8. A statistics course had two lecture sections, taught by the same instructor. One section used a typical variety of examples from different subject areas, while the other section used only sports-themed examples. The lecture sections were labelled as such, so that a student knew which type of section they were enrolling in, and could choose the one they preferred. The sections are labelled Regular and Sports in the data file. The sports-themed section had its lectures earlier in the day than the regular section. The data are in https://www.utsc.utoronto.ca/~butler/c32/SportsExamples.csv. For each student, the total number of points earned in the course was recorded, as well as which section they were in and their letter grade.

A)(a) (4 marks) Read the data into SAS, and display the mean TotalPoints for each lecture section.

filename myurl url "https://www.utsc.utoronto.ca/~butler/c32/SportsExamples.csv";

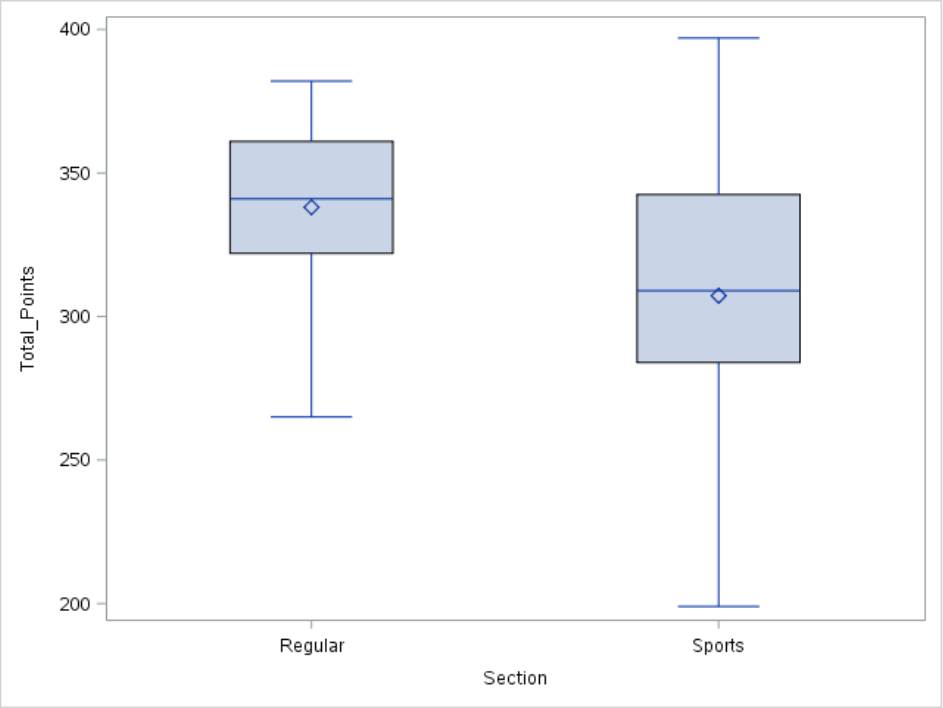
```
proc import
   datafile=myurl
   dbms=csv
   out=mydata
   replace;
   getnames=yes;

proc means;
   var Total_Points;
   class Section;
```

The MEANS Procedure

Analysis Variable : Total_Points						
Section	N Obs	N	Mean	Std Dev	Minimum	Maximum
Regular	29	29	338.0689655	30.6045410	265.0000000	382.0000000
Sports	28	28	307.2500000	52.5107573	199.0000000	397.0000000

(b) (3 marks) Obtain a boxplot of total points for each section. Does one of the sections have a clearly higher average than the other? Explain briefly.

```
proc sgplot;
   vbox Total_Points / category=Section;
```



Regular Students have a much higher median and mode than Sports. Also the spread in students in sports is much larger than students taking a Regular Class.

(c) (2 marks) What is it about the design of this study that would make you hesitant to say that having only sports examples is a bad idea? Explain briefly.

You are testing two different treatments on two diferent populations. For example, the students taking the Regular courses may have taken because that was the time slot that could be filled. They were not randomly chosen to be in a section, instead the students picked which one they could be in.

9. The “ecological footprint” of a person, city or country is defined by the World Wildlife Federation as . . . the impact of human activities measured in terms of the area of biologically productive land and water required to produce the goods consumed and to assimilate the wastes generated. More simply, it is the amount of the environment necessary to produce the goods and services necessary to support a particular lifestyle. See http://wwf.panda.org/knowledge_hub/teacher_resources/webfieldtrips/ecological_balance/eco_footprint/. The units of an ecological footprint for a country is hectares per capita. Data on 66 countries from the Americas, Europe and Asia (mainly the western part of Asia) are given in http://www.utsc.utoronto.ca/~butler/c32/footprint.txt. There are three columns: the name of the country, with spaces removed, the continent (called Region), with S standing for Asia, and the ecological footprint.

(a) (3 marks) Read the data into SAS and display the first 20 rows.

filename myurl url "http://www.utsc.utoronto.ca/~butler/c32/footprint.txt";

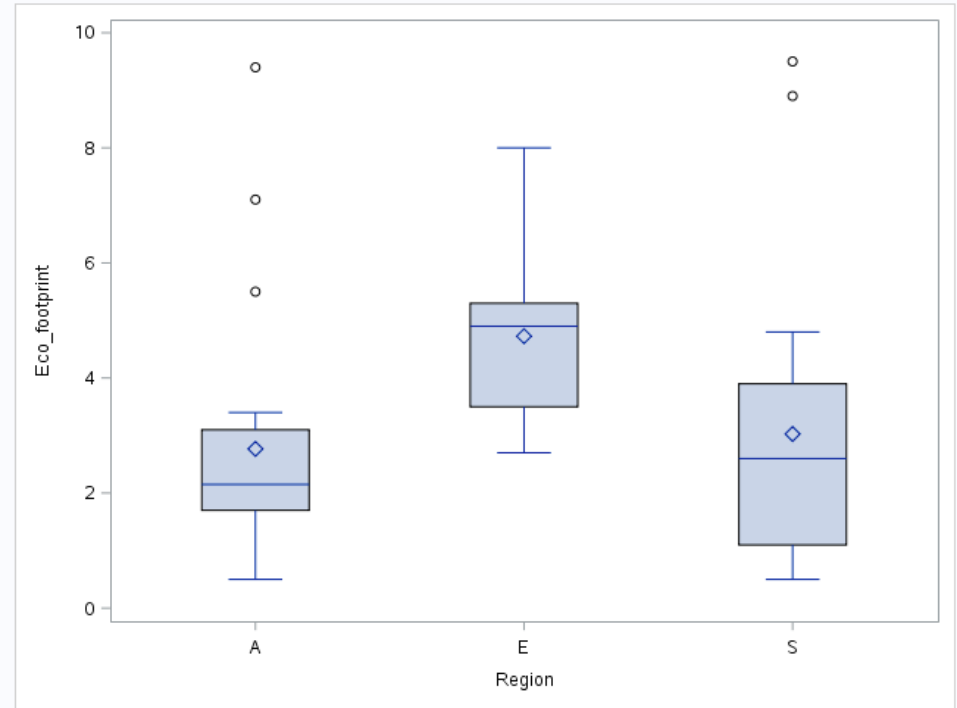
```
proc import
   datafile=myurl
   dbms=dlm
   out=footprint
   replace;
   delimiter=' ';
   getnames=yes;
```

```
proc print data=footprint(obs=20);
```

Obs	Country	Region	Eco_footprint
1	Argentina	A	2.5
2	Bolivia	A	2.1
3	Brazil	A	2.4
4	Chile	A	3
5	Colombia	A	1.8
6	CostaRica	A	2.3
7	Cuba	A	1.8
8	DominicanRep	A	1.5
9	Ecuador	A	2.2
10	ElSalvador	A	1.6
11	Guatemala	A	1.5
12	Haiti	A	0.5
13	Honduras	A	1.8
14	Jamaica	A	1.1
15	Mexico	A	3.4
16	Nicaragua	A	2
17	Panama	A	3.2
18	Paraguay	A	3.2
19	Peru	A	1.6
20	TrinidadToba	A	2.1

(b) (2 marks) Make a suitable graph of the ecological footprint values for each Region. Note that in SAS, variable names are not case-sensitive.

```
proc sgplot;
   vbox Eco_footprint / category=Region;
```



(c) (2 marks) Find the mean and median ecological footprint by region.

```
proc means medians;
   var Eco_footprint;
   class Region;

proc means;
   var Eco_footprint;
   class Region;
```

The MEANS Procedure

Analysis Variable : Eco_footprint						
Region	N Obs	N	Mean	Std Dev	Minimum	Maximum
A	24	24	2.7666667	1.9776285	0.5000000	9.4000000
E	23	23	4.7260870	1.2764359	2.7000000	8.0000000
S	19	19	3.0263158	2.5220895	0.5000000	9.5000000

The MEANS Procedure

Analysis Variable : Eco_footprint		
Region	N Obs	Median
A	24	2.1500000
E	23	4.9000000
S	19	2.6000000

(d) (3 marks) Explain briefly how your graph and your calculations are consistent. (There are a number of different approaches you can take; anything that offers insight into how the graph and the calculations are telling the same story is good.)

We can see from the graph how mean and medians stack up. The line is the median and diamond is the mean. For example, region A has a higher mean than median(2.7 > 2.1 and shown in the graph). E has a higher median than Mean(4.7 < 4.9). and S has a higher mean than media(3.026 > 2.6).