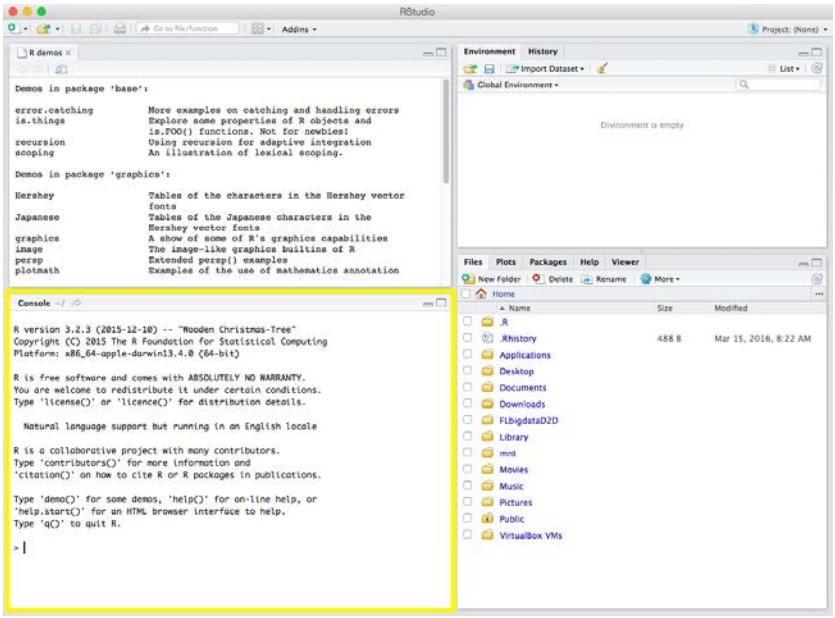
Instructions: Housing data dimension reduction

In this exercise you will use RStudio and H2O to explore data from the AER package and build a principle component analysis model (PCA).

1) Open RStudio. Enter each command in the steps below, one line at a time, into your RStudio console.



a) Load R packages. Each time we open RStudio we need to load our packages for that session.

```
library(AER)
library(h2o)
```

b) Start the H2O server locally.

```
localH2o = h2o.init(ip = "127.0.0.1", port = 54321)
```

c) Load the house prices data provided in the AER library.

```
data("HousePrices")
```

d) Inspect the data frame.

```
HousePrices
```

2) Prepare the house prices data for our PCA model.

a) Load the house prices data into the variable 'housing_data'.

```
housing_data <- as.h2o(HousePrices)
```

b) Load the response data into the variable 'response'.

```
response <- housing_data[,1]
```

c) Create a dataset 'covariates' that excludes the response variable.

```
covariates <- housing_data[,-1]
```

3) Build the PCA model.

```
pca_model <- h2o.prcomp(training_frame = covariates,
                        k=11,
                        max_iterations = 1000,
                        transform="STANDARDIZE",
                        pca_method="GramSVD"
                        )
```

4) Examine the model.

```
summary(pca_model)
```

How many of the principal components would you use to build the predictive model for price?

The Rotation in the H2O Flow output shows the composition of the principal components. What does the composition suggest to you about important features?

