Cars.com (part 1)

XX ADD NAMES HERE

Here is sample code that reads in your data, displays some summary statistics and creates some multivariate graphics. You will be editing this file to tell an interesting story with your data (then publish the results to Rpubs).

PART 1: A SUMMARY OF WHAT WE FOUND:

SOLUTION:

```
# be sure to upload cars.csv to RStudio in the same folder as this file
ds <- read.csv("cars-example.csv")
ds <- ds %>%
  mutate(price = readr::parse_number(price))
names(ds)
```

```
## [1] "car"      "model"    "price"    "year"     "mileage"  "location"
```

```
glimpse(ds)
```

```
## Observations: 40
## Variables: 6
## $ car      <fctr> Toyota Prius, Toyota Prius, Toyota Prius, Toyota Pri...
## $ model    <fctr> Two, Two, Four, Four, Five, Two, Two, Five, Two, Two...
## $ price    <dbl> 21950, 20887, 19998, 18800, 18499, 17999, 17979, 1699...
## $ year     <int> 2016, 2016, 2013, 2014, 2015, 2016, 2016, 2013, 2015,...
## $ mileage  <int> 6255, 9997, 30322, 46219, 31579, 21740, 18295, 38045,...
## $ location <fctr> Atlanta, Atlanta, Atlanta, Atlanta, Atlanta, Atlanta...
```

```
# summary statistics (please delete or comment out lines that you are not using)
tally(~ car, data=ds)
```

```
## car
## Toyota Prius
##           40
```

```
tally(~ model, data=ds)
```

```
## model
##              Five      Four     Three Three SE       Two
##         3         6         5        10         1        15
```

```
tally(~ year, data=ds)
```

```
## year
## 2011 2012 2013 2014 2015 2016
##    1    4   14    8    9    4
```

```
tally(~ location, data=ds)
```

```
## location
## Atlanta
##      40
```

```
favstats(~ price, data=ds)
```

```
##    min    Q1 median    Q3   max     mean       sd  n missing
##  13999 14596  14981 15999 21950 15743.92 1965.711 40       0
```

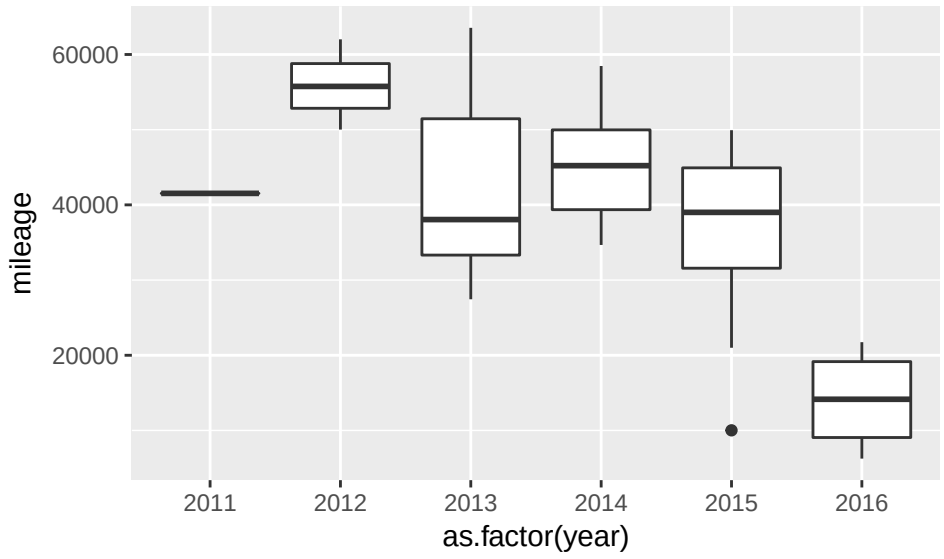
```
favstats(~ mileage, data=ds)
```

```
##   min       Q1 median    Q3   max     mean       sd  n missing
##  6255 32688.25  40002 49950 63546 40057.38 14701.67 40       0
```

PART 2: AN INTERESTING PLOT (WITH INTERPRETATION)

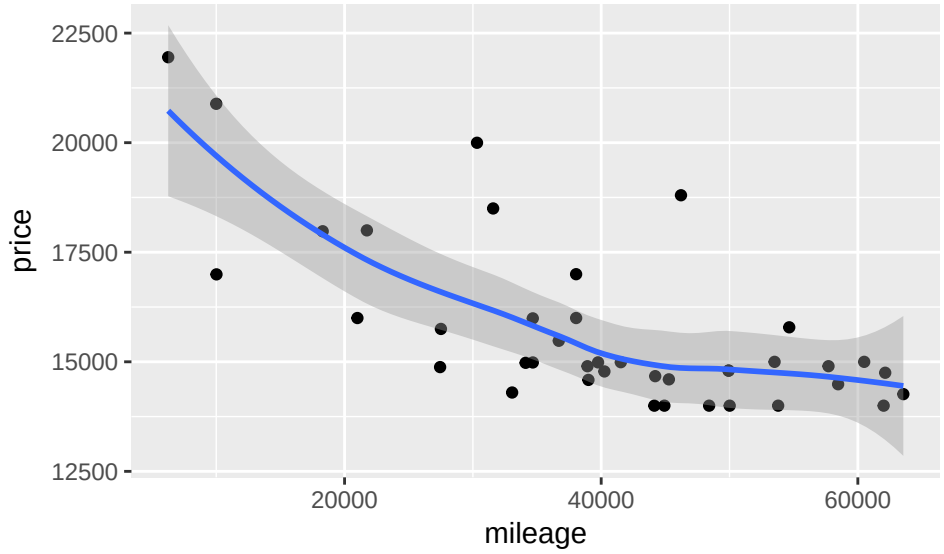
```
# here are some of my ideas: please explore and elaborate using some different plots
# please delete or comment out lines that you are not using
# hint: use the | or group= operator
gf_boxplot(price ~ as.factor(year), data=ds)
```


```
gf_boxplot(mileage ~ as.factor(year), data=ds)
```




```
gf_point(price ~ mileage, data=ds) %>%
  gf_smooth()
```

```
## `geom_smooth()` using method = 'loess'
```




```
gf_point(price ~ year, data=ds) %>%
  gf_smooth()
```

```
## `geom_smooth()` using method = 'loess'
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 2014
```

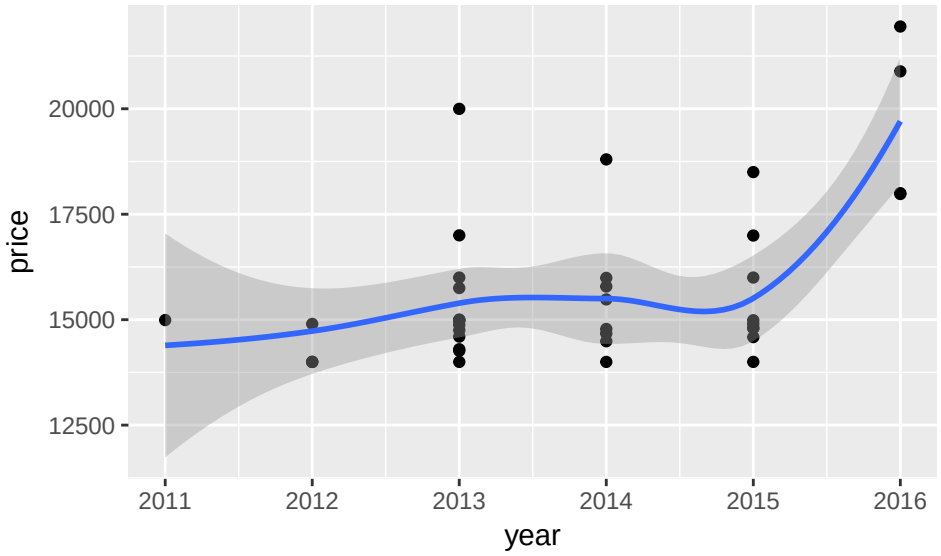
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 2014
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 1
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 0
```



SOLUTION:

```
# your new plot goes here
```

PART 3: RESULTS FROM YOUR MULTIPLE REGRESSION MODEL

along with interpretation of the coefficients

```
# here's my example (with just one predictor)
lm1 <- lm(price ~ year, data=ds)
coef(lm1)
```

```
##   (Intercept)          year
## -1561842.4679      783.3878
```

```
# your model should have at least two predictors
```

SOLUTION: