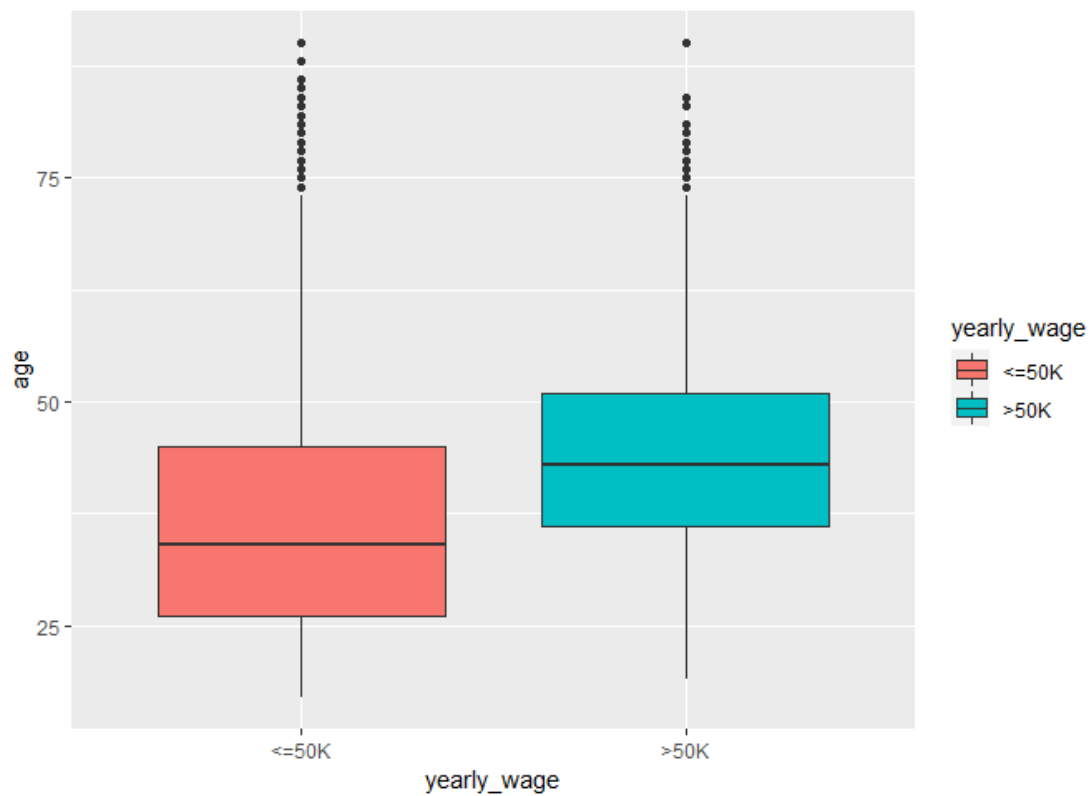
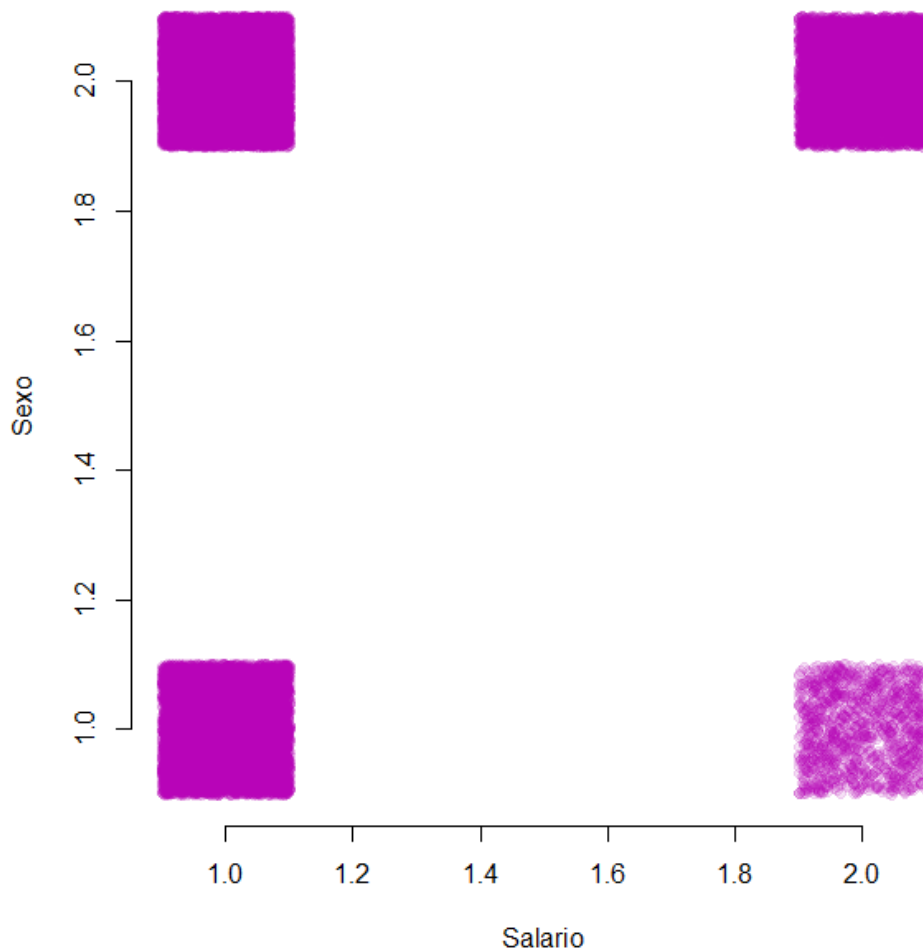


Relatório EDA

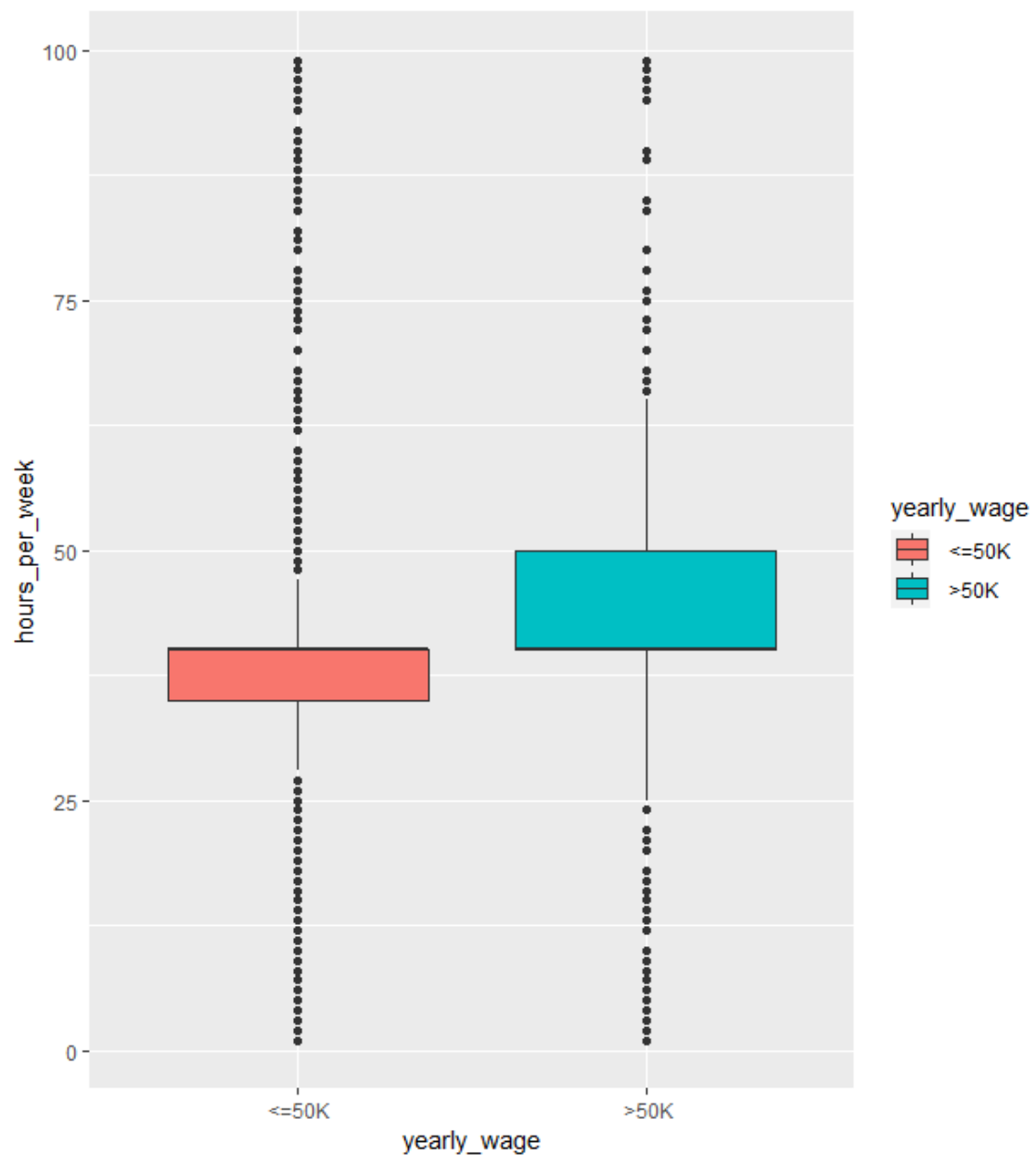
Inicialmente, podemos perceber padrões a partir da idade, percebemos a tendencia pessoas mais velhas em média, receberem mais. Essa é uma análise exploratória simples, com variável resposta sendo salário e a variável explicativa sendo a idade, esse gráfico nos ajuda a entender a dispersão dos nossos dados e como as médias se comportam.



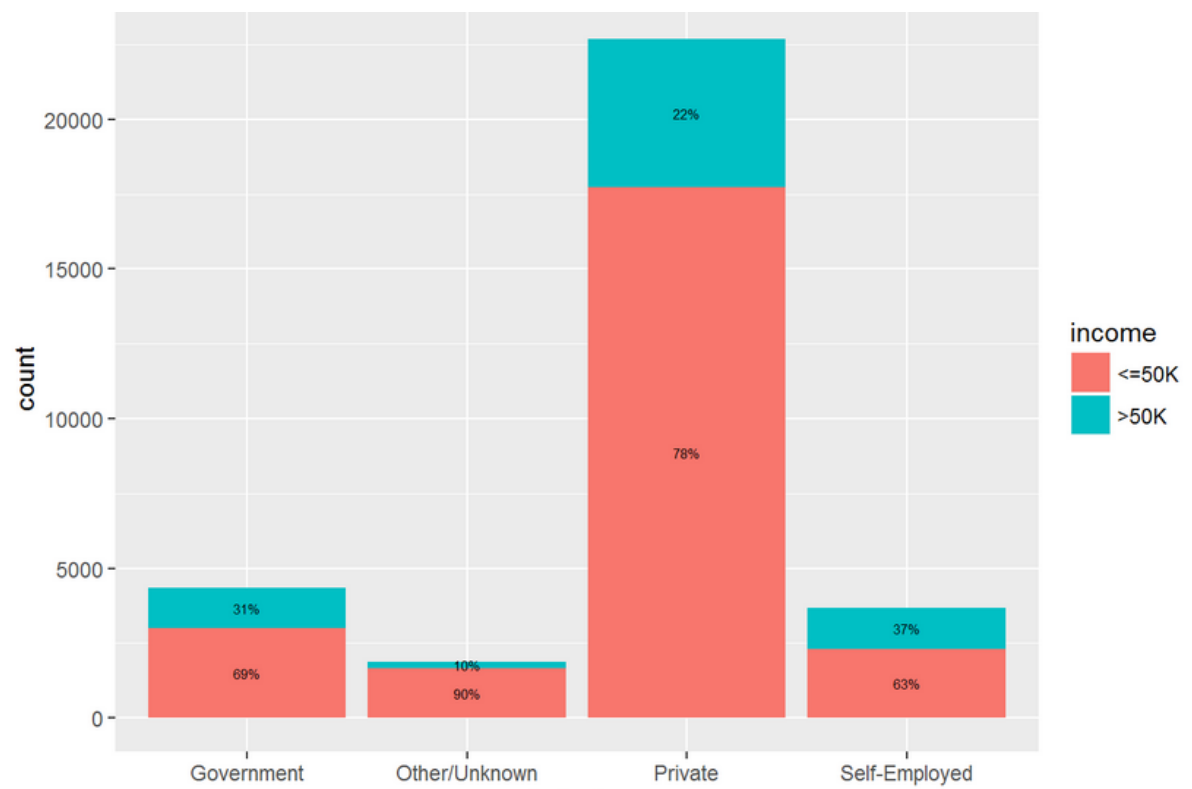
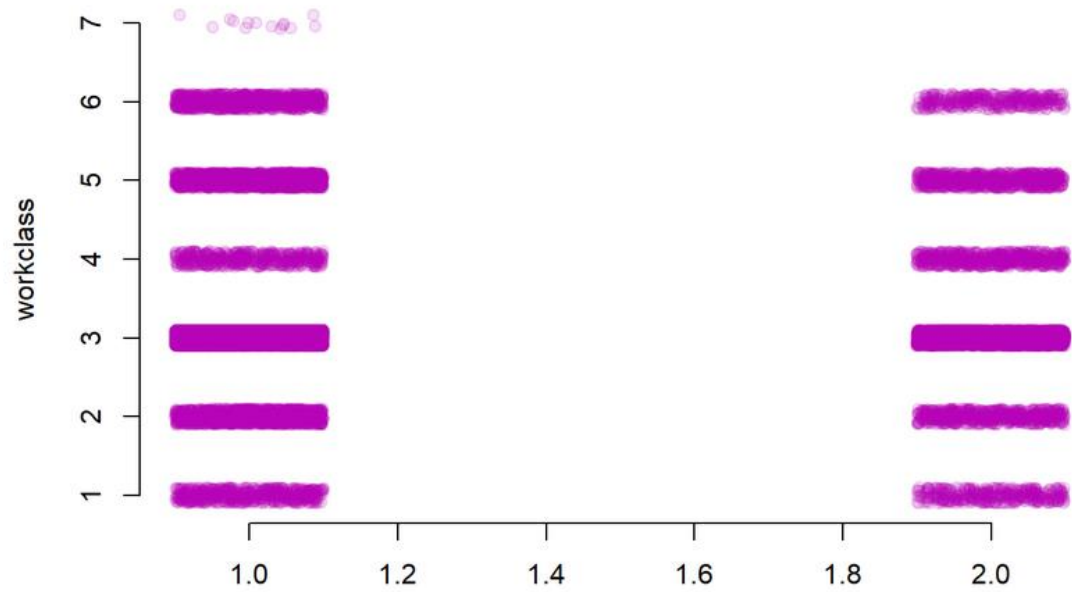
Existe também diferenças entre os sexos, existe uma tendência a homens possuírem salários mais altos, com variável resposta sendo salário (1 1 corresponde a “≤50K” e 2 corresponde a “>50K”) e a variável explicativa sendo o sexo, plotados em níveis sendo 1 o sexo feminino e 2 o sexo masculino.



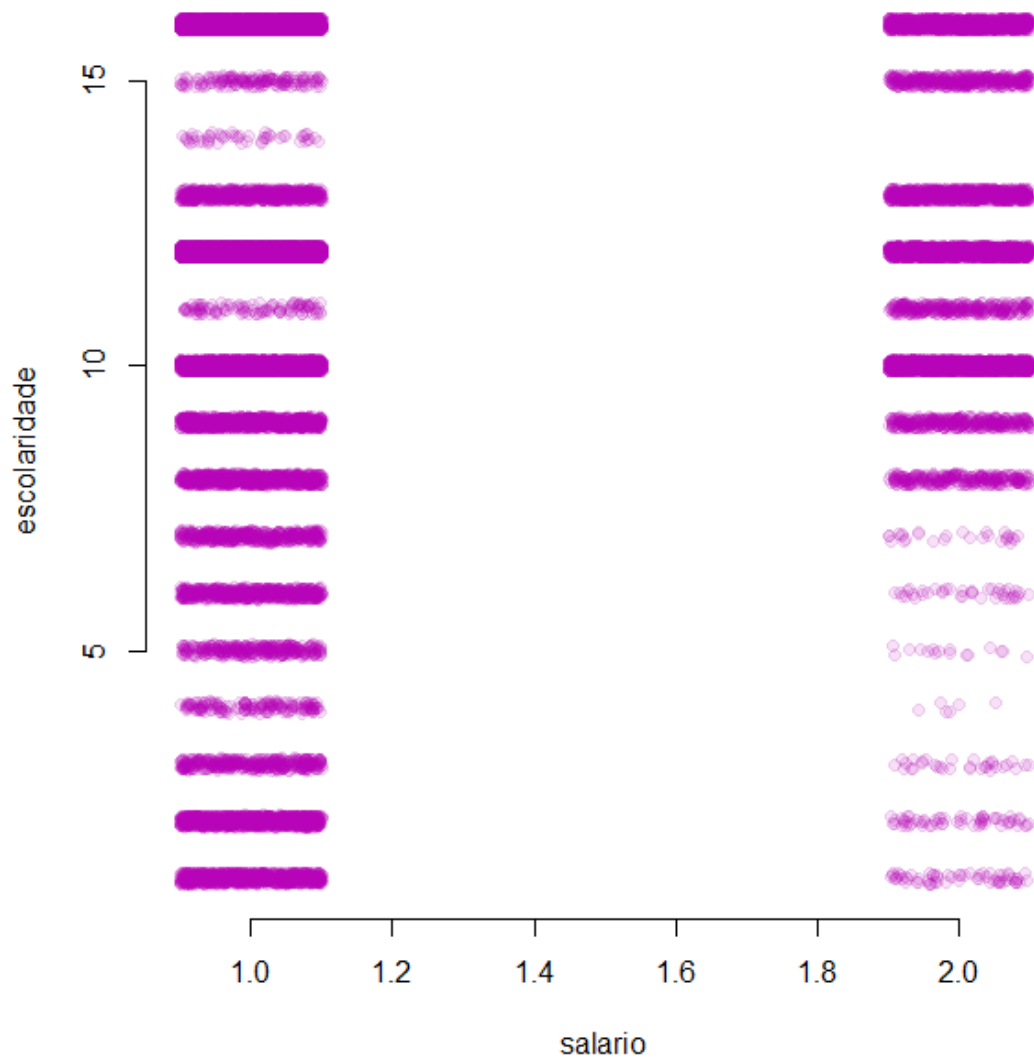
As horas trabalhadas também influenciaram no salário, quanto mais horas mais a chance de se receber mais de 50 mil. A variável explicativa é o número de horas trabalhadas e a variável resposta o salário.



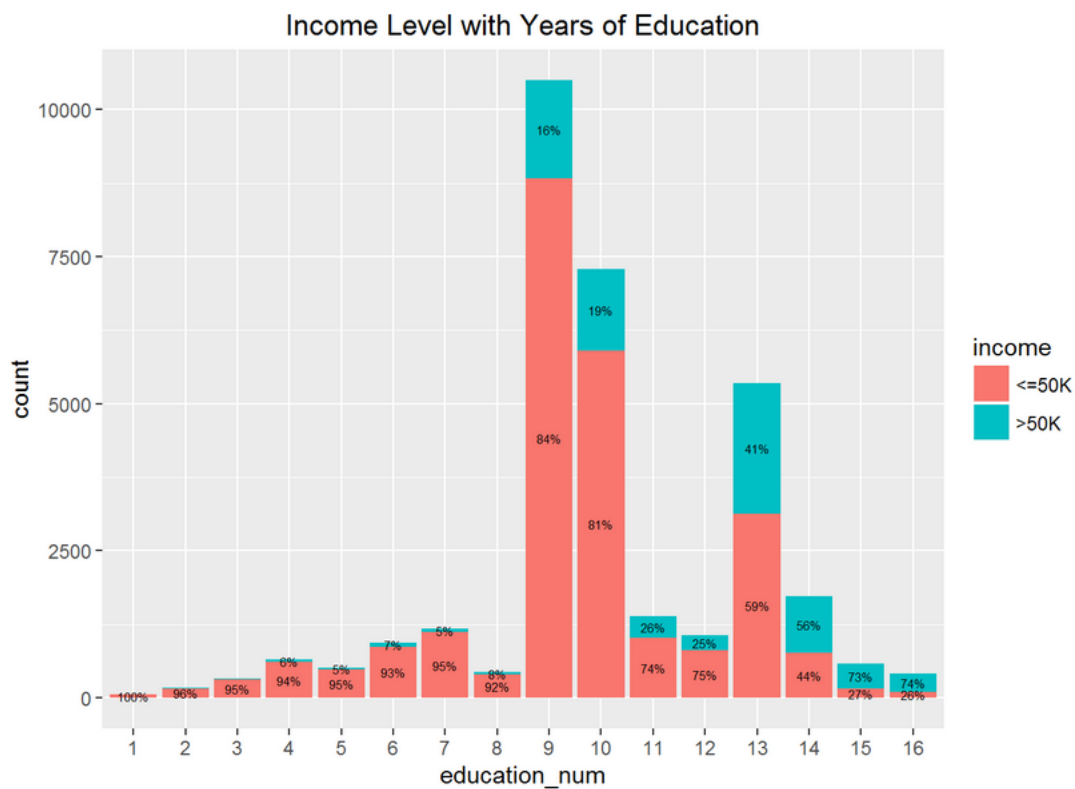
Em relação ao tipo de emprego também existe correlação com o salário, nesse gráfico enumerei cada uma das classes (1= "Federal-gov", 2= "Local-gov", 3= "Private", 4= "Self-emp-inc", 5= "Self-emp-not-inc", 6= "State-gov", 7= "Without-pay"), sendo usadas como variáveis explicativas, e o salário como variável resposta. A maior proporção de pessoas recebendo acima de 50 mil é a de autônomos (self-employed) com 37% das pessoas recebendo acima de 50 mil



Em relação a escolaridade, podemos perceber que quanto maior o nível de escolaridade, existe a tendência de o salário também ser maior.



A categoria 16 (equivalente ao título de doutor) possui a maior proporção de pessoas recebendo acima de 50 mil por ano (74%)



Etnia também é uma variável explicativa, pessoas brancas e asiáticas tem maior tendencia a receber valores acima de 50 mil ao ano.

