

# Investigating the infection rates in counties categorized under the CDC community transmission levels

Sina Mokhtar

## Abstract

---

In September 2020, the Centers for Disease Control and Prevention (CDC) introduced a framework for monitoring COVID-19 community transmission levels based on four categories: high, substantial, moderate, and low. This framework aimed to reduce the risk of severe illness and healthcare strain by using two key metrics: the total number of new cases per 100,000 persons in the past 7 days and the percentage of positive Nucleic Acid Amplification Test results during the same period. These community transmission levels have been instrumental in guiding setting-specific guidance and layered prevention strategies, such as screening testing in schools and implementing masking policies.

This study focuses on investigating the infection rates in counties categorized under the four different CDC COVID-19 community transmission levels. Additionally, we compare the infection rates of the counties with different risk level three weeks later to evaluate the effectiveness of the community transmission risk level classification in predicting future infection trends.

## METHODS

---

### Data Sources

The data for this study was obtained from the following sources

1- [Covid Act Now API](#): This API provides comprehensive COVID-19 data, including metrics related to transmission risk and infection rates.

2- CDC, National Center for Health Statistics [NCHS](#) This classification system categorizes counties based on their urban or rural characteristics, providing important contextual information for analyzing COVID-19 transmission patterns.

### Statistical Analysis

These statistics provide an overview of the infection rate values for different risk levels, both in the present and 3 weeks later. They give insights into the mean and median infection rates within each CDC Transmission risk level.

To determine if there is a significant difference in the distribution of infection rate numbers among counties with different CDC Transmission Risk levels, we used two statistical tests: ANOVA (Analysis of Variance) and Kruskal-Wallis test.

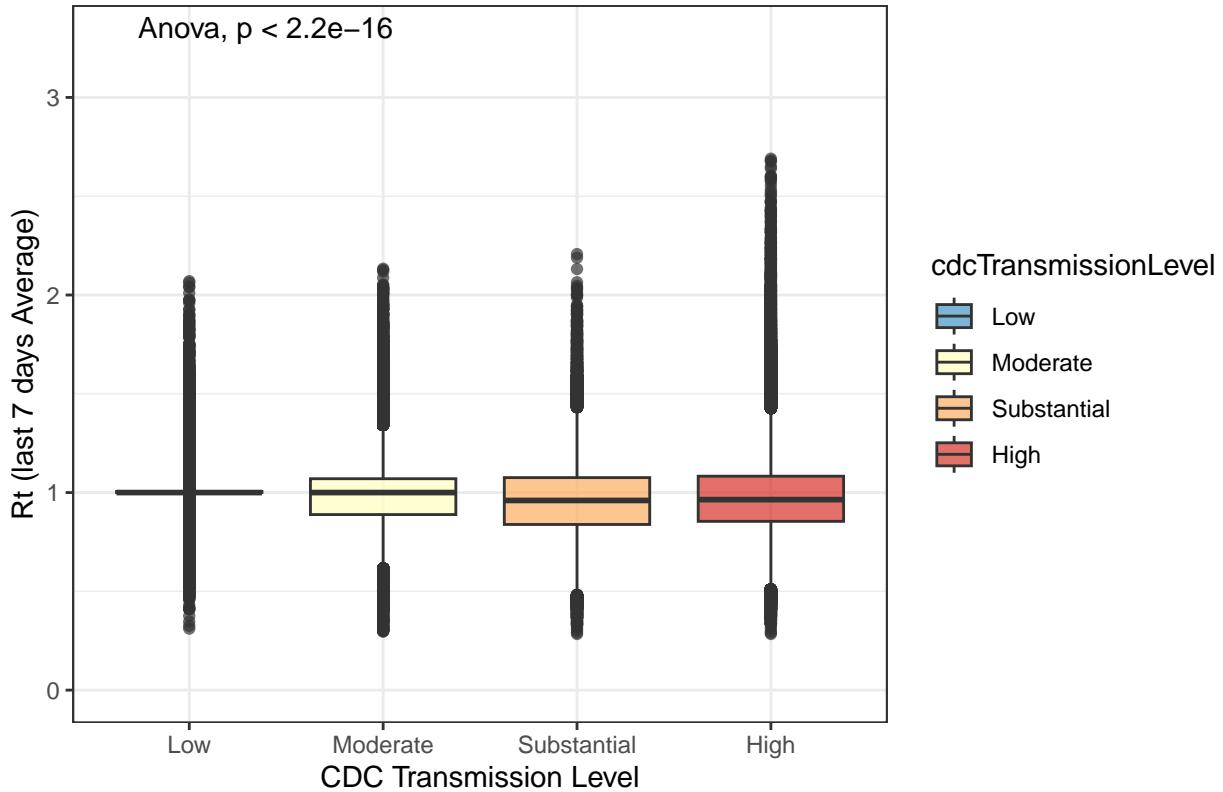
One-way **ANOVA** model that is used to test whether there is a significant difference in means between the groups defined by the CDC Transmission Risk level. The null hypothesis is that there is no significant difference in means between the groups, and the alternative hypothesis is that at least one group mean is different from the others.

Table 1: Summary statistics by CDC transmission level

| Risk Level  | Counties | Mean Rt   | Mean Rt (3 Weeks) | Median Rt | Median Rt (3 Weeks) |
|-------------|----------|-----------|-------------------|-----------|---------------------|
| Low         | 215725   | 1.0064128 | 1.0981825         | 1.0000000 | 1.021429            |
| Moderate    | 577466   | 0.9857376 | 1.0758070         | 1.0000000 | 1.050000            |
| Substantial | 513856   | 0.9638845 | 1.0094027         | 0.9600000 | 1.000000            |
| High        | 1968276  | 0.9739620 | 0.9235935         | 0.9642857 | 0.920000            |

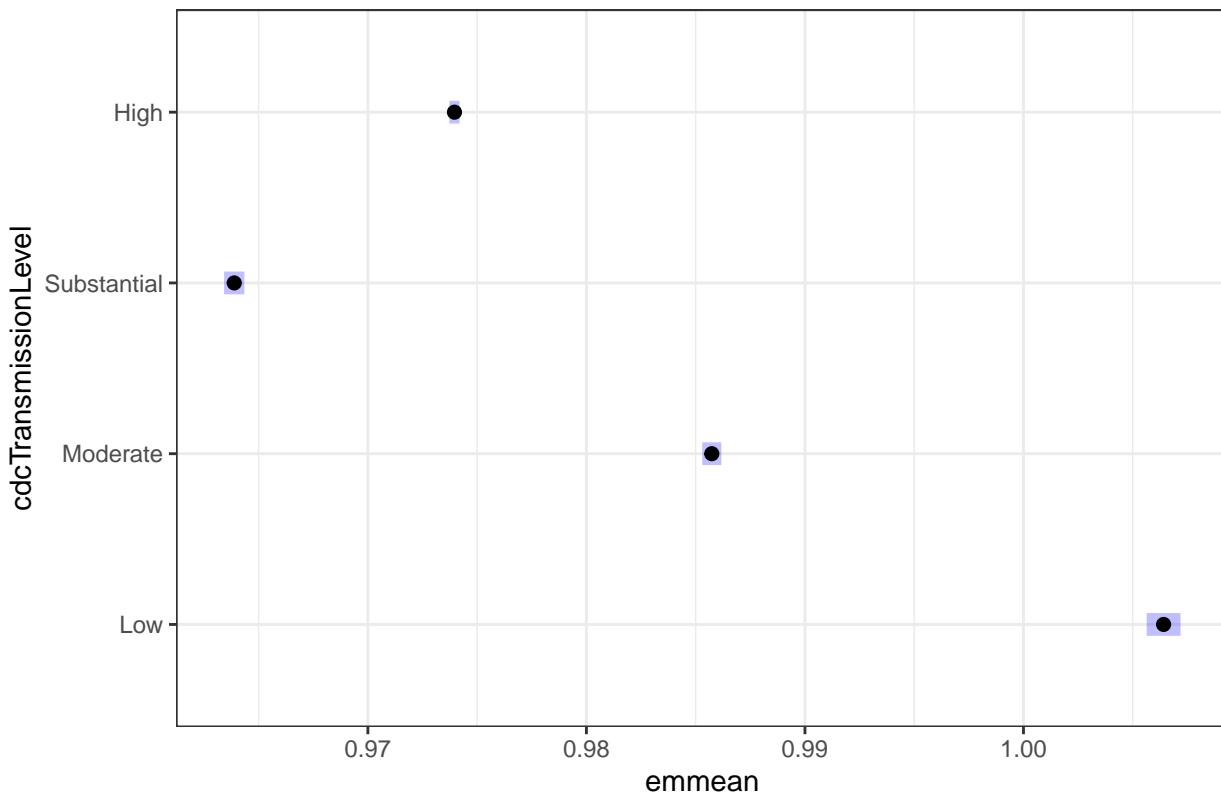
```
## Analysis of Variance Table
##
## Response: mean_last_7_days
##                               Df Sum Sq Mean Sq F value    Pr(>F)
## cdcTransmissionLevel      3   306   101.928     3569 < 2.2e-16 ***
## Residuals                 3225940  92130    0.029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Distribution of Rt by CDC transmission level



In this case, the p-value is 2e-16, which is smaller than the conventional significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a significant difference in means between the groups. To further analyze the specific group differences, we calculate the estimated marginal means, allowing us to identify significant differences among the groups and gain insights into the average response across different groups.

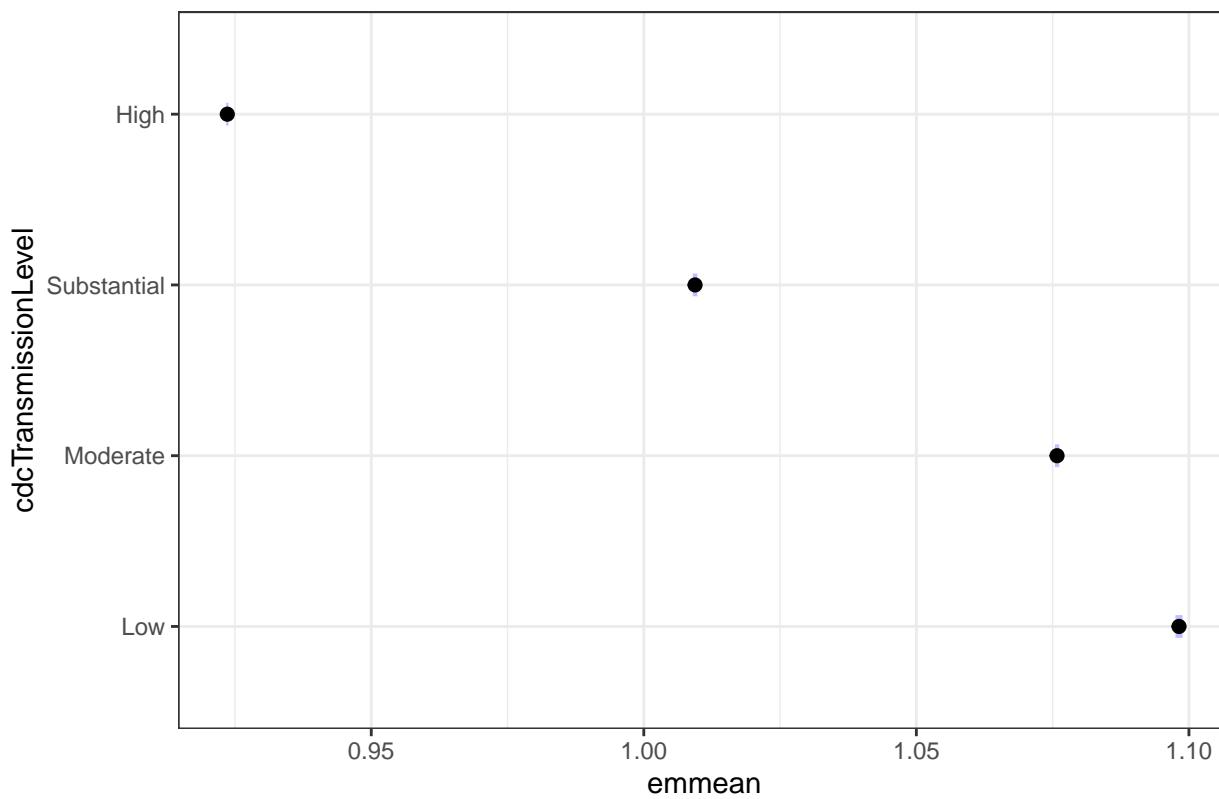
### Rt estimated marginal means for each CDC risk level



```
##   cdcTransmissionLevel    emmean          SE      df lower.CL upper.CL
##   Low                  1.00641  0.00039598 3225940  1.00564  1.00719
##   Moderate             0.98574  0.00022338 3225940  0.98530  0.98618
##   Substantial           0.96388  0.00023616 3225940  0.96342  0.96435
##   High                 0.97396  0.00012073 3225940  0.97373  0.97420
##
##   ## Confidence level used: 0.95
```

The results show that the Low transmission group has the highest mean, followed by the Moderate and High transmission groups, and the Substantial transmission group has the lowest mean. The p-values for all the groups are very small ( $p < 0.001$ ), indicating that the differences in means are statistically significant.

Rt estimated marginal means for each CDC risk level 3 weeks later

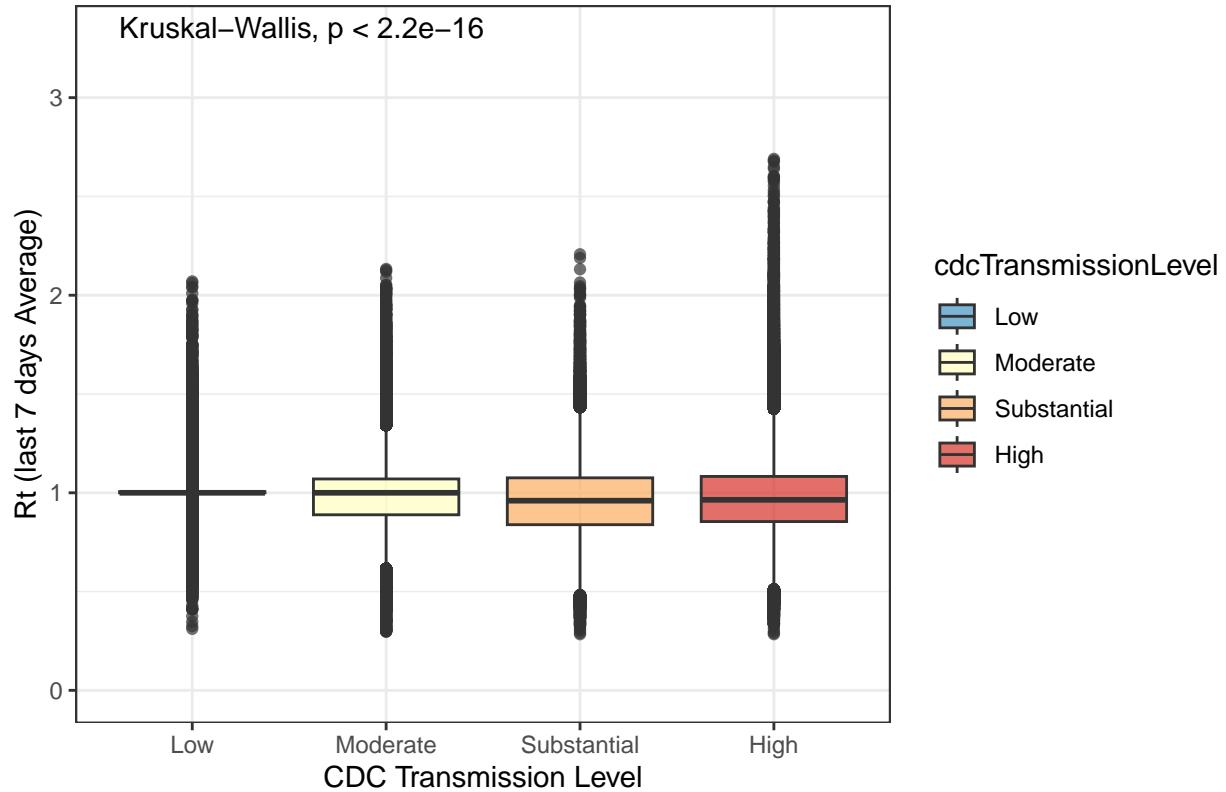


```
##   cdcTransmissionLevel    emmean          SE      df lower.CL upper.CL
##   Low                  1.09818 0.00033231 3275319  1.09753  1.09883
##   Moderate              1.07581 0.00020311 3275319  1.07541  1.07621
##   Substantial            1.00940 0.00021532 3275319  1.00898  1.00982
##   High                 0.92359 0.00011002 3275319  0.92338  0.92381
##
##   ## Confidence level used: 0.95
```

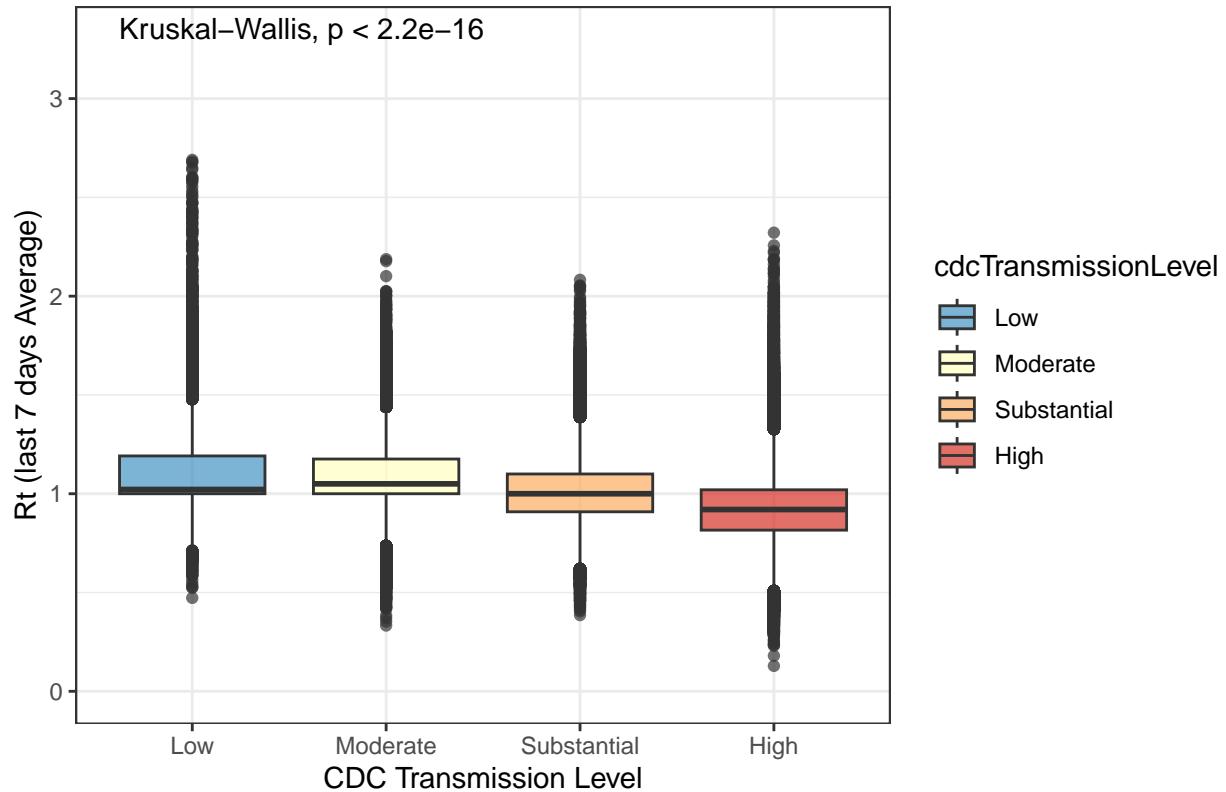
The **Kruskal-Wallis** test is a nonparametric statistical test used to determine whether three or more independent groups have different medians. It is an alternative to the one-way ANOVA test, which assumes that the data are normally distributed and have equal variances. Since in our case these assumptions are not met the Kruskal-Wallis test is more appropriate.

```
##  
##  Kruskal-Wallis rank sum test  
##  
##  data: Rt3NextWeeks by cdcTransmissionLevel  
##  Kruskal-Wallis chi-squared = 560965, df = 3, p-value < 2.2e-16
```

### Distribution of Rt by CDC transmission level



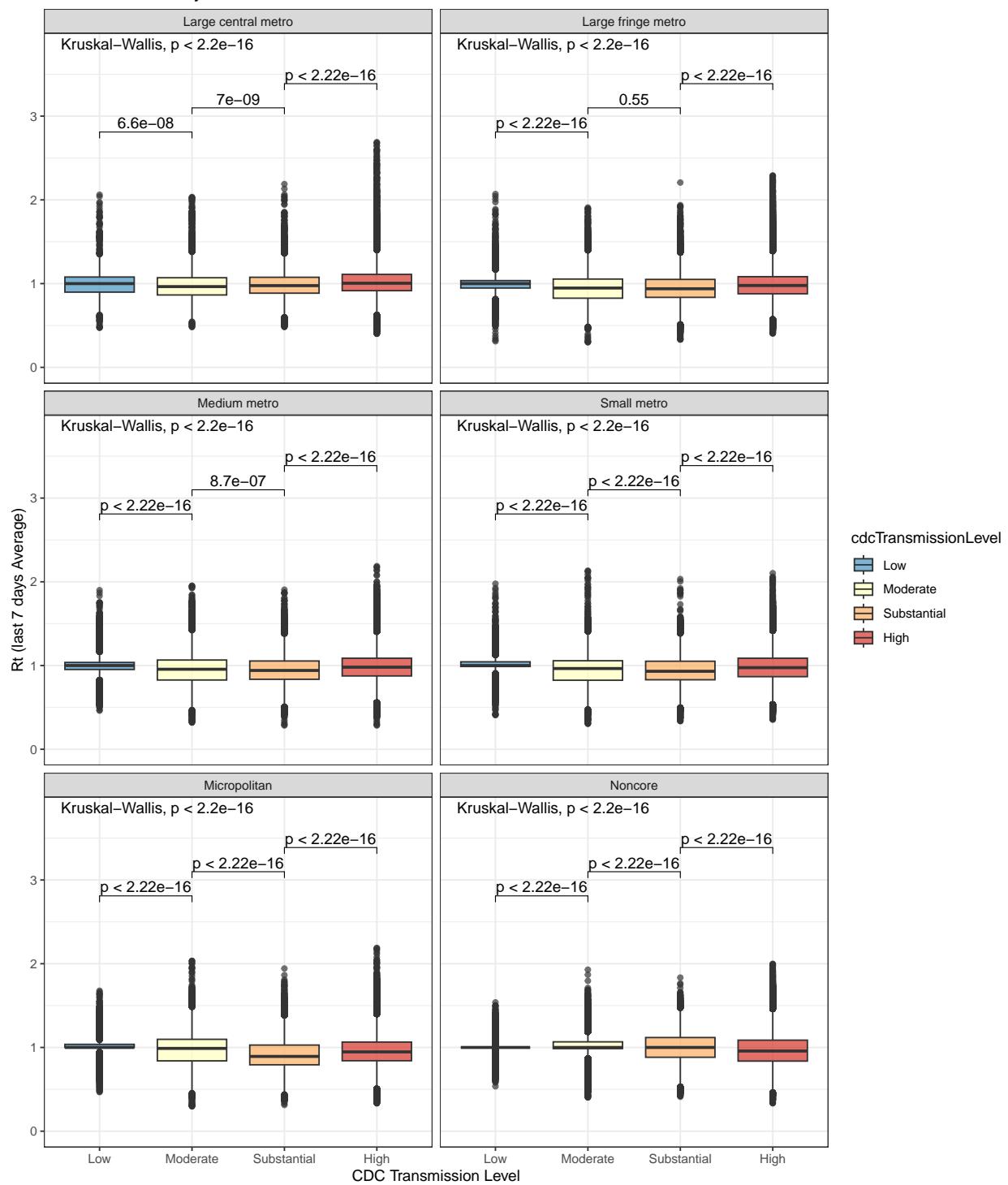
## Distribution of Rt by CDC transmission level at 21 days later



The output of the test suggests that the p-value is less than the significance level of 0.05, indicating strong evidence against the null hypothesis that there are no differences between the groups. Therefore, we can conclude that there is a significant difference in the distribution of the Rt number across the four levels of CDC Transmission risk level both at the same day and 21 days later.

Then we applied the Kruskal-Wallis same test in each NCHS category to see if there is significant difference in the distribution of the Rt number across CDC Transmission risk level groups both at the same day and 21 days later.

Distribution of Rt by CDC transmission level



Distribution of Rt by CDC transmission level at 21 days later

