

## Introduction

The Canada-Wide Early Learning & Child Care (CWELCC) is a five-year government strategy aiming “to make child care more accessible and affordable; support the early childhood workforce; and improve accessibility by addressing barriers to providing inclusive and flexible child care.”<sup>1</sup>. One critical aspect of accessible and flexible child care is the availability of spaces across early development: from infancy to the age children are enrolled in full time school.

Centres can be classified into three auspice sectors: non-profit, commercial, and public. The primary objective of this data analysis is to characterize the current spaces available for infants, toddlers, preschool/kindergarten, children and grade one and up. Moreover, we aim to assess whether CWELCC subsidy spaces differ across (1) age group, and (2) age groups across each auspice

The following data analysis is driven by dual research questions:

**Research Question 1:** Is there a difference in the average number of spaces available for early childhood age groups with a subsidy contract?

**Research Question 2:** Are there significant differences in the spaces available across auspice and age group?

## 1. Data

Licensed Child Care Centres is an open source dataset of 1063 licensed centres in the City of Toronto including their capacities by age group, CWELCC membership, and subsidy status.

- 926 documented centers participating in CWELCC, of which 705 are have a fee subsidy contract. Note: all Public programs are subsidized.
- For the scope of this analysis, I am not interested in the 13 centers receiving subsidies that are **not** in the CWELCC.

Thus, the following analysis will focus on a subset of the data containing ( $N = 705$ ) child-care centers meeting criterion: (i) registered under the CWELCC and (ii) has a fee subsidy contract.

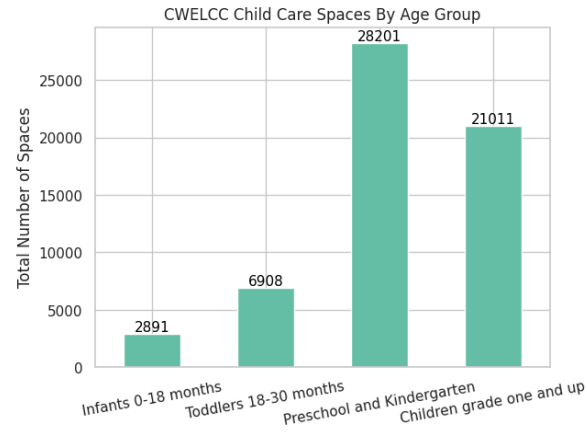
### 1.1 Data Transformations and Variables of Interest

A new categorical column for ‘Preschool and Kindergarten’ is created as the initial data coding scheme overlaps in definitions for PG and KG age groups. The merged columns are then dropped from the dataframe, leaving four age groups (Figure 1).

*Figure 1. Count of CWELCC Child Care Spaces*

---

<sup>1</sup>[www.toronto.ca/community-people/community-partners/early-learning-child-care-partners/canada-wide-early-learning-child-care-agreement-cwelcc/](http://www.toronto.ca/community-people/community-partners/early-learning-child-care-partners/canada-wide-early-learning-child-care-agreement-cwelcc/)



## 2. Methods

### 2.1 One Way ANOVA

To tackle **Research Question 1**, the following hypothesis is tested:

$H_0$ : The mean subsidized space for each age group is equal ( $\mu_1 = \mu_2 = \mu_3 = \mu_4$ )

$H_A$ : There is a difference between mean subsidized space available for each age group.

#### 2.1.0

Figure 2. Boxplot for Age Group

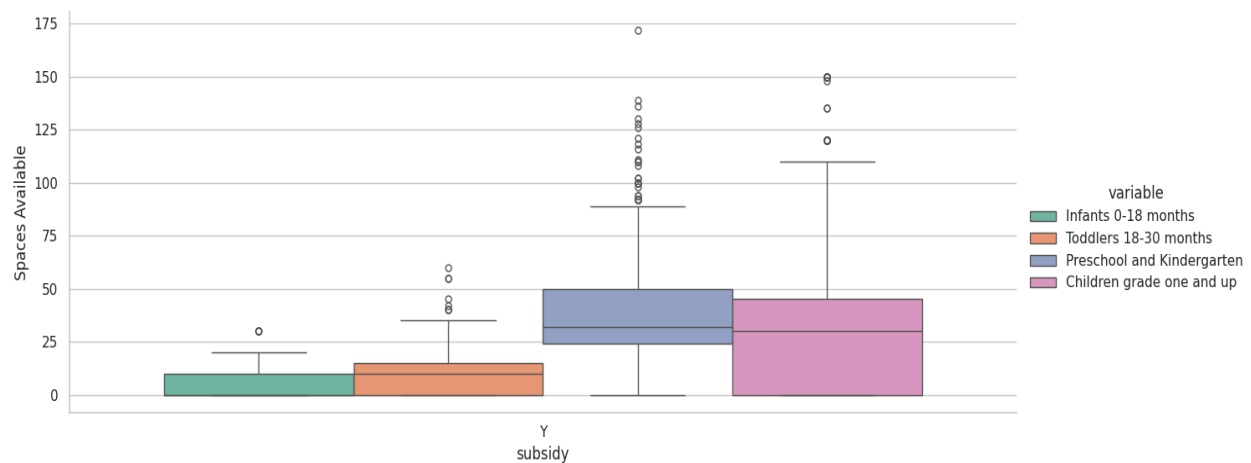


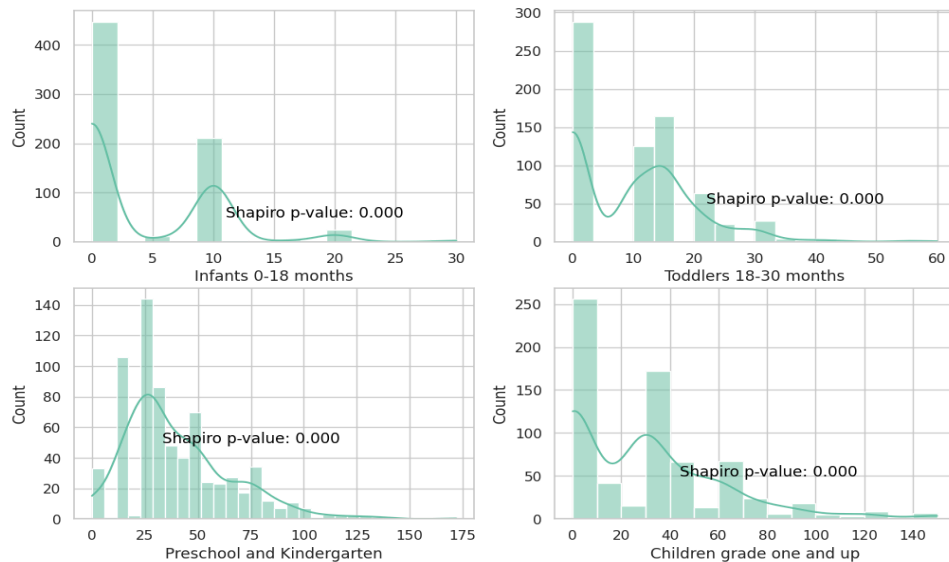
Table 1. Describing CWELCC Subsidized Spaces By Age Group

	Infants 0-18 months	Toddlers 18-30 months	Preschool and Kindergarten	Children grade one and up
mean	4.1	9.8	40.0	29.8
median	0.0	10.0	32.0	30.0
std	5.92	9.92	25.45	31.41
min	0.0	0.0	0.0	0.0
max	30.0	60.0	172.0	150.0

### 2.1.1 Assumptions

#### ✗ Assumption 1: Normality

The Shapiro-Wilk test tests the null hypothesis that each group was drawn from a normal distribution. A p-value less than  $\alpha = .05$  suggests that we should reject our null hypothesis— the assumption of normality is violated for every age group.



#### ✗ Assumption 2 : Homoscedasticity

Levene's test is used to test the the null hypothesis that each group has equal variance. Levene's statistic is 286.07 with a p-value  $< .005$ . The test result is significant, suggesting non-homogeneity of variances shown in Figure 2.

#### ✓ Assumption 3 : Independence

Samples are drawn from independent groups.

### 2.1.2 Inferential Statistics

**One Way ANOVA Results:** The one-way ANOVA test revealed a highly significant difference in the average number of spaces available among the four age groups ( $F = 451.72, p < .001$ ). The extremely low p-value suggests that there is strong evidence to reject the null hypothesis, indicating that at least one age group has a significantly different mean number of spaces.

### 2.1.3 Post Hoc Tests

The Tukey HSD test was conducted to compare the mean differences in the number of spaces available across the four age groups. The results are summarized below:

*Table 2. Tukey Pairwise Test for Age Groups*

Group 1	Group 2	Mean Difference	95% CI	p < .05
Children grade one and up	Infants 0-18 months	-25.7021	(-28.58, -22.82)	Yes
Children grade one and up	Preschool and Kindergarten	10.1986	(7.32, 13.07)	Yes
Children grade one and up	Toddlers 18-30 months	-20.0043	(-22.88, -17.12)	Yes

Infants 0-18 months	Preschool and Kindergarten	35.9007	(33.03, 38.7)	Yes
Infants 0-18 months	Toddlers 18-30 months	5.6979	(2.81, 8.57)	Yes
Preschool and Kindergarten	Toddlers 18-30 months	-30.2028	(-33.08, -27.32)	Yes

The results of the Tukey HSD test indicate significant mean differences in the number of spaces available among all pairs of age groups within the CWELCC subsidy program. Each comparison resulted in a rejection of the null hypothesis, suggesting that the mean number of spaces available is significantly different between the compared age groups.

## 2.2 Two Way ANOVA

To answer our second research question, the following three alternate hypothesis are tested using a two way analysis of covariance on the independent variables 3(Auspice) \* 4(Age Groups), creating twelve distinct treatment groups.

$H_{A1}$  : There is a difference between the groups in the mean space for **age group**.

$H_{A2}$  : There is a difference between the groups in the mean space for **auspice**.

$H_{A3}$  : There is a **joint effect of age group and auspice** on space available.

### 2.2.0 Assumptions

Assumptions for the distribution of data in each of the categorical levels for 'Auspice' are tested. For assumptions on the distribution of 'Age Groups' see 2.1.0.

✗ Assumption 1: **Normality**

Shapiro-Wilk test for Non Profit: ( $p < .001$ ), Commercial: ( $p < .001$ ), Public: ( $p < .001$ )

✗ Assumption 2 : **Homoscedasticity**

Levene's test for equality of variances (Levene's test statistic = 19.571,  $p < .001$ )

✓ Assumption 3 : **Independence**

Samples are drawn from independent Auspices. .

### 2.2.2 Inferential Statistics

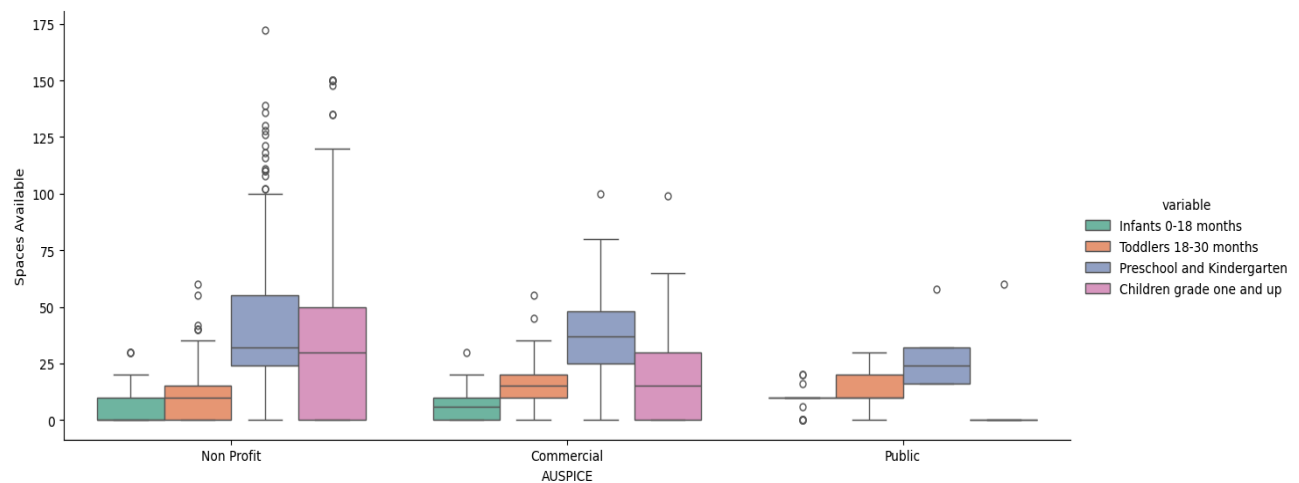
Table 3. Two Way ANOVA Results

	df	sum_sq	F	PR(>F)
<b>Auspice</b>	2.0	14402.58	17.15	3.94e-08
<b>Age Group</b>	3.0	598953.67	475.52	7.95e-250
<b>Auspice x Age Group</b>	6.0	51258.73	20.35	1.98e-23
<b>Residual</b>	2808.0	1178962.53		

- **Main Effect of Age Group:** The main effect of 'AgeGroup' is highly statistically significant ( $F = 475.52$ ,  $p < 0.05$ ), as indicated in our One Way ANOVA and follow up test.
- **Main Effect of Auspice:** The main effect of 'AUSPICE' is statistically significant ( $F = 17.15$ ,  $p < 0.05$ ). This suggests that there are significant differences in the mean space available' among different levels of 'AUSPICE.'

- **Interaction Effect (Auspice x AgeGroup):** The interaction effect between 'AUSPICE' and 'AgeGroup' is statistically significant ( $F = 20.35$ ,  $p < 0.05$ ).

*Figure 4. Boxplot of Spaces by Age Group and Auspice*

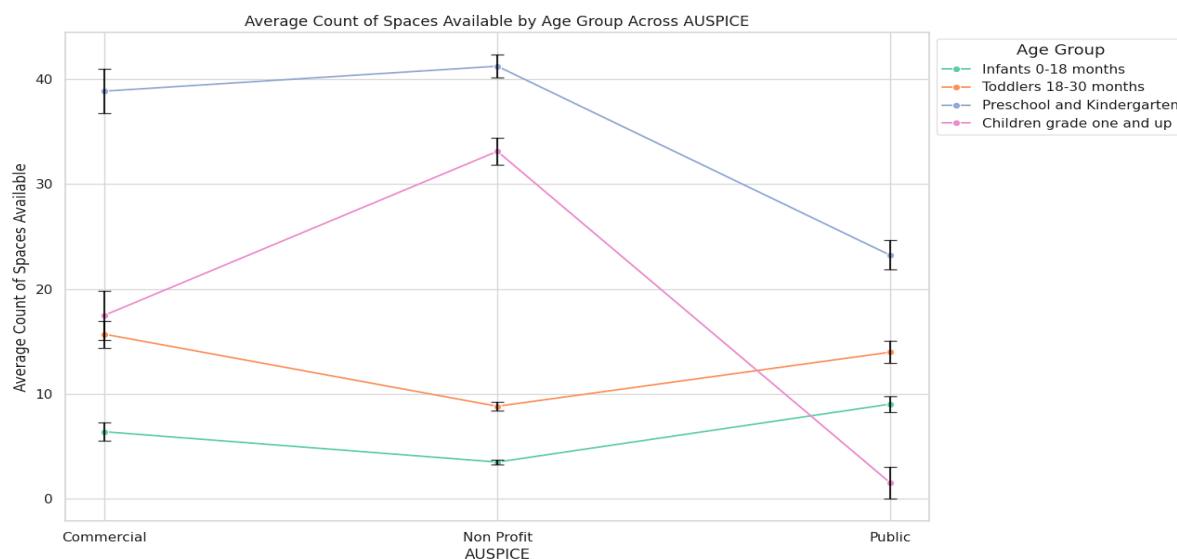


Moving from the commercial and non-profit to public sector, the joint effect of auspice and age group have a significant influence on the number of spaces for children grade one and up and preschoolers/kindergarten, with fewer available spaces. This interaction is visually indicated by the intersecting lines and reversal of magnitude in the number of spaces moving across sectors.

### 2.2.3 Post Hoc Tests

The post hoc test (Tukey's HSD) was conducted to explore the pairwise differences in mean spaces available following a significant interaction effect in the two-way ANOVA. For brevity, results are visualized with an interaction plot below. The full table including F-statistics and mean differences can be found in the corresponding code.

*Figure 5. Interaction Plot (Auspice\*Age Group)*



### 3. Discussion and Conclusion

The objective of our data analysis was to gain insights into the landscape of licensed childcare centers participating in the Canada-Wide Early Learning & Child Care (CWELCC) program with a fee subsidy contract.

#### → *Research Question 1:*

Analysis revealed a substantial difference in the average number of subsidized spaces available among Infants 0-18 months, Toddlers 18-30 months, Preschool and Kindergarten, and Children grade one and up. **Post hoc tests further identified significant mean differences between each pair of age groups**, indicating varied space allocations across early childhood stages. Notably, the number of spaces for Preschool and Kindergarten-aged children was significantly higher than those for other age groups, while Children grade one and up had fewer available spaces in comparison.

#### → *Research Question 2:*

Both age group and auspice significantly influenced the mean number of spaces available. Second, there was a significant interaction effect between auspice and age group, indicating that **the joint influence of these factors is essential in understanding space allocations**.

#### 3.1 Implications and Future Directions

Our findings shed light on critical disparities in the availability of subsidized childcare spaces. Follow-up studies could delve deeper into the underlying factors contributing to these observed variations, including regional, population, and demographic influences. Additionally, assessing the impact of these disparities on the accessibility and quality of childcare services for families would provide valuable insights. Further analyses could explore the ratio of spaces to the population of children in each ward, household income rates, and proportions of allocated childcare spaces using statistical tests of proportions.

#### 3.2 Limitations

This data analysis is limited in its failure to examine outliers in the skewed dataset. While our study provides valuable insights, we acknowledge its limitations, including violations of normality and homogeneity of variance assumptions. Violation of these assumptions reduces our statistical power, and increases the likelihood of a false positive. Hence, results should be interpreted with caution. Future iterations may consider employing alternative tests or non-parametric statistics to account for unequal variances.