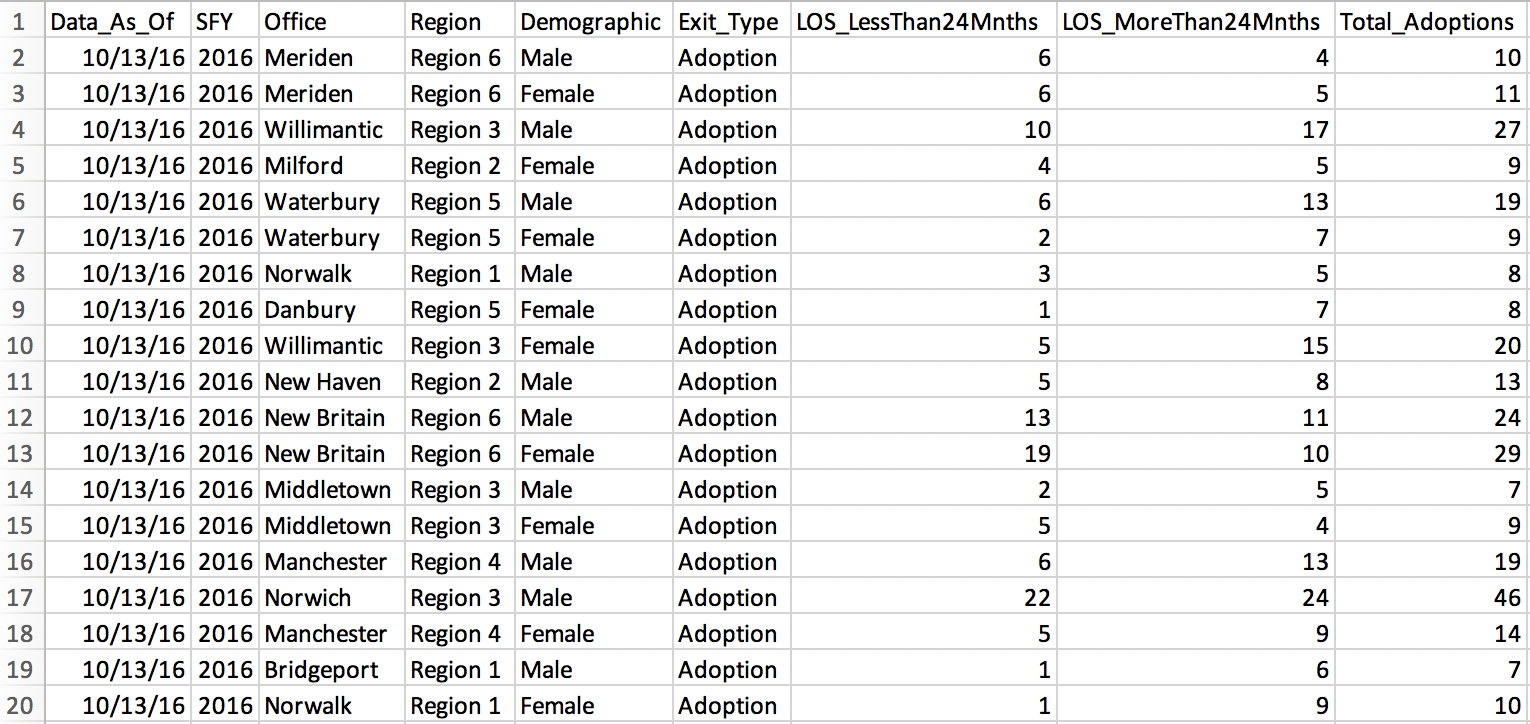
The dataset that I found for two samples hypothesis testing is Adoption by Department of Children and Families Office. In this dataset, it concerns the number of children that exited DCF care to an adoption. Below is a preview of the data table.



One of the claims is that children stay in DCF care more than 24 months had better chance to be adopted. Therefore in this form we are going to concentrate on columns: Length of Stay Less Than 24 Months and LOS\_MoreThan24Mnths. We are going to collect data in this two columns aggregated by different office and compare the the mean adoption numbers for two LOS by conducting a two sample t-test. We let μ1 and μ2 represent the population mean LOS less and more than 24 months. If the claim is true, then the mean LOS less than 24 months will be less than the mean in LOS more than 24 months. Thus, the elements of the test are as follows:

H0: (μ1 - μ2) = 0

H1: (μ1 - μ2) < 0

Then we are going to use t.test() in R to get test statistics (z value, p value and critical values). Then are will compare z value versus critical value and p value versus significance level(α) to decide whether we should reject H0.

I will continue to calculate all the test statistics in R later in this week's assignment and get the results if the samples provide sufficient evidence to reject H0. Therefore we would know if LOS more than 24 months had better chance to get the adoption.

[Adoptions\_by\_SFY\_\_DCF\_Office\_\_Gender\_and\_Length\_of\_Stay.csv](https://northeastern.blackboard.com/courses/1/ALY6015.80877.201935/db/_12168040_1/embedded/Adoptions_by_SFY__DCF_Office__Gender_and_Length_of_Stay.csv)