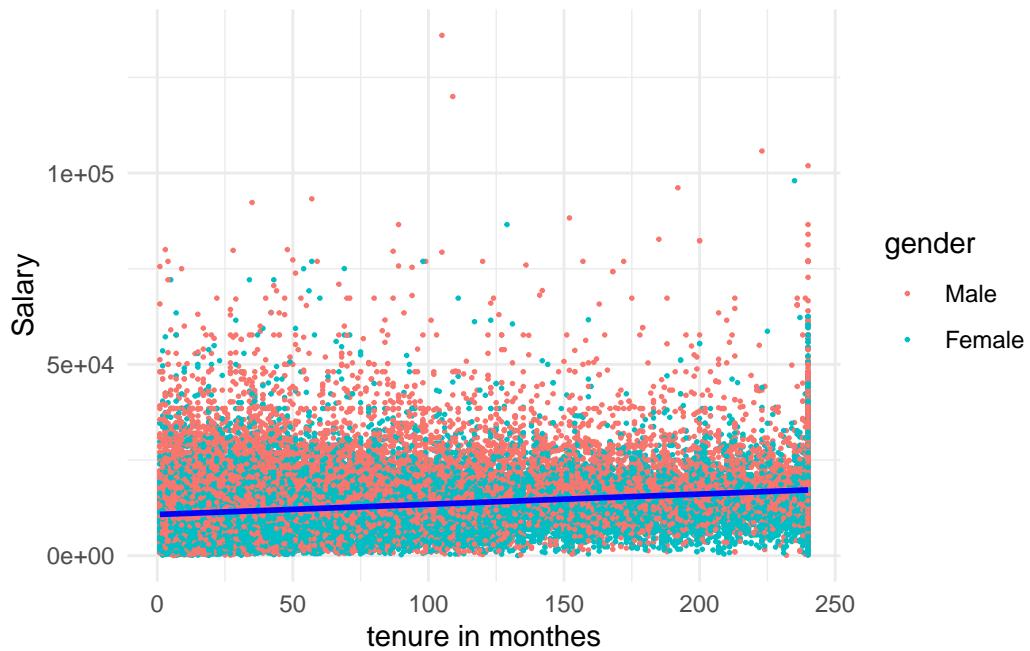
\$`Immig,not more 10y`
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
180 7200 10350 12091 15384 76924 2971

\$`Immig,more 10y`
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
134.6 8076.0 11800.0 13707.1 17679.0 136000.0 9934

\$`Non-immigrant`
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
16.1 7692.0 11745.0 13190.0 17308.0 120003.0 44755



Wages vs tenure (using linear regression)



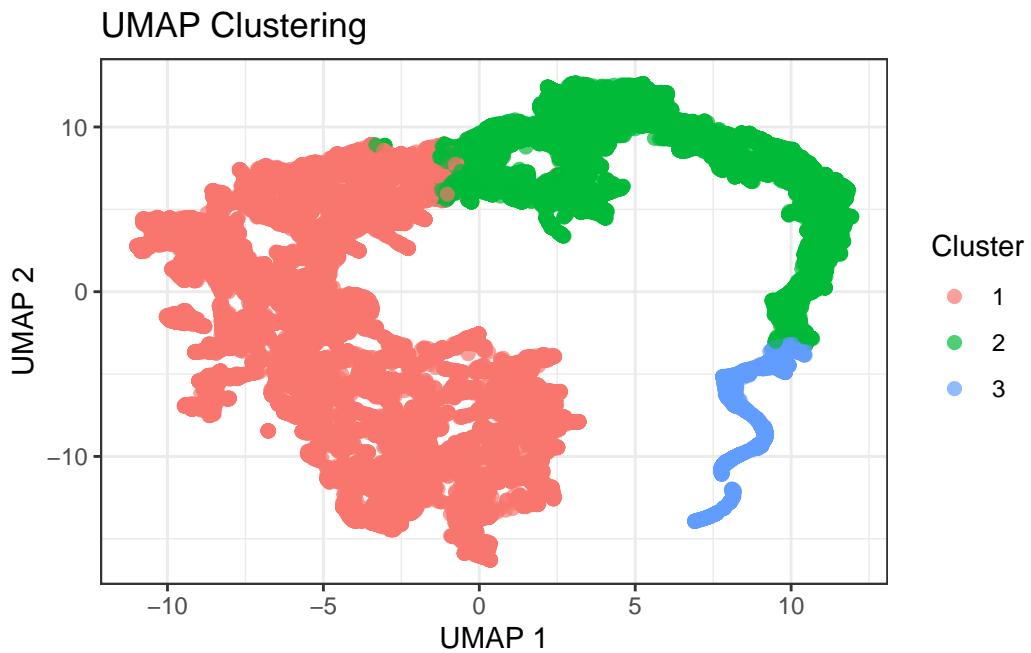
Model-based analysis: What factors are crucial for wages

Wages clustering : using UMAP and K-means

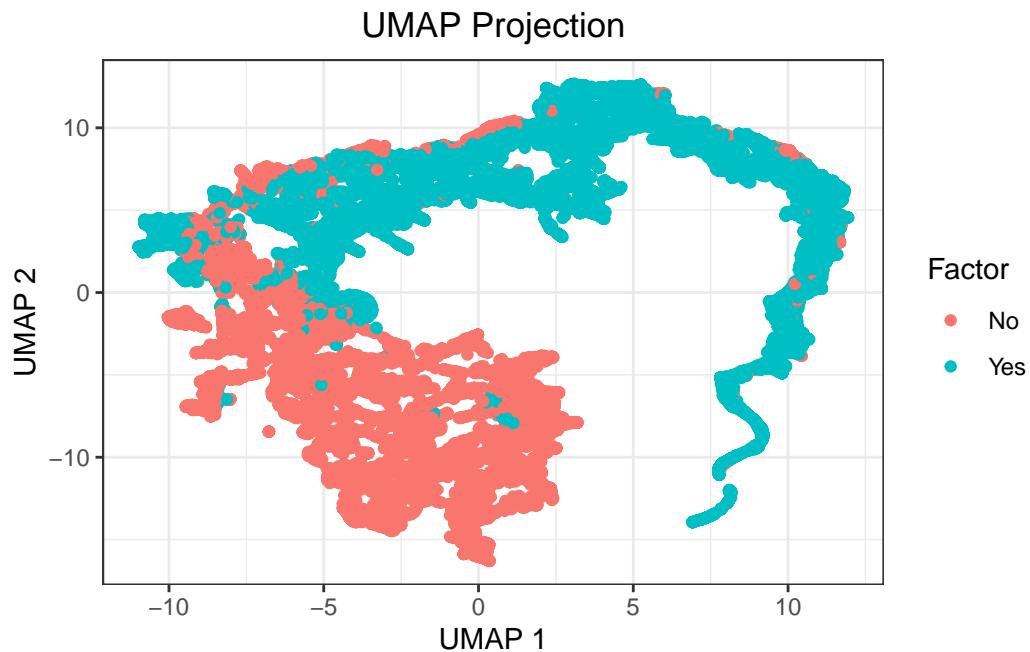
Table 1: Correlation of UMAP dimensions with other variables

	UMAP1	UMAP2
hrlyearn	0.69	0.28
paidot	0.04	0.09
unpaidot	0.17	0.03
xtrahrs	0.14	0.09
noc_43	-0.25	-0.24
noc_10	-0.21	-0.19
naics_21	0.02	0.00
utothrs	0.15	0.36
atothrs	0.18	0.32
hrsaway	0.00	0.06
tenure	0.20	0.25
age_12	0.11	0.19

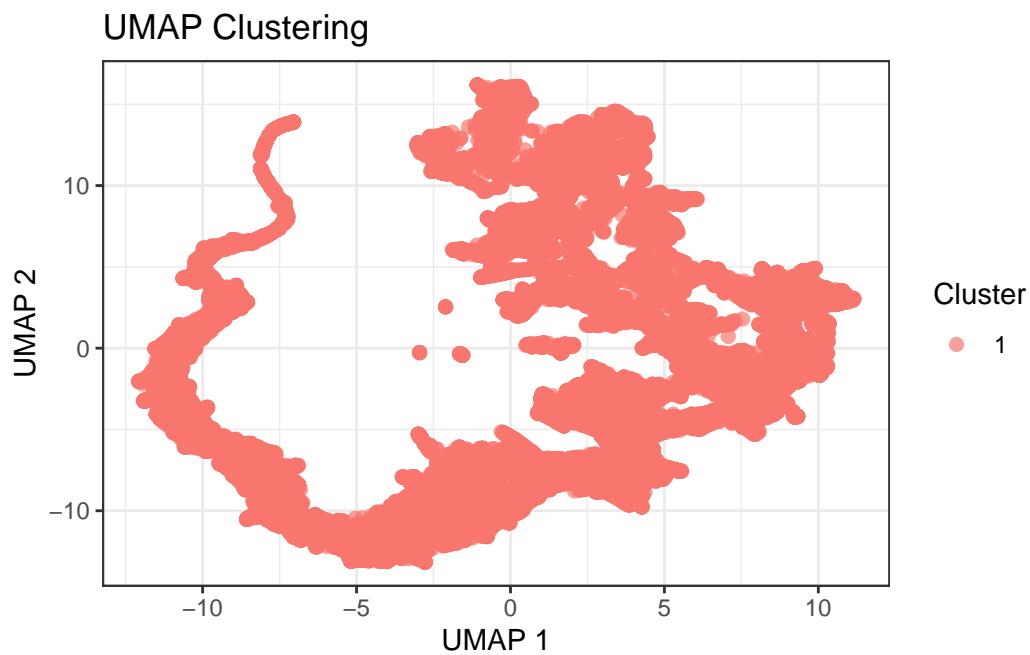
	UMAP1	UMAP2



```
[1] "Factor: wage based on utothrs > 10500 "
```



Clustering using DBSCAN



Linear regression

Wages are determined by the hourly rate (**hrlyearn**) and the typical number of hours worked (or the actual total hours - fields **utothrs**, **atothrs paidot**, **unpaidot**). Both hourly rate and hours worked exhibit a higher correlation with the UMAP embedding coordinates, suggesting they contribute significantly to the underlying variance structure of the dataset.

In the next step we check what factors define the hourly rate (**hrlyearn**) and usual hours (**utothrs**).

Model on the hourly rate

Call:

```
lm(formula = hrlyearn ~ utothrs + gender + educ + cowmain + agyownk +
    age_12 + noc_10 + prov, data = complete_labor)
```

Residuals:

Min	1Q	Median	3Q	Max
-5380.1	-1006.9	-217.9	712.2	16933.3

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4913.3524	168.5842	29.145	< 2e-16 ***
utothrs	0.6001	0.1340	4.477	7.60e-06 ***
genderFemale	-794.0552	28.1053	-28.253	< 2e-16 ***
educ1	132.2180	148.1653	0.892	0.372205
educ2	296.4799	134.9756	2.197	0.028064 *
educ3	393.2543	147.1816	2.672	0.007548 **
educ4	587.2434	132.5090	4.432	9.39e-06 ***
educ5	1212.0601	133.8732	9.054	< 2e-16 ***
educ6	1575.5548	135.5485	11.624	< 2e-16 ***
cowmain2	-371.1191	28.0222	-13.244	< 2e-16 ***
agyownk2	14.8653	32.0512	0.464	0.642798
agyownk3	100.5919	41.9116	2.400	0.016399 *
agyownk4	8.8160	50.2448	0.175	0.860718
age_12	80.2107	10.4798	7.654	2.03e-14 ***
noc_102	-2015.7933	45.7771	-44.035	< 2e-16 ***
noc_103	-1114.5310	51.4464	-21.664	< 2e-16 ***
noc_104	-1722.7262	52.5422	-32.787	< 2e-16 ***
noc_105	-1741.2820	47.6808	-36.520	< 2e-16 ***
noc_106	-2202.0038	111.6999	-19.714	< 2e-16 ***

```

noc_107      -2822.9601    48.3166 -58.426 < 2e-16 ***
noc_108      -2229.5923    49.1567 -45.357 < 2e-16 ***
noc_109      -1846.3864    100.2324 -18.421 < 2e-16 ***
noc_1010     -2471.1826    68.2268 -36.220 < 2e-16 ***
prov11       -717.1945    100.0314 -7.170 7.76e-13 ***
prov12       -400.3133    77.4120 -5.171 2.35e-07 ***
prov13       -467.9615    73.1458 -6.398 1.61e-10 ***
prov24       -33.3800     58.4122 -0.571 0.567697
prov35       201.5857     56.3541  3.577 0.000348 ***
prov46       -431.9166    67.2236 -6.425 1.35e-10 ***
prov47       -184.3017    70.4273 -2.617 0.008879 **
prov48       342.6239     65.7848  5.208 1.92e-07 ***
prov59       190.9888    62.8390  3.039 0.002374 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 1679 on 21470 degrees of freedom
(3718 observations deleted due to missingness)
Multiple R-squared: 0.3258, Adjusted R-squared: 0.3248
F-statistic: 334.6 on 31 and 21470 DF, p-value: < 2.2e-16

Model on usual total working hours

```

Call:
lm(formula = atothrs ~ gender + educ + cowmain + agyownk + naics_21 +
prov, data = complete_labor)

Residuals:
    Min      1Q  Median      3Q      Max 
-529.52  -60.91   27.37   88.09  759.11 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 438.109    15.681  27.940 < 2e-16 ***
genderFemale -69.644     2.277 -30.588 < 2e-16 ***
educ1        -5.794    12.931  -0.448 0.654084  
educ2         7.338    11.900   0.617 0.537486  
educ3         1.920    12.969   0.148 0.882338  
educ4         6.311    11.688   0.540 0.589197  
educ5         1.373    11.780   0.117 0.907191  
educ6        20.042    11.913   1.682 0.092504 .  

```

cowmain2	12.430	3.772	3.295	0.000986	***
cowmain3	58.033	6.228	9.319	< 2e-16	***
cowmain4	10.539	6.187	1.703	0.088537	.
cowmain5	75.587	12.446	6.073	1.27e-09	***
cowmain6	-29.907	5.334	-5.607	2.08e-08	***
cowmain7	-82.373	36.196	-2.276	0.022871	*
agyownk2	19.117	2.558	7.474	8.06e-14	***
agyownk3	29.748	2.897	10.269	< 2e-16	***
agyownk4	38.071	3.063	12.431	< 2e-16	***
naics_212	-7.666	19.446	-0.394	0.693421	
naics_213	-237.132	25.598	-9.264	< 2e-16	***
naics_214	-35.054	11.692	-2.998	0.002719	**
naics_215	-86.065	13.623	-6.318	2.70e-10	***
naics_216	-70.701	10.057	-7.030	2.12e-12	***
naics_217	-50.582	10.590	-4.777	1.79e-06	***
naics_218	-50.707	10.710	-4.735	2.21e-06	***
naics_219	-54.651	10.998	-4.969	6.77e-07	***
naics_2110	-79.291	10.159	-7.805	6.19e-15	***
naics_2111	-56.786	10.363	-5.480	4.30e-08	***
naics_2112	-72.802	10.412	-6.992	2.78e-12	***
naics_2113	-52.598	12.125	-4.338	1.44e-05	***
naics_2114	-65.738	10.063	-6.533	6.58e-11	***
naics_2115	-83.332	11.161	-7.466	8.53e-14	***
naics_2116	-115.812	10.529	-10.999	< 2e-16	***
naics_2117	-77.732	9.961	-7.803	6.26e-15	***
naics_2118	-96.218	11.152	-8.628	< 2e-16	***
naics_2119	-92.518	11.138	-8.306	< 2e-16	***
naics_2120	-82.794	10.779	-7.681	1.63e-14	***
naics_2121	-72.835	10.715	-6.797	1.09e-11	***
prov11	-19.090	8.901	-2.145	0.031986	*
prov12	-48.971	7.038	-6.959	3.52e-12	***
prov13	-5.084	6.651	-0.764	0.444624	
prov24	-19.307	5.348	-3.610	0.000307	***
prov35	-52.239	5.158	-10.129	< 2e-16	***
prov46	-6.715	6.129	-1.096	0.273202	
prov47	-4.793	6.332	-0.757	0.449081	
prov48	-2.678	5.946	-0.450	0.652454	
prov59	-26.037	5.666	-4.595	4.35e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 160.3 on 25174 degrees of freedom

Multiple R-squared: 0.1074, Adjusted R-squared: 0.1058

F-statistic: 67.29 on 45 and 25174 DF, p-value: < 2.2e-16

[1] 380404.8

[1] 380668

Analysis of Variance Table

Response: hrlyearn

	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
utothrs	1	6.0284e+08	602844604	213.797	< 2.2e-16 ***						
gender	1	2.7134e+09	2713409268	962.304	< 2.2e-16 ***						
educ	6	1.0719e+10	1786543590	633.593	< 2.2e-16 ***						
cowmain	1	1.0633e+09	1063343505	377.112	< 2.2e-16 ***						
agyownk	3	5.4014e+08	180045051	63.852	< 2.2e-16 ***						
age_12	1	3.5661e+08	356607717	126.470	< 2.2e-16 ***						
noc_10	9	1.1785e+10	1309481641	464.404	< 2.2e-16 ***						
prov	9	1.4695e+09	163277112	57.906	< 2.2e-16 ***						
Residuals	21470	6.0539e+10	2819701								

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'..'	0.1	' '	1

#Chi-Squared and ANOVA Tests Dependence Between Categorical Variables

Pearson's Chi-squared test

data: table(labor\$age_12, labor\$marstat)
X-squared = 68827, df = 55, p-value < 2.2e-16

Pearson's Chi-squared test

data: table(labor\$agyownk, labor\$marstat)
X-squared = 1050.7, df = 15, p-value < 2.2e-16

Pearson's Chi-squared test

data: table(labor\$agyownk, labor\$age_12)
X-squared = 22809, df = 33, p-value < 2.2e-16

Pearson's Chi-squared test

```
data: table(labor$naics_21, labor$noc_10)
X-squared = 192817, df = 180, p-value < 2.2e-16
```

Pearson's Chi-squared test

```
data: table(labor$naics_21, labor$noc_43)
X-squared = 299653, df = 840, p-value < 2.2e-16
```

```
      eta.sq eta.sq.part
prov   0.01513395  0.01537168
gender 0.01512012  0.01535785
```

```
[1] 0.1968539
```