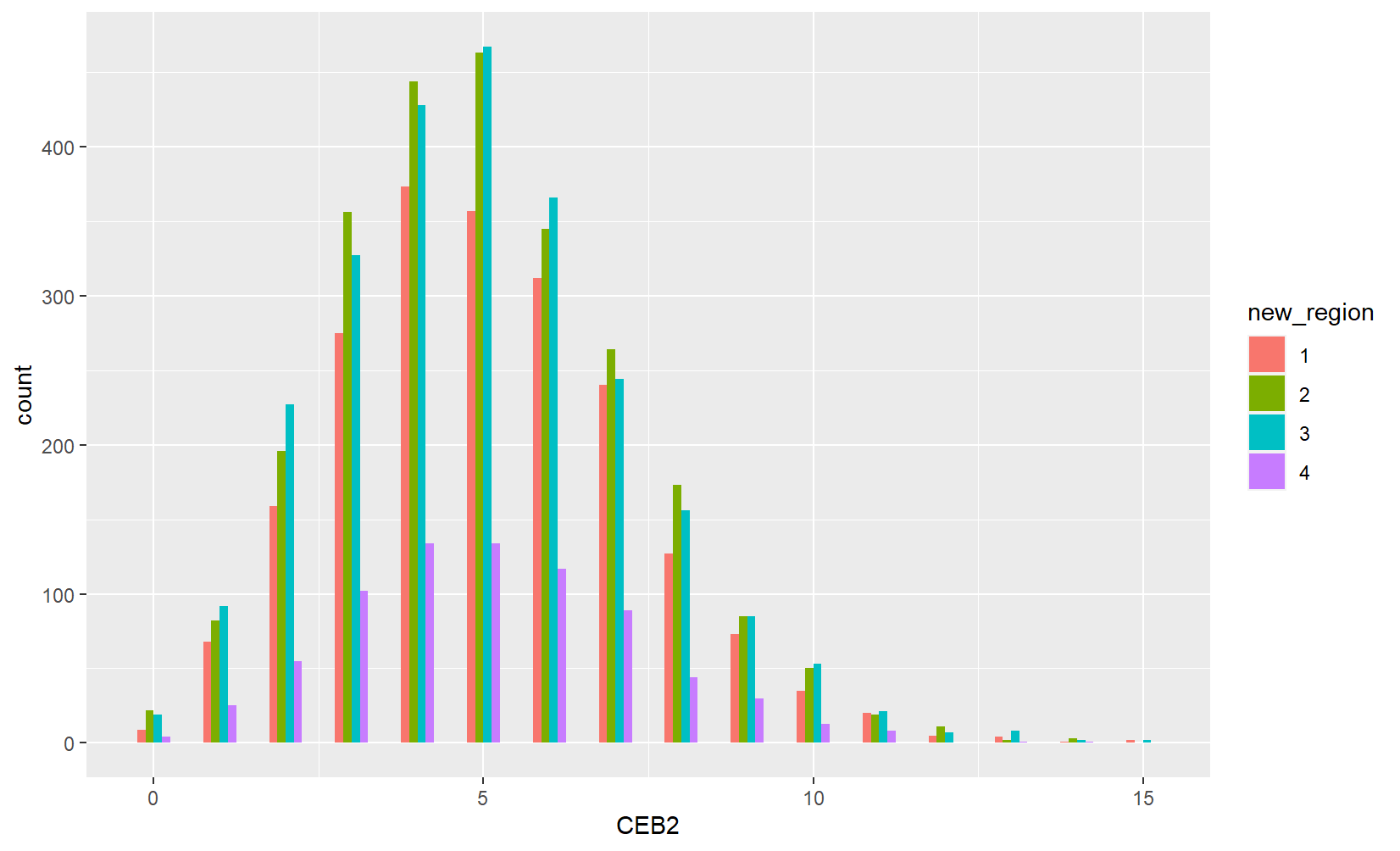
**Homework 2**

Name: Krishna Prasad Palvai

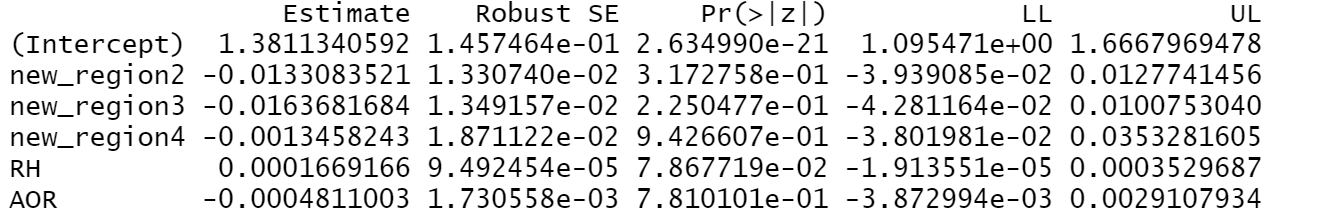
ID: 922672795

HW2.csv file is stored in d1. Categorical variable region is used to create new categorical variable with four levels. New variable is named new\_region. Mean and variance values of CEB1 by new region are calculated, over dispersion is observed so it is fitted by negative binomial regression. Mean and variance values of CEB2 by new region are calculated, both means and variance are equal so it is better fitted by Poisson regression.

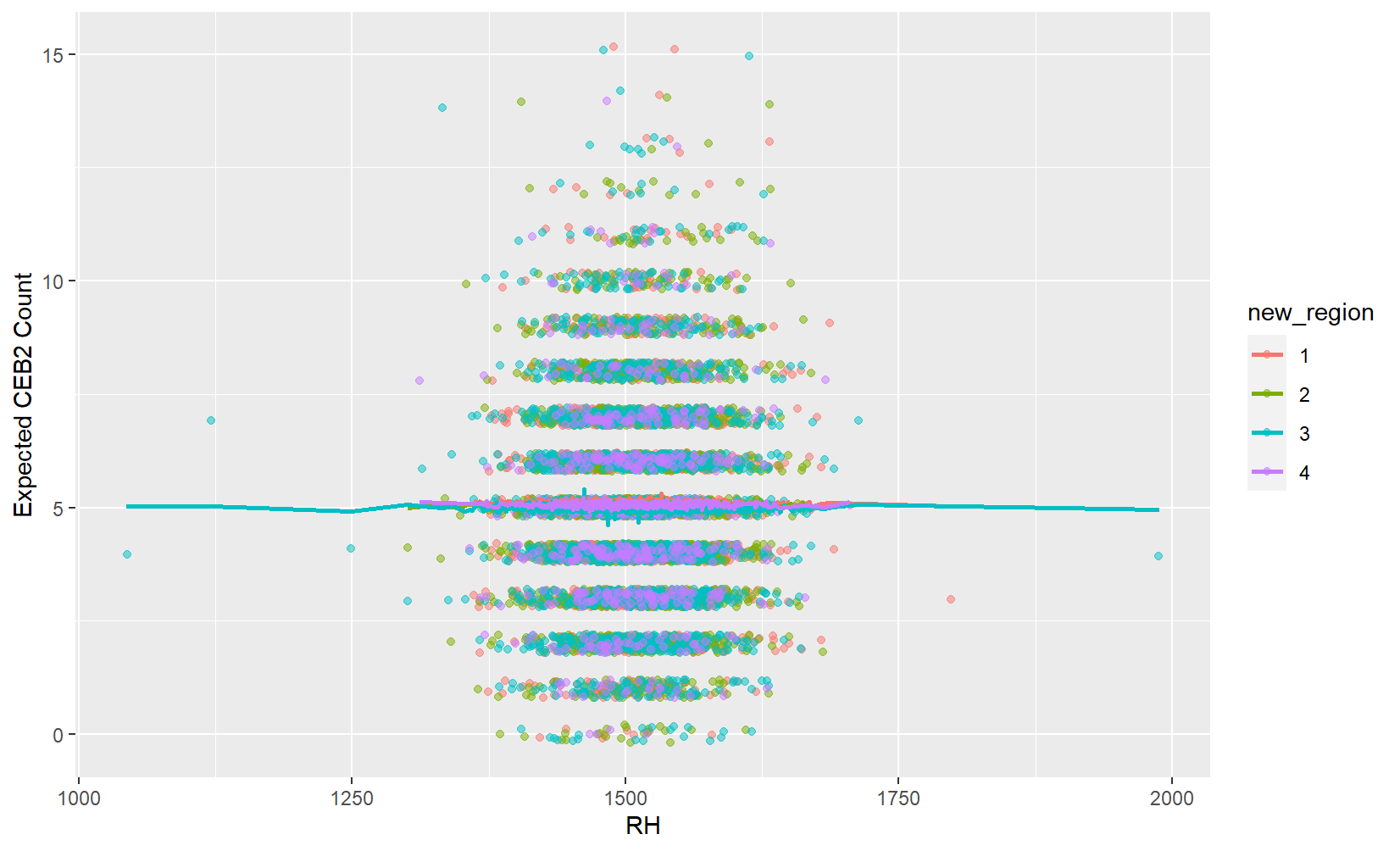


Above plot showing counts of observations for each CEB2 value for each new region category.

Poisson regression is fitted with CEB2 as response variable and RH, AOR and new\_region as covariates.

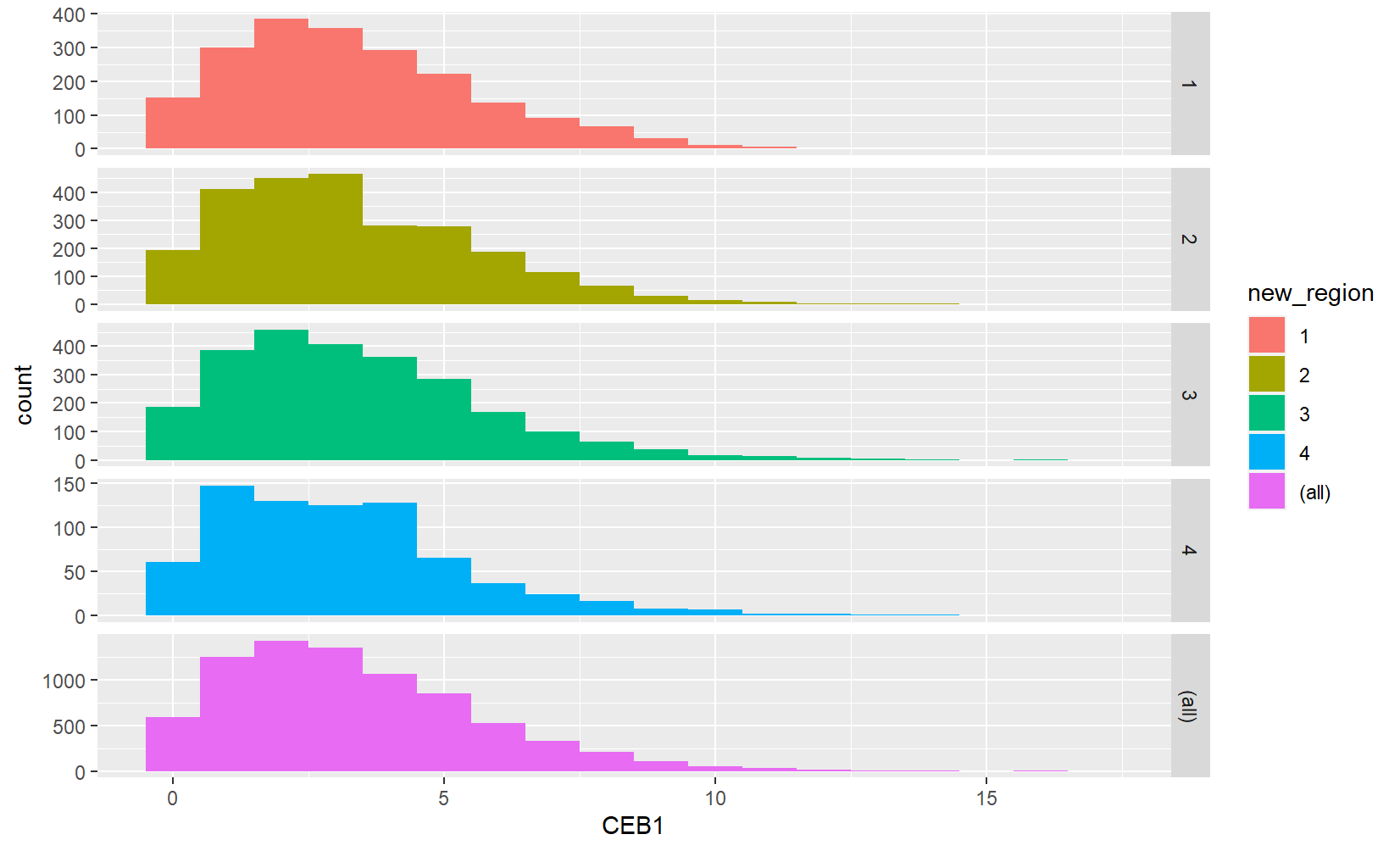


Coefficient of new\_region2 is -0.013308, this tells holding everything else constant if new region value changes from 1 to 2 expected log count decreases by 0.013308. With 1 unit increase in RH, expected log count increases by 0.0001669.



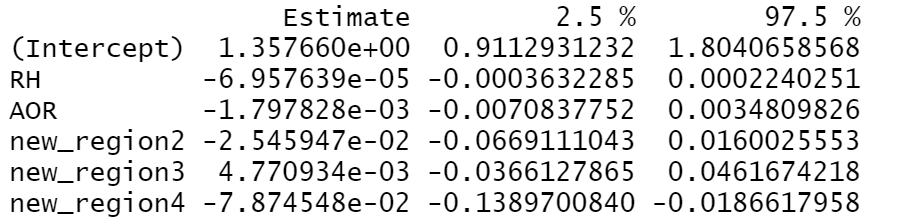
Plot showing expected CEB2 by line, points are colour coded based on new region value.

Negative binomial regression is fitted with CEB1 as response variable and RH, AOR and new\_region as covariates.

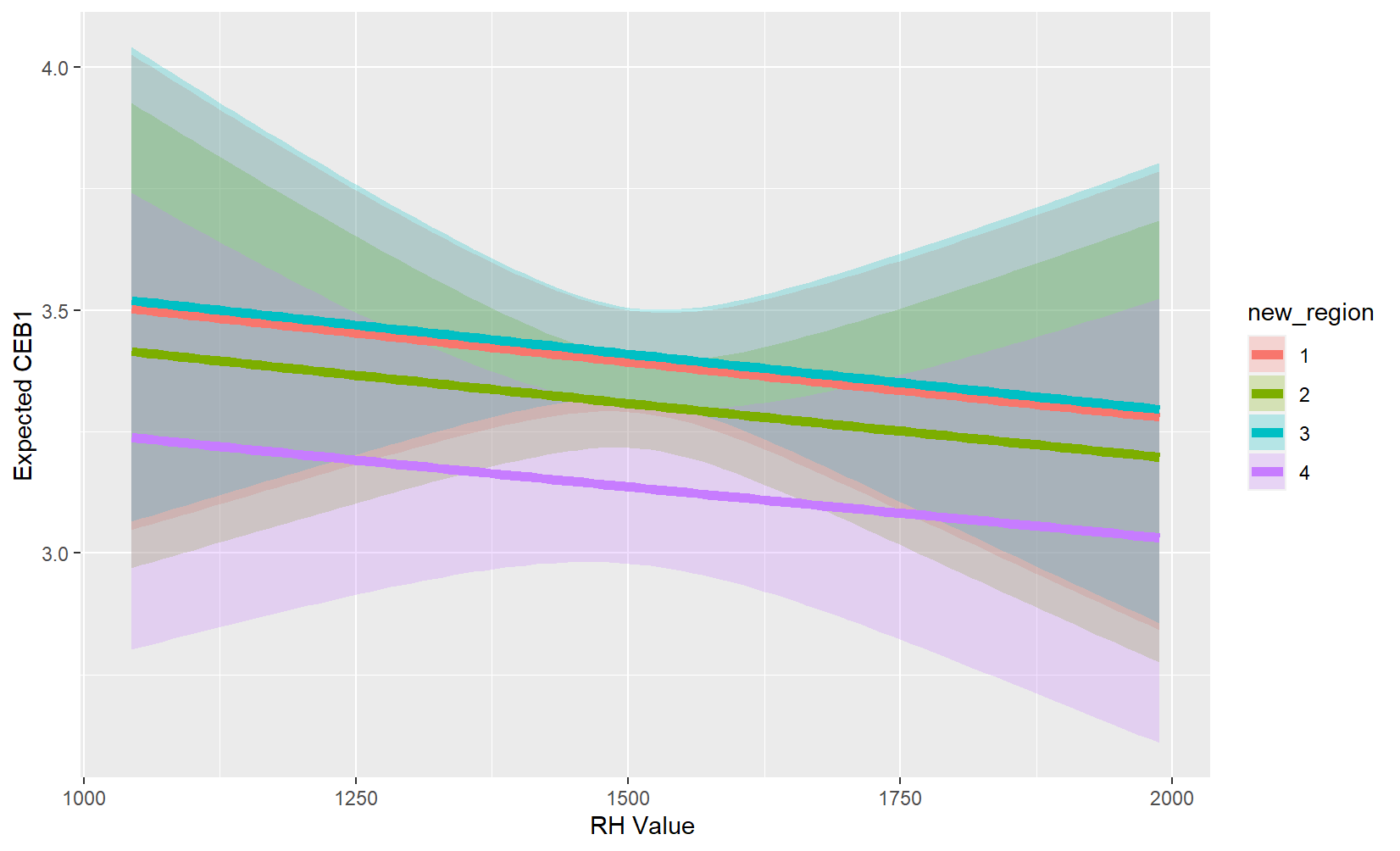


Histograms of CEB1 for each category and also bottom histogram shows counts of all categories of new region.

Output of negative binomial regression is below.



Coefficient of new\_region2 is -0.02546 this tells holding everything else constant if new region value changes from 1 to 2 expected log count decreases by 0.02546. With 1 unit increase in RH, expected log count decreases by 0.00006958.



Plot showing expected CEB1 value along with 95% confidence interval for expected CEB1 value.