Instructions

- Here is the list of packages you can use for the exam:

```
> search()
 [1] ".GlobalEnv"        "package:writexl"   "package:readxl"
 [4] "tools:rstudio"     "package:stats"     "package:graphics"
 [7] "package:grDevices" "package:utils"     "package:datasets"
[10] "package:methods"   "Autoloads"         "package:base"
```

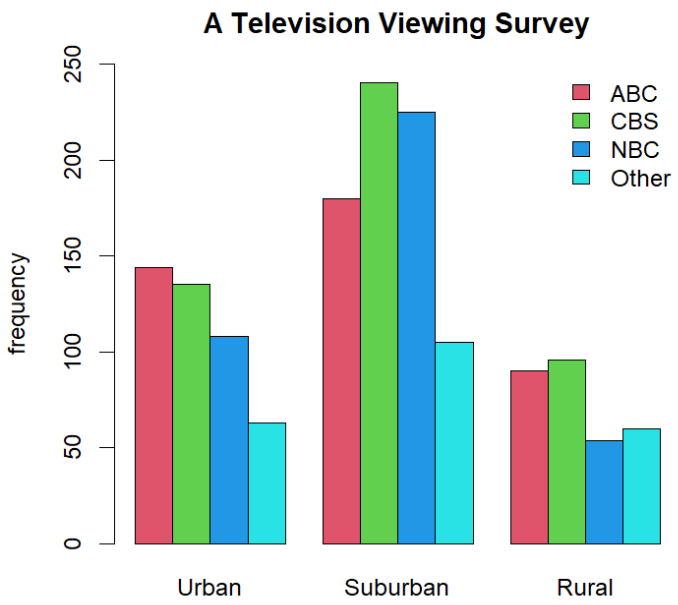
- Submit the file (r-script) on Canvas, Assignments, Midterm: Take-home exam

Problems (10 points each)

1. (10 points) A television viewing survey was conducted on 1,500 individuals. Each individual surveyed was asked to state his or her place of residence and network preference for national news. The results of the survey are shown in the table.

Network Preference	Residence		
	Urban	Suburban	Rural
ABC	144	180	90
CBS	135	240	96
NBC	108	225	54
Other	63	105	60

File **survey.xlsx** contains the data.
Construct a comparative bar graph that shows the **counts** of viewers for each network preference ABC, CBS, NBC, and Other across all places of residence Urban, Suburban, and Rural. The type of information is often used by advertisers to determine which networks' programs they will place their commercials in.
Your plot should be similar to the provided one.
When it comes to the color options, I used $col = 2{:}5$. However, I do not mind if you choose another color palette.



2. (10 points) University officials periodically review the distribution of undergraduate majors within the colleges of the university to help to determine a fair allocation of resources to departments within the colleges. At one such review, the following data were obtained:

College	Number of Majors
Agriculture (Agr)	1,500
Arts and Science (A&S)	11,000
Business Administration (BA)	7,000
Education (Edu)	2,000
Engineering (Engr)	5,000

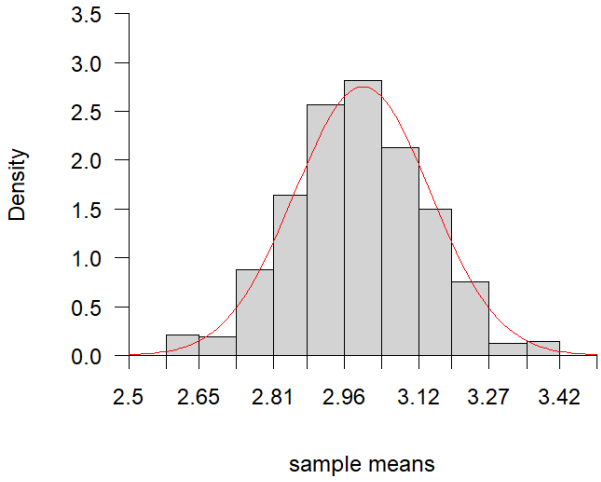
Construct a pie chart for these data. The label for each category should contain the corresponding percentage rounded to one decimal place. For example, "Edu, 28%".

3. (10 points) According to the Central Limit Theorem: For any random sample $X_1, X_2, \ldots, X_m$ from a distribution with mean μ and standard deviation σ and large m $(m > 40)$, $\bar{X}$ has approximately $N\left(\mu, \frac{\sigma^2}{m}\right)$.

   a. Generate 200 random samples from the binomial distribution with the following parameters: the number of trails is 10, the probability of a success in a single trial is .3. The size of each sample is $m = 100$.

   b. Find the sample mean for each of the 200 samples.

   c. Construct a density histogram for $\bar{x}$'s from b. **Use $\sqrt{200} \approx$ 14 classes**. Some of the 14 bars will not be shown on the plot as the corresponding densities are too low. Make sure that your graph looks neat. That is, the lengths of the axes are sufficient, the axes are labeled.

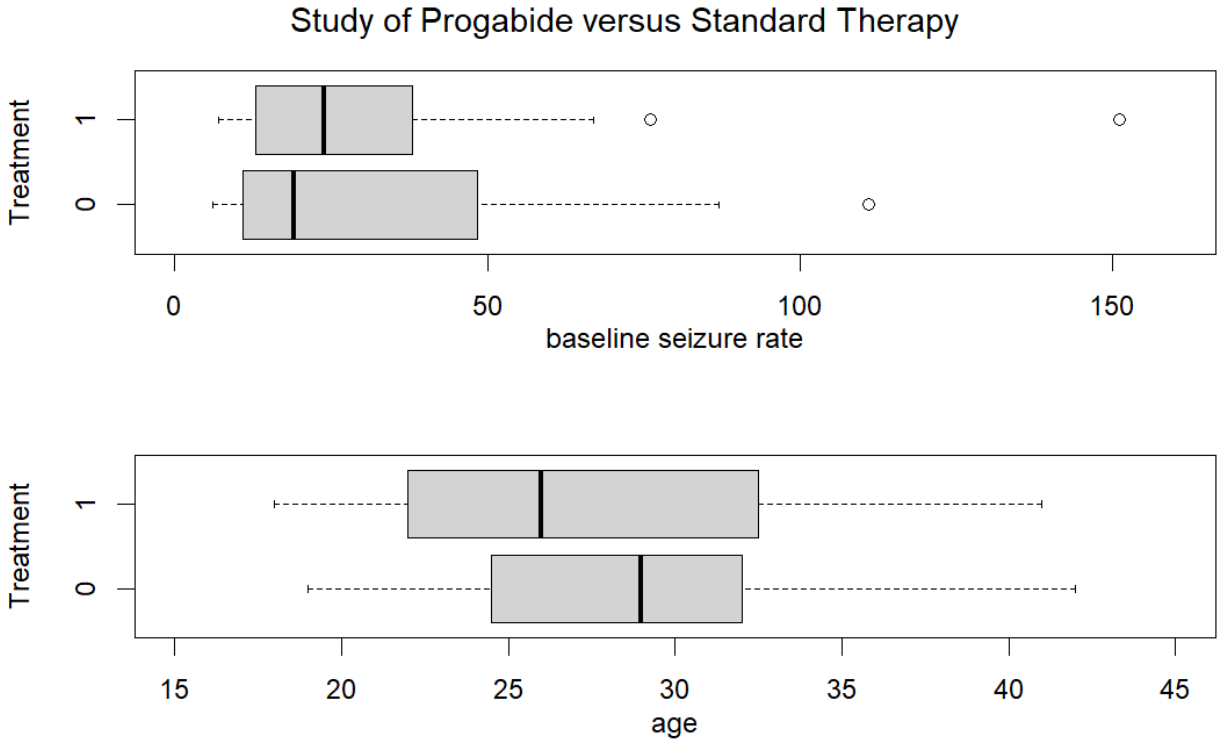


   d. Add the normal curve (density curve) to the histogram with the mean $\mu = 10(.3) = 3$ and the standard deviation $\frac{\sigma}{\sqrt{m}} = \sqrt{2.1/100}$

4. (10 points) Thall and Vail (1990) described a study to evaluate the effectiveness of the anti-epileptic drug **progabide** as an adjuvant to standard therapy. A group of 59 patients was selected to participate in the clinical trial. The patients were randomly assigned to receive either progabide or a placebo. At each of four successive clinical visits, the number of seizures occurring over the previous 2 weeks was reported. The measured variables were
   - $y_i, i = 1, 2, 3, 4 -$ the seizure counts recorded at four clinical visits,
   - $Trt - 0$ is the placebo,
   - $Trt - 1$ is progabide,
   - Base, the baseline seizure rate,
   - Age, the patients age in years.

   The data are given in the file Epilepsy.xlsx.

   - Draw comparative boxplots to compare base number of seizures in the placebo and the progabide groups.
   - Draw comparative boxplots to compare the ages of the participants assigned to the placebo and the progabide groups.
   - Combine both plots to obtain a plot similar to the following one. Please pay attention to the details.



   - Comment on any interesting features you observe (outliers, spreads, medians).