

# Solutions APPM2720 HW5

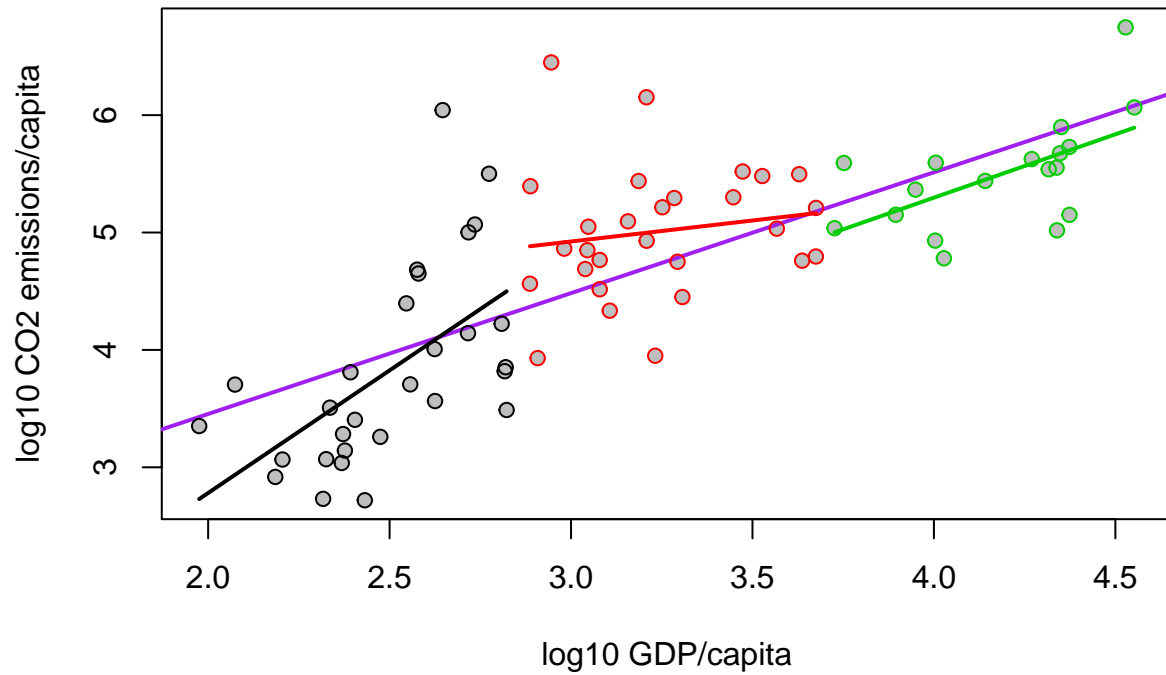
2/15/2017

Repeat the line fitting from the Audi A4 example for the CO2 data and GDP per cap in the **WorldBankCO2** data set.

```
data( WorldBankCO2)
x<- log10(WorldBankCO2$GDP.cap)
y<- log10(WorldBankCO2$CO2.cap)

plot( x,y, pch=16, col="grey", cex=1.0,
      xlab="log10 GDP/capita",
      ylab="log10 CO2 emissions/capita")
title("75 Most Populous Countries
      piecewise linear fits")
fit0<- lm( y~ x)
abline( fit0, col="purple", lwd=2)
nCut<- 3
groupID<- as.numeric(cut( x,nCut))
nGroup<- max( groupID)
for( k in 1: nGroup){
  ind<- groupID == k
  fit<- lm( y[ind]~ x[ind])
  # way in class: abline( fit, col=k, lwd=2)
  # Extra credit: just draw lines where they are fit to
  # the subsets of data.
  xr<- range( x[ind]) # min and max of x's in group
  # value of lines at these endpoints
  yr<- fit$coefficients[1] + xr*fit$coefficients[2]
  lines(xr, yr, col=k, lwd=2)
  points( x[ind], y[ind], col=k)
}
```

## 75 Most Populous Countries piecewise linear fits



*How many cuts do you suggest to best represent the relationship?*

I think 2 or 3 cuts seems adequate especially if the 2 cut solution cuts the data at around 2.8.

E.g.

```
cut(x, breaks=c( min(x), 2.8, max(x) ),
     include.lowest=TRUE)
```

But clearly a single linear fit is not appropriate.

# 75 Most Populous Countries 2 piecewise linear fits

