Exploring Child care in Ontario in Toronto

1. Introduction

We are looking into the reading, math and general knowledge scores for fall 1998 and spring 1999 measurements, evaluating kindergarten students over the span of several months, based on the total house income.

1. **Research Question 1:** Are students' total house income affects their scores?
2. **Research Question 2:** Are students' total house income affects their general scores through months?

2. Exploratory Data Analysis (EDA)

	fallreadingscore	fallmathscore	fallgeneralknowledgescore	springreadingscore	springmathscore	springgeneralknowledgescore	totalhouseholdincome	incomeinthousands	incomegroup
0	36.58	39.54	33.822	49.37	50.10	40.533	140000.0	140.0	3
1	50.82	44.44	38.147	83.50	58.93	37.409	120000.0	120.0	3
2	40.68	28.57	28.108	40.41	32.81	29.312	90000.0	90.0	3
3	32.57	23.57	15.404	34.14	35.25	27.382	50000.0	50.0	2
4	31.98	19.65	18.727	32.84	23.60	26.977	55000.0	55.0	2
...	...	...	...	...	...	...	...	...	...
11928	32.51	25.68	11.694	45.63	24.27	20.189	39000.0	39.0	1
11929	31.96	25.73	21.461	47.84	40.35	32.357	26000.0	26.0	1
11930	40.44	30.50	16.836	49.93	41.00	19.803	45000.0	45.0	2
11931	35.51	28.96	28.864	51.60	50.27	35.991	28000.0	28.0	1
11932	23.47	14.30	15.256	31.41	17.49	18.283	13000.0	13.0	1

We proceeded with a comprehensive EDA to leverage insight that could potentially lead to interesting research questions. We started by describing our quantitative data as shown in *Figure 1* below. Additionally, we employed boxplots seen in *Figure 2* to visually represent the distribution of these features, after removing outliers. This descriptive analysis offered a clearer understanding of the general trends within each feature

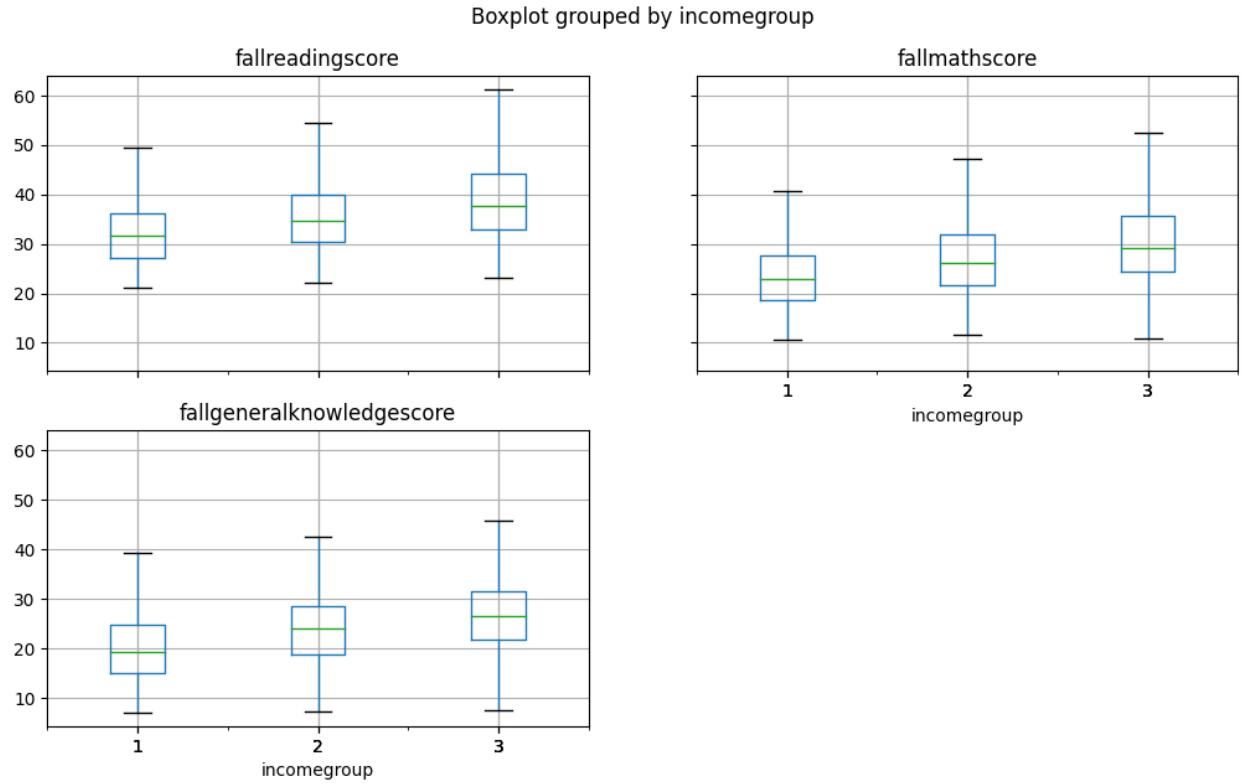


Figure1- fall scores

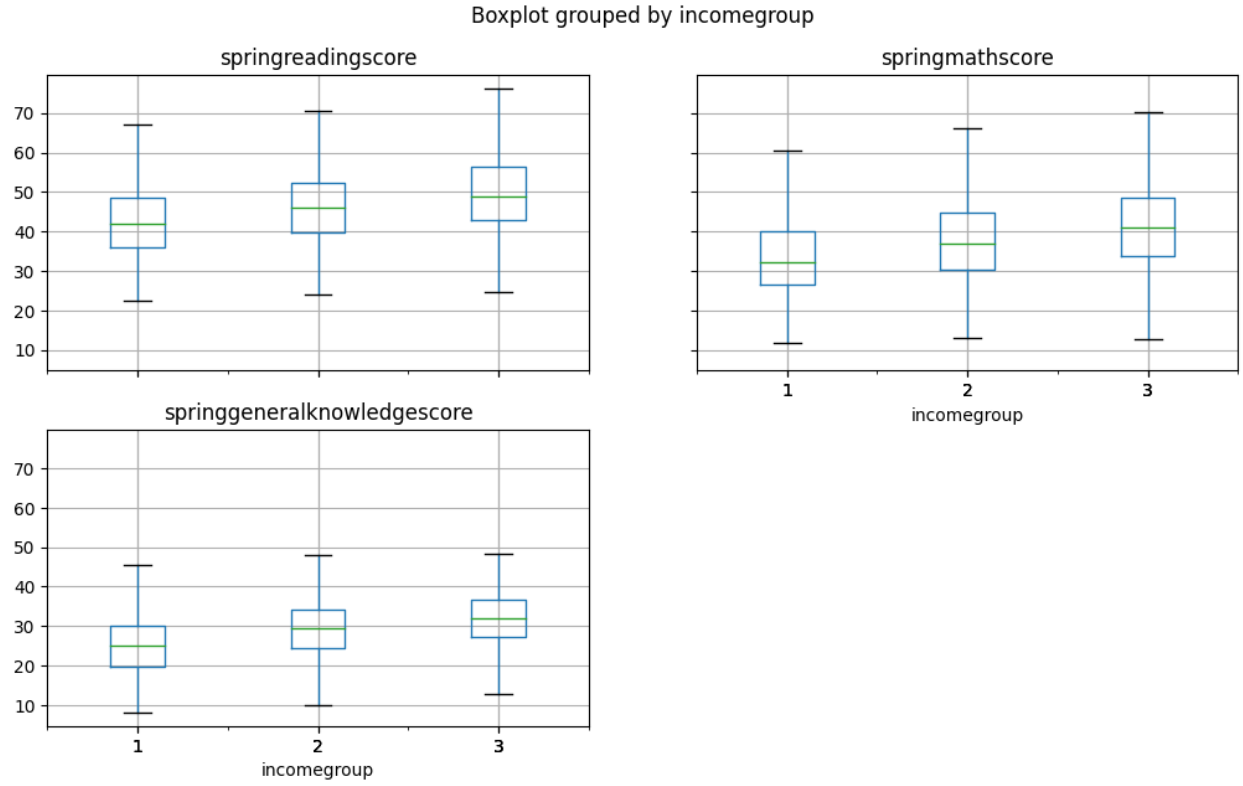


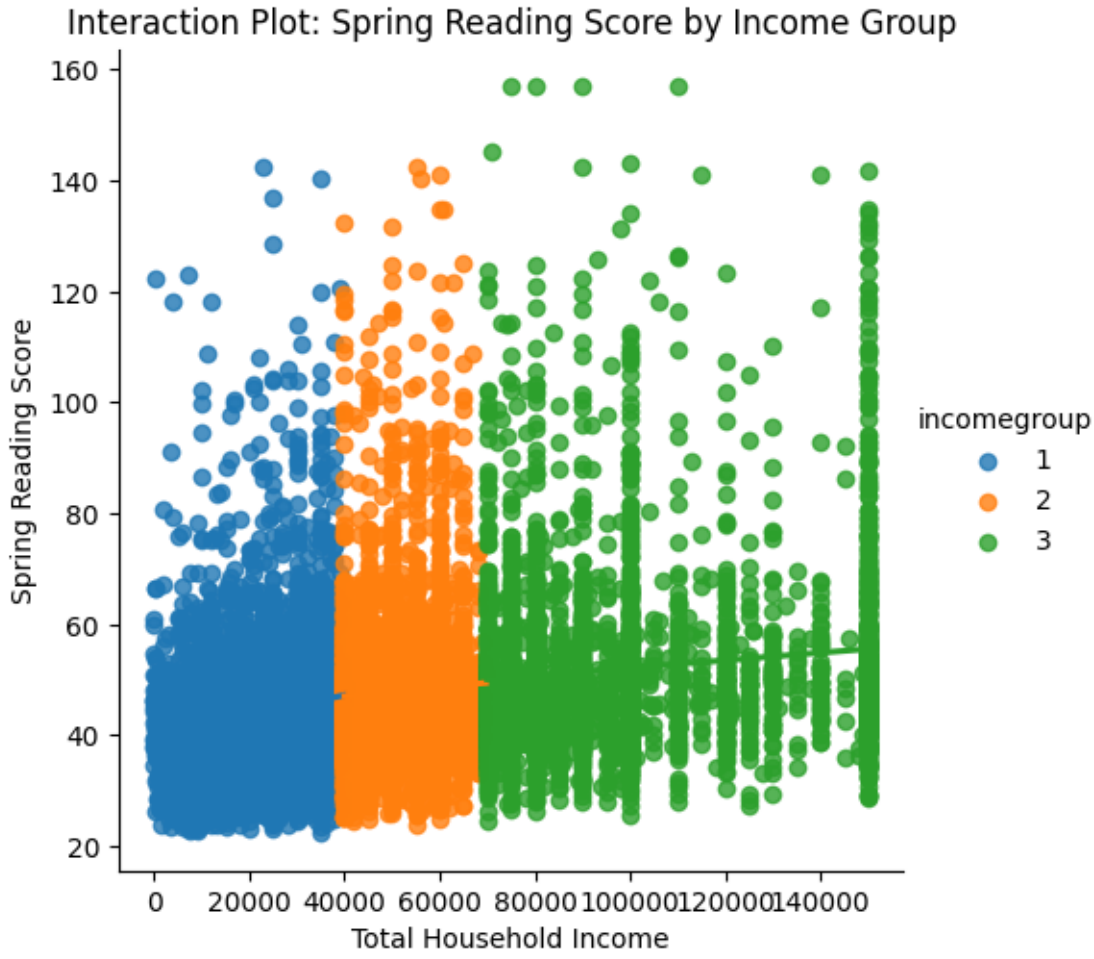
Figure2 – spring scores

Based on the boxplot above, we can see that students has higher total house income has better average scores and all students have better average scores through month

3. Ancova for spring

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     springreadingscore   R-squared:                       0.071
Model:                            OLS   Adj. R-squared:                  0.071
Method:                 Least Squares   F-statistic:                     304.1
Date:                Sun, 24 Mar 2024   Prob (F-statistic):          2.89e-190
Time:                        03:11:39   Log-Likelihood:                -48259.
No. Observations:               11933   AIC:                         9.653e+04
Df Residuals:                   11929   BIC:                         9.656e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                 41.7466      0.259    161.430      0.000      41.240      42.253
C(incomegroup)[T.2]        1.7547      0.374      4.691      0.000       1.022       2.488
C(incomegroup)[T.3]        1.6613      0.661      2.514      0.012       0.366       2.957
totalhouseholdincome   8.713e-05    7.4e-06     11.775      0.000    7.26e-05       0.000
==============================================================================
Omnibus:                     6092.761   Durbin-Watson:                   1.762
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            50336.400
Skew:                           2.317   Prob(JB):                         0.00
Kurtosis:                      11.931   Cond. No.                     3.73e+05
==============================================================================
```

Here is an example Ancova result, C(incomegroup)[T.2] and C(incomegroup)[T.3]: These represent the differences in the mean spring reading score between each level of the incomegroup variable and the reference group (usually level 1). For example, the coefficient 1.7547 for incomegroup level 2 indicates that, on average, the spring reading score for level 2 is 1.7547 units higher than the reference group (level 1) after controlling for the effect of the covariate and other predictors.



4. Anova for Fall compare to Spring

0 秒

OLS Regression Results

Dep. Variable:	fallgeneralknowledgescore	R-squared:	0.733
Model:	OLS	Adj. R-squared:	0.733
Method:	Least Squares	F-statistic:	8196.
Date:	Sun, 24 Mar 2024	Prob (F-statistic):	0.00

```
Time:                        03:43:07   Log-Likelihood:              -32926.
No. Observations:               11933   AIC:                       6.586e+04
Df Residuals:                   11928   BIC:                       6.590e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================================
                                    coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Intercept                        -0.4464      0.139     -3.217      0.001      -0.718      -0.174
C(incomegroup)[T.2]               0.4068      0.104      3.907      0.000       0.203       0.611
C(incomegroup)[T.3]               0.6040      0.183      3.300      0.001       0.245       0.963
totalhouseholdincome           8.379e-06   2.07e-06      4.052      0.000    4.33e-06    1.24e-05
springgeneralknowledgescore       0.8061      0.005    161.236      0.000       0.796       0.816
==============================================================================
Omnibus:                       71.653   Durbin-Watson:                   1.881
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              104.912
Skew:                          -0.028   Prob(JB):                     1.65e-23
Kurtosis:                       3.456   Cond. No.                     3.78e+05
==============================================================================
```

The model has an R-squared value of 0.733, indicating that approximately 73.3% of the variance in fall general knowledge scores is explained by the independent variables included in the model.
The F-statistic is highly significant ($p < 0.001$), indicating that the overall model is statistically significant.
All coefficients, except for the intercept, are statistically significant ($p < 0.05$). This suggests that income group, total household income, and spring general knowledge score significantly predict fall general knowledge score.
The omnibus test result indicates that the model significantly deviates from the assumption of normality ($p < 0.001$). However, this might not be a concern, especially with a large sample size.
The Durbin-Watson statistic is approximately 1.881, indicating no significant autocorrelation among residuals.
The Jarque-Bera test result suggests that the residuals do not follow a normal distribution ($p < 0.001$).
Assumptions:

While the model is statistically significant, it's essential to consider potential violations of assumptions, such as normality of residuals and homogeneity of variances.
Overall, the ANCOVA model indicates that income group, total household income, and spring general knowledge score are significant predictors of fall general knowledge score.

However, further analysis and diagnostics are necessary to ensure that the model meets the underlying assumptions of ANCOVA.

5. Conclusion

Through quantitative analysis and visual exploration, we gained valuable insights about higher house income in total, the better scores students gain.