

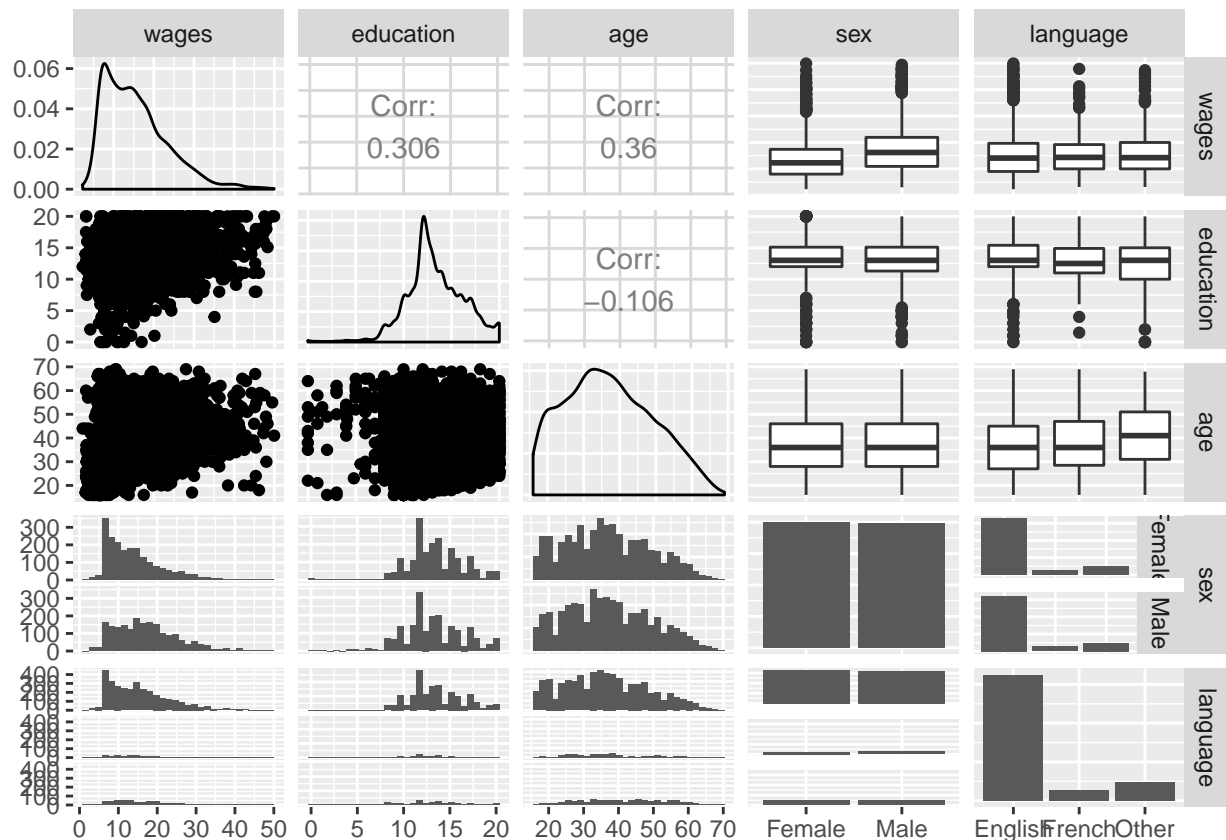
Exercise 1 - Linear models for Gaussian data

Anders Christiansen Sørby, Edvard Hove

We will work on the dataset `carData`, which consists of 3987 observations on the following 5 variables:

- **wages**, composite hourly wage rate from all jobs
- **education**, number of years in schooling
- **age**, in years
- **sex**, Male or Female
- **language**, English, French or Other

Using the function `ggpairs` we get the following diagnostic plot matrix:



The plot suggests that **wages** increase with both **education** and **age**, despite the apparent slight decrease of **education** with **age**. It also seems like males receive higher wages than females, despite similar levels of age and education. There is some relationship between **language**