Midterm 2

Data

For this midterm you need to use two datasets:

"chinalanduse_MODIS_2012.nc" contains four layers with land cover data for China. The data were derived from MODIS satellite data for the year 2012. Each layer gives the fraction of the grid cell that has a specific land cover type: urban (layer 1), cropland (layer 2), grassland (layer 3) and forest (layer 4).

"ch_adm.*" with polygons for the provinces of China.

Q1. Read in the land use data as a SpatRaster get the polygons as a SpatVector (2 points)

```
#load terra package and set working directory
library(terra)
```

```
## terra 1.7.9
```

```
##
## Attaching package: 'terra'
```

```
## The following object is masked from 'package:knitr':
##
##     spin
```

```
setwd("C:/Users/Benny Panjaitan/Documents/GitHub/esp106-Naomi/midterm2")

#store Land Use data as spatial raster and Province data as spatial vector
landuse <- rast("chinalanduse_MODIS_2012.nc")
province <- vect("chn_adm.shp")
```

```
#checking the Province vector data
head(province)
```

```
##       NAME_1   GID_1 GID_0 COUNTRY VARNAME_1 NL_NAME_1    TYPE_1    ENGTYPE_1
## 1      Anhui CHN.1_1   CHN   China     Ānhuī 安徽|安徽     Shěng     Province
## 2    Beijing CHN.2_1   CHN   China   Běijīng 北京|北京 Zhíxiáshì Municipality
## 3 Chongqing CHN.3_1   CHN   China Chóngqìng 重慶|重庆 Zhíxiáshì Municipality
## 4     Fujian CHN.4_1   CHN   China    Fújiàn      福建     Shěng     Province
## 5      Gansu CHN.5_1   CHN   China     Gānsù 甘肅|甘肃     Shěng     Province
## 6 Guangdong CHN.6_1   CHN   China Guǎngdōng 廣東|广东     Shěng     Province
##   CC_1 HASC_1 ISO_1
## 1 <NA>  CN.AH CN-AH
## 2 <NA>  CN.BJ CN-BJ
## 3 <NA>  CN.CQ CN-CQ
## 4 <NA>  CN.FJ CN-FJ
## 5 <NA>  CN.GS CN-GS
## 6 <NA>  CN.GD CN-GD
```

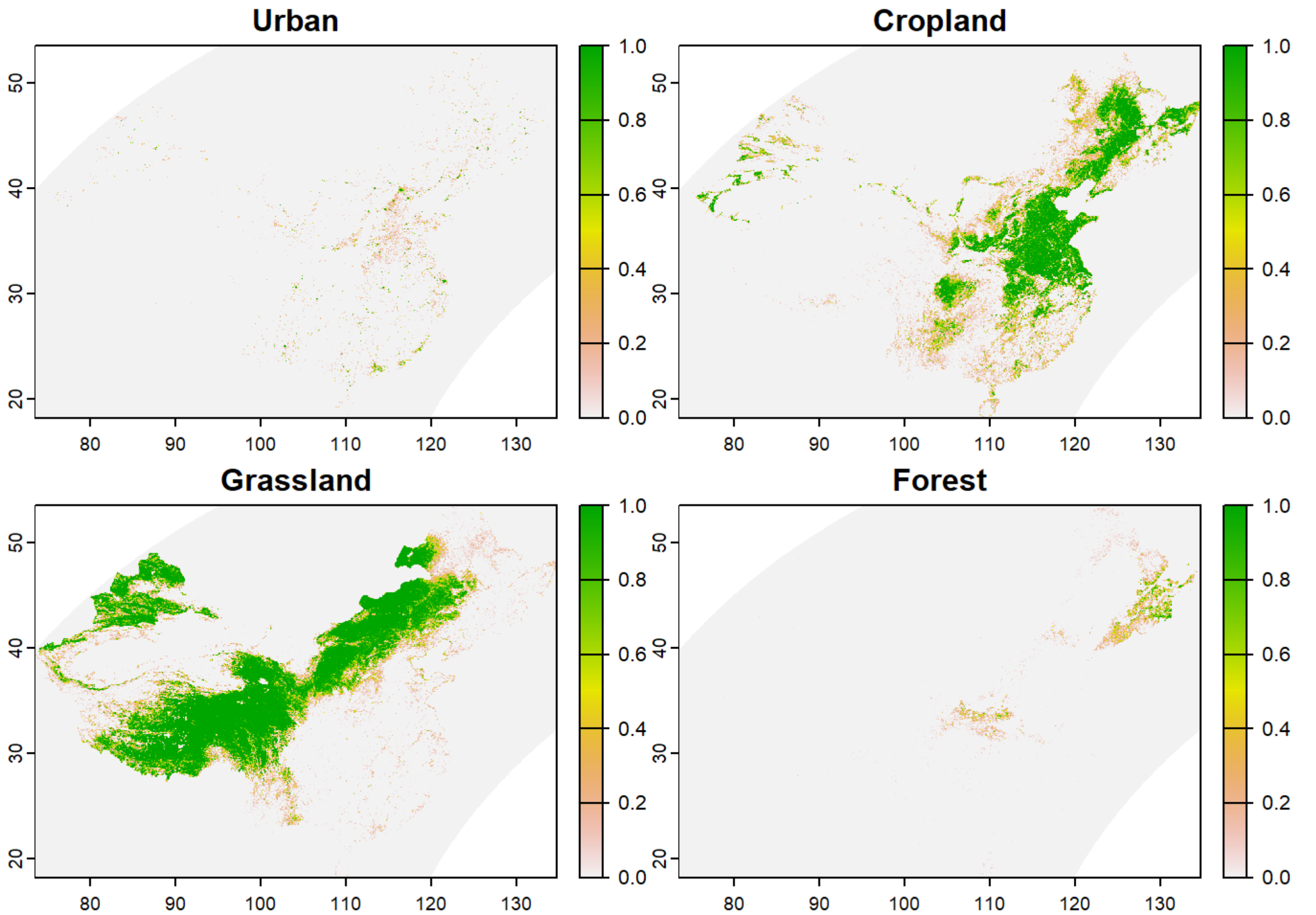
Q2a. Crop the land use SpatRaster to the same extent as the SpatVector of Chinese provinces (1 point), and set all grid cells outside of China to `NA`

```
#cropping the Land Use data according to the Province spatial range
e <- ext(province)
chlu <- crop(landuse, e, snap="in", extend=FALSE)
```

Q2b. Rename the layers in the SpatRaster so they provide information about what data is in each of the 4 layers (2 points)

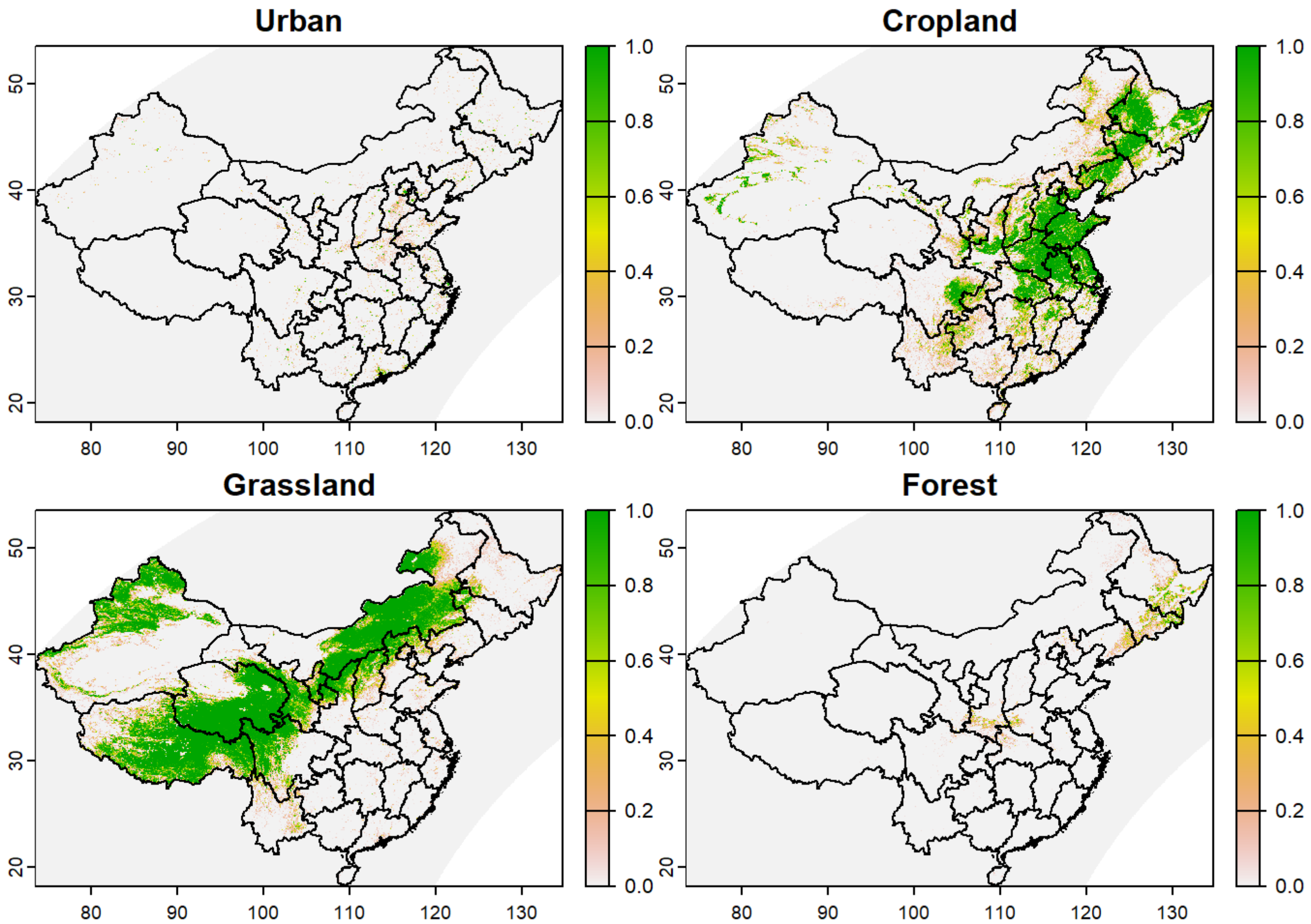
```
#renaming Land Use layers
names(chlu) <- c("Urban", "Cropland", "Grassland", "Forest")

#plotting to check
plot (chlu)
```



Q3. Make a figure showing each SpatRaster layer on one of the panels and overlay the polygons of the Chinese provinces. Title each panel with the type of land use it shows. (4 points)

```
#plotting the Land Use classification by China provinces map
plot(chlu, fun=function() lines(province))
```



Q4a. Use `terra::extract` to find the fraction of each province in each of the four land use classes. [For this question you can assume all the grid cells have the same size] (3 points)

```
#calculating the fraction of each land use type in every provinces
china.df <- extract(chlu, province, fun="mean")
china.df$ID <- province$NAME_1
head(china.df)
```

```
##          ID       Urban   Cropland   Grassland       Forest
## 1     Anhui 0.014851682 0.64640573 0.003847900 1.387219e-03
## 2   Beijing 0.115371393 0.29972989 0.144125230 3.093923e-03
## 3 Chongqing 0.006521529 0.20318474 0.001481789 5.244315e-03
## 4    Fujian 0.026005605 0.06143423 0.004364550 3.092684e-05
## 5     Gansu 0.003996819 0.08768062 0.400894396 3.527299e-03
## 6 Guangdong 0.042238786 0.11982247 0.009257352 1.260333e-04
```

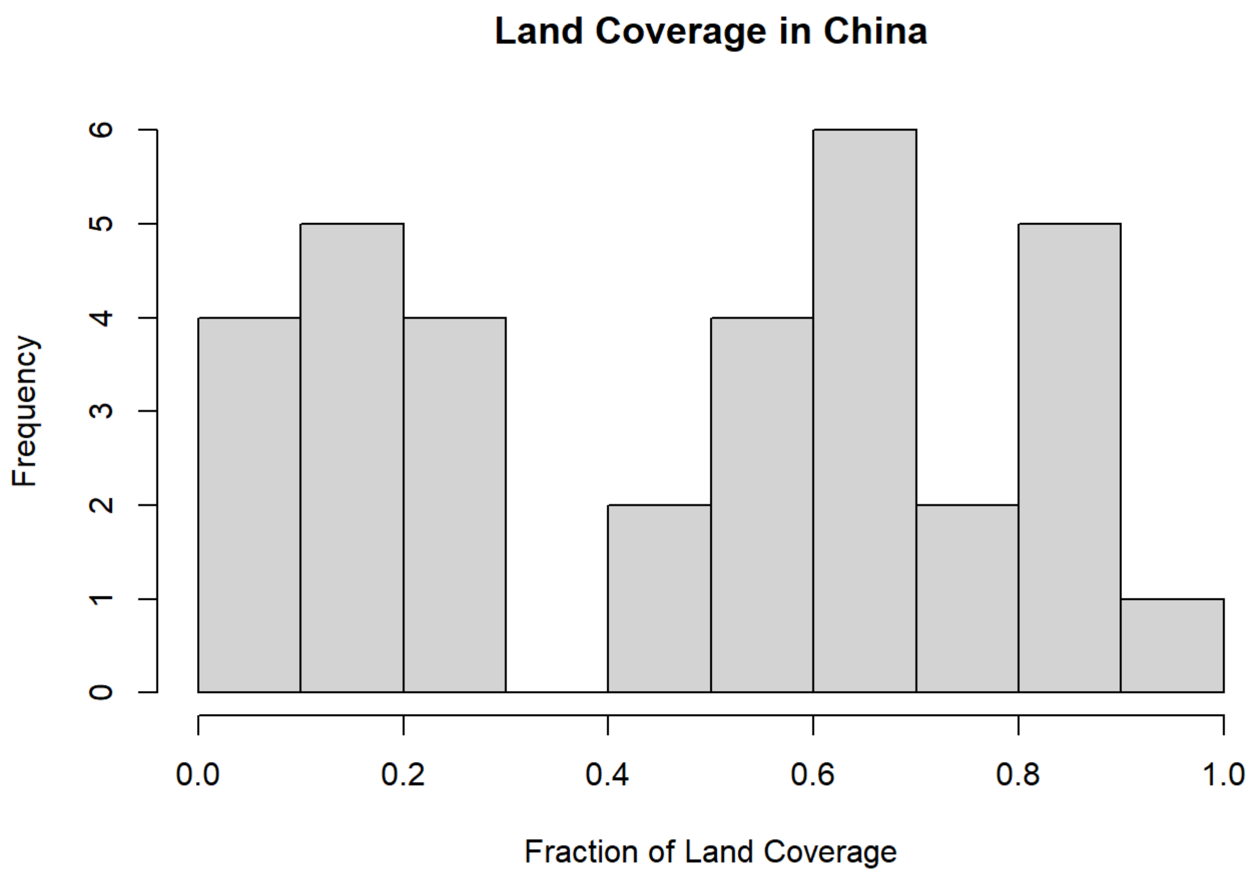
Q4b. Describe the potential problem with the area assumption made in 4a. How might it affect the calculation in that step? What could we do if we didn't want to make that assumption? (You don't have to do it, just describe in theory) (2 points)

Answer: Because all grid cells are assumed to have same size, they may be overlapping in each province's sizes. This may make the fractions not accurate. The alternative way other than making that assumption is by computing the size of grid cells by terra::area function.

Q4c. Sum up the fractions in the four land cover classes for each province and plot these as a histogram. (2 points)

```
#calculating the total of land coverage specified by 2012 MODIS satellite data
china.df$Total <- rowSums(china.df[,2:5])

#plotting the frequencies of land coverage in China
hist(china.df$Total, main="Land Coverage in China", xlab="Fraction of Land Coverage")
```



Q5. Add a new variable called “other” to the data.frame created with terra::extract. This variable should represent the fraction of all other land cover classes. Assign it the appropriate values. (2 points)

```
#Calculating the "Other" type of land than 4 land use type defined
china.df$Other <- 1-china.df$Total
head(china.df)
```

```
##          ID       Urban   Cropland   Grassland       Forest      Total
## 1     Anhui 0.014851682 0.64640573 0.003847900 1.387219e-03 0.66649253
## 2   Beijing 0.115371393 0.29972989 0.144125230 3.093923e-03 0.56232044
## 3 Chongqing 0.006521529 0.20318474 0.001481789 5.244315e-03 0.21643237
## 4    Fujian 0.026005605 0.06143423 0.004364550 3.092684e-05 0.09183531
## 5     Gansu 0.003996819 0.08768062 0.400894396 3.527299e-03 0.49609914
## 6 Guangdong 0.042238786 0.11982247 0.009257352 1.260333e-04 0.17144464
##       Other
## 1 0.3335075
## 2 0.4376796
## 3 0.7835676
## 4 0.9081647
## 5 0.5039009
## 6 0.8285554
```

Q6. Make barplots showing the breakdown of urban, cropland, grassland, forest, and other for each province. The barplots should be “stacked” (a single bar for each province, showing land cover with a color) and “horizontal” (province names on the vertical axis).

Q6a) Use graphics::barplot, make sure to include a legend. (4 points)

First, I prepare the matrix for barplotting using R base

```
#removing the "Total" column and renaming the columns' name
china.df <- subset(china.df, select=-c(Total))
colnames(china.df) <- c("Province", "Urban", "Cropland", "Grassland", "Forest", "Other")

#convert the data frame into matrix for barplotting and renaming the first column's name
china.mat <- as.matrix(subset(china.df, select=-c(Province)))
rownames(china.mat) <- china.df$Province

#checking the matrix
head(china.mat)
```

```
##                 Urban   Cropland   Grassland       Forest     Other
## Anhui     0.014851682 0.64640573 0.003847900 1.387219e-03 0.3335075
## Beijing   0.115371393 0.29972989 0.144125230 3.093923e-03 0.4376796
## Chongqing 0.006521529 0.20318474 0.001481789 5.244315e-03 0.7835676
## Fujian    0.026005605 0.06143423 0.004364550 3.092684e-05 0.9081647
## Gansu     0.003996819 0.08768062 0.400894396 3.527299e-03 0.5039009
## Guangdong 0.042238786 0.11982247 0.009257352 1.260333e-04 0.8285554
```

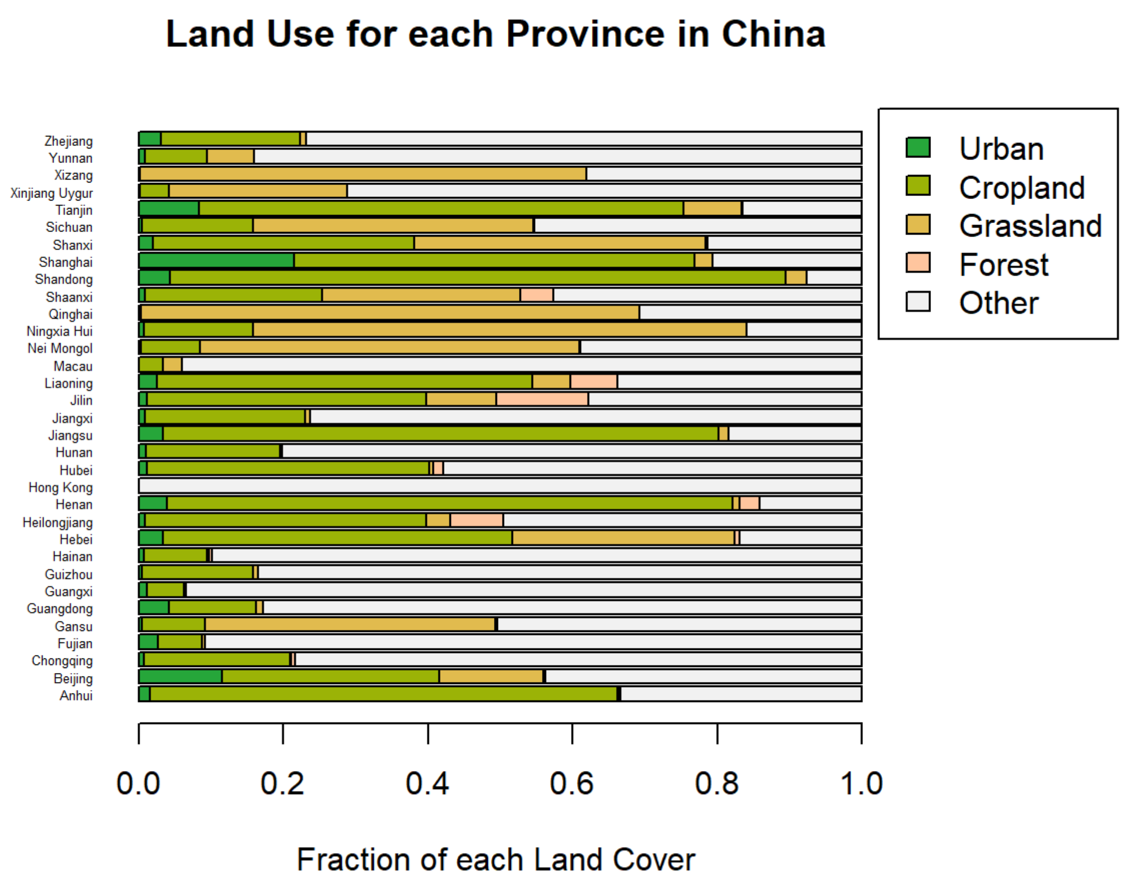
Then, I plot the matrix into barplot.

```
#barplotting using base R
library(colorspace)
```

```
##
## Attaching package: 'colorspace'
```

```
## The following object is masked from 'package:terra':
##
##     RGB
```

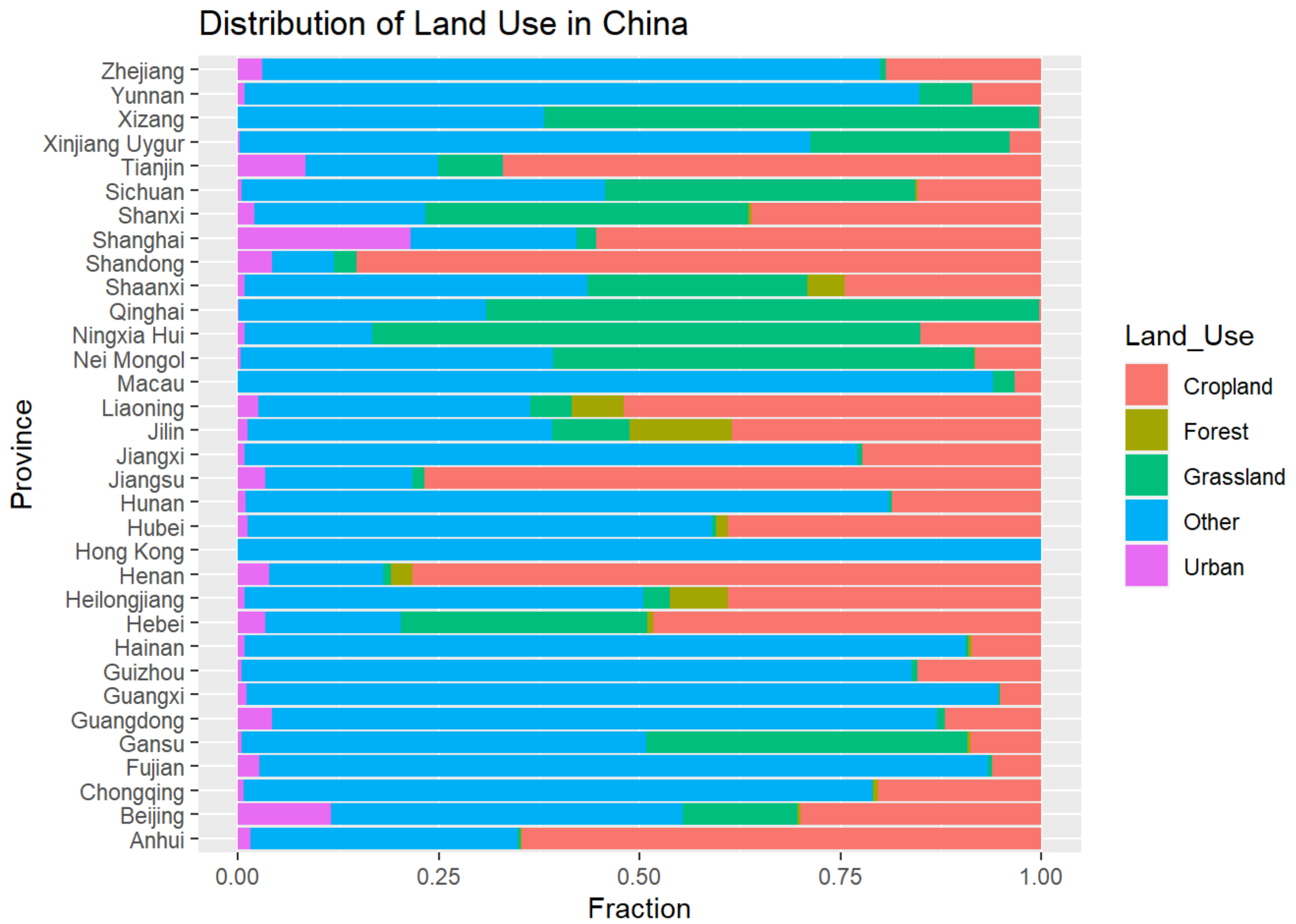
```
par(mar=c(5, 8, 4, 8), xpd=TRUE)
barplot(t(china.mat),
        main="Land Use for each Province in China",
        xlab="Fraction of each Land Cover",
        cex.names=0.4,
        legend.text=T,
        args.legend = list(x = "topright", inset = c(-0.35, 0)),
        beside=F,
        horiz=T,
        las=1,
        col=terrain_hcl(5)
        )
```



Q6b) Use ggplot. (4 points)

```
ch.df <- tidyr::pivot_longer(china.df, cols=2:6, values_to="Fraction", names_to="Land_Use")

library(ggplot2)
ggplot(ch.df, aes(x=Province, y=Fraction, fill=Land_Use)) +
  geom_bar(position="stack", stat="identity")  +
  coord_flip() +
  labs(title="Distribution of Land Use in China")
```



Q7. Upload your R markdown file, and your knitted output to Canvas. Push the R markdown file to your Github repository. (2 points)