**MINI PROJECT #1**

**Name:** Omkar Nandkumar Dixit (ond170030)

**Name of group members:** Sushrut Patnaik (sxp175331)

**Contributions of each group members:**

We discussed both the questions together, but Omkar coded the first question and Sushrut coded the second question.

**Section 1:**

**Question 1 -**

1. If you see barplot you can easily observe that the numbers of runners from maine are way more than the ones that are not from maine(away), approximately thrice. If you see in the code where we are finding the summary of the runners we get

*(maine.summary = summary(maine))*

*Maine*

*Away :1417*

*Maine:4458*

**Section 2: R Code**

*> # Reading roadrace.csv*

*> roadrace = read.csv("Documents/Fall18/Stats/R\_Mini\_Projects/MP2/roadrace.csv", na.strings = "\*")*

*> summary(roadrace)*

*Place Division.Place Division.Entrants Division Age Sex State.Country Time..seconds. Mile.pace..seconds. From.USA*

*Min. : 1 Min. : 1.0 Min. : 3.0 F3034 : 471 Min. : 7.00 : 1 ME :4458 Min. :1667 Min. : 269.0 No : 74*

*1st Qu.:1470 1st Qu.: 62.0 1st Qu.:235.0 F3539 : 426 1st Qu.:29.00 F:2951 MA : 535 1st Qu.:2987 1st Qu.: 481.0 Yes:5801*

*Median :2938 Median :139.0 Median :333.0 M4044 : 411 Median :39.00 M:2923 NH : 166 Median :3421 Median : 551.0*

*Mean :2938 Mean :156.1 Mean :311.1 F2529 : 397 Mean :38.83 NY : 116 Mean :3486 Mean : 561.6*

*3rd Qu.:4406 3rd Qu.:232.0 3rd Qu.:397.0 F4044 : 394 3rd Qu.:48.00 CT : 78 3rd Qu.:3869 3rd Qu.: 623.0*

*Max. :5875 Max. :471.0 Max. :471.0 M4549 : 357 Max. :86.00 VT : 64 Max. :9130 Max. :1470.0*

*NA's :1 NA's :1 (Other):3419 NA's :1 (Other): 458*

*Maine Time..minutes.*

*Away :1417 Min. : 27.78*

*Maine:4458 1st Qu.: 49.78*

*Median : 57.02*

*Mean : 58.11*

*3rd Qu.: 64.48*

*Max. :152.17*

*>*

*> #Storing the Maine column of the data*

*> a = table(roadrace[["Maine"]])*

*>*

*> # Question 1(a)*

*> # Creating bar graph for variable Maine*

*> barplot(a, main="Question 1(a)", xlab="Maine or Away", ylim = c(0,5000))*

*> maine = subset(roadrace, select = "Maine")*

*> (maine.summary = summary(maine))*

*Maine*

*Away :1417*

*Maine:4458*

*>*

*> # Question 1(b)*

*> # Creating a subset of the values of Column "Maine" to create two separate histograms*

*> b.maine = subset(roadrace, roadrace$Maine=="Maine")*

*> b.away = subset(roadrace, roadrace$Maine=="Away")*

*>*

*> # Storing just the running times (in minutes)*

*> b.maine.time = b.maine$Time..minutes.*

*> b.away.time = b.away$Time..minutes.*

*>*

*> # Plotting Histograms*

*> hist(b.maine.time, main = "Question 1(b)", xlab = "Running Time (Maine)", ylim = c(0,2000), col=rgb(0.1,0.1,0.1,0.5))*

*> hist(b.away.time, main = "Question 1(b)", xlab = "Running Time (Away)", ylim=c(0,2000), add=T, col=rgb(0.8,0.8,0.8,0.5))*

*> box()*

*>*

*>*

*> # Five number summary*

*> (b.maine.time.summary = summary(b.maine.time))*

*Min. 1st Qu. Median Mean 3rd Qu. Max.*

*30.57 50.00 57.03 58.20 64.24 152.17*

*> b.maine.time.q1 = b.maine.time.summary[2]*

*> b.maine.time.q3 = b.maine.time.summary[5]*

*> (b.away.time.summary = summary(b.away.time))*

*Min. 1st Qu. Median Mean 3rd Qu. Max.*

*27.78 49.15 56.92 57.82 64.83 133.71*

*> b.away.time.q1 = b.away.time.summary[2]*

*> b.away.time.q3 = b.away.time.summary[5]*

*>*

*> # mean*

*> (b.maine.time.mean = mean(b.maine.time))*

*[1] 58.19514*

*> (b.away.time.mean = mean(b.away.time))*

*[1] 57.82181*

*>*

*> # standard deviation*

*> (b.maine.time.sd = sd(b.maine.time))*

*[1] 12.18511*

*> (b.away.time.sd = sd(b.away.time))*

*[1] 13.83538*

*>*

*> # range*

*> (b.maine.time.range = range(b.maine.time))*

*[1] 30.567 152.167*

*> (b.away.time.range = range(b.away.time))*

*[1] 27.782 133.710*

*>*

*> # median*

*> (b.maine.time.median = median(b.maine.time))*

*[1] 57.0335*

*> (b.away.time.median = median(b.away.time))*

*[1] 56.92*

*>*

*> # inter-quartile range*

*> (b.maine.time.iqr = IQR(b.maine.time))*

*[1] 14.24775*

*> (b.away.time.iqr = IQR(b.away.time))*

*[1] 15.674*

*>*

*> # Question 1(c)*

*> boxplot(b.maine.time, b.away.time, main="Running Time (Maine and Away)")*

*> #boxplot(b.maine.time, main="Running Time (Maine)")*

*> #boxplot(b.away.time, main="Running Time (Away)")*

*> (b.maine.time.outliers = subset(b.maine, ((b.maine$Time..minutes. < (b.maine.time.q1 - 1.5 \* b.maine.time.iqr) ) | (b.maine$Time..minutes. > (b.maine.time.q3 + 1.5 \* b.maine.time.iqr)))))*

*Place Division.Place Division.Entrants Division Age Sex State.Country Time..seconds. Mile.pace..seconds. From.USA Maine Time..minutes.*

*5685 5685 389 397 F2529 26 F ME 5143.2 828 Yes Maine 85.720*

*5689 5689 409 412 M4044 41 M ME 5164.3 832 Yes Maine 86.072*

*5690 5690 378 394 F4044 40 F ME 5165.2 832 Yes Maine 86.087*

*5694 5694 232 235 M5559 57 M ME 5175.0 833 Yes Maine 86.250*

*5695 5695 452 471 F3034 34 F ME 5175.4 833 Yes Maine 86.257*

*5697 5697 380 394 F4044 44 F ME 5182.3 834 Yes Maine 86.372*

*5698 5698 417 426 F3539 35 F ME 5191.2 836 Yes Maine 86.520*

*5699 5699 128 142 F5559 55 F ME 5197.9 837 Yes Maine 86.632*

*5700 5700 62 72 F6064 63 F ME 5201.9 838 Yes Maine 86.698*

*5701 5701 129 142 F5559 58 F ME 5206.1 838 Yes Maine 86.768*

*5702 5702 233 235 M5559 58 M ME 5206.2 838 Yes Maine 86.770*

*5703 5703 8 14 F7074 71 F ME 5209.0 839 Yes Maine 86.817*

*5704 5704 329 333 M5054 53 M ME 5209.2 839 Yes Maine 86.820*

*5705 5705 330 333 M5054 50 M ME 5209.3 839 Yes Maine 86.822*

*5706 5706 233 237 F2024 23 F ME 5209.4 839 Yes Maine 86.823*

*5707 5707 310 325 F4549 49 F ME 5223.7 841 Yes Maine 87.062*

*5708 5708 130 142 F5559 55 F ME 5225.5 841 Yes Maine 87.092*

*5709 5709 311 325 F4549 46 F ME 5229.2 842 Yes Maine 87.153*

*5710 5710 3 4 M8099 83 M ME 5229.4 842 Yes Maine 87.157*

*5711 5711 381 394 F4044 43 F ME 5232.0 842 Yes Maine 87.200*

*5712 5712 140 144 M6064 63 M ME 5252.2 846 Yes Maine 87.537*

*5713 5713 418 426 F3539 35 F ME 5254.3 846 Yes Maine 87.572*

*5715 5715 58 62 M6569 66 M ME 5264.7 848 Yes Maine 87.745*

*5716 5716 419 426 F3539 35 F ME 5265.0 848 Yes Maine 87.750*

*5718 5718 454 471 F3034 34 F ME 5268.9 848 Yes Maine 87.815*

*5719 5719 312 325 F4549 45 F ME 5270.6 849 Yes Maine 87.843*

*5720 5720 23 29 F6569 66 F ME 5276.5 850 Yes Maine 87.942*

*5721 5721 217 229 F5054 52 F ME 5276.5 850 Yes Maine 87.942*

*5723 5723 351 357 M4549 48 M ME 5279.4 850 Yes Maine 87.990*

*5724 5724 59 62 M6569 67 M ME 5279.8 850 Yes Maine 87.997*

*5729 5729 24 29 F6569 69 F ME 5287.5 851 Yes Maine 88.125*

*5730 5730 383 394 F4044 41 F ME 5290.1 852 Yes Maine 88.168*

*5732 5732 131 142 F5559 55 F ME 5297.3 853 Yes Maine 88.288*

*5733 5733 220 229 F5054 53 F ME 5313.2 856 Yes Maine 88.553*

*5735 5735 384 394 F4044 43 F ME 5318.3 856 Yes Maine 88.638*

*5736 5736 385 394 F4044 42 F ME 5319.0 857 Yes Maine 88.650*

*5737 5737 386 394 F4044 44 F ME 5319.7 857 Yes Maine 88.662*

*5738 5738 457 471 F3034 31 F ME 5324.5 857 Yes Maine 88.742*

*5739 5739 458 471 F3034 32 F ME 5328.5 858 Yes Maine 88.808*

*5740 5740 392 397 F2529 26 F ME 5349.8 861 Yes Maine 89.163*

*5741 5741 393 397 F2529 28 F ME 5350.9 862 Yes Maine 89.182*

*5742 5742 221 229 F5054 53 F ME 5370.1 865 Yes Maine 89.502*

*5743 5743 64 72 F6064 64 F ME 5375.0 866 Yes Maine 89.583*

*5744 5744 459 471 F3034 31 F ME 5383.7 867 Yes Maine 89.728*

*5745 5745 9 14 F7074 70 F ME 5391.3 868 Yes Maine 89.855*

*5747 5747 60 62 M6569 69 M ME 5392.7 868 Yes Maine 89.878*

*5752 5752 314 325 F4549 47 F ME 5400.6 870 Yes Maine 90.010*

*5754 5754 22 23 M7074 74 M ME 5426.0 874 Yes Maine 90.433*

*5758 5758 420 426 F3539 38 F ME 5436.7 875 Yes Maine 90.612*

*5759 5759 352 357 M4549 46 M ME 5439.4 876 Yes Maine 90.657*

*5760 5760 33 33 F0114 13 F ME 5439.8 876 Yes Maine 90.663*

*5761 5761 315 325 F4549 46 F ME 5440.2 876 Yes Maine 90.670*

*5762 5762 333 335 M3539 38 M ME 5440.5 876 Yes Maine 90.675*

*5763 5763 10 14 F7074 70 F ME 5447.4 877 Yes Maine 90.790*

*5764 5764 332 333 M5054 50 M ME 5467.7 880 Yes Maine 91.128*

*5767 5767 133 142 F5559 59 F ME 5496.0 885 Yes Maine 91.600*

*5768 5768 167 174 F1519 15 F ME 5499.5 886 Yes Maine 91.658*

*5769 5769 410 412 M4044 43 M ME 5511.6 887 Yes Maine 91.860*

*5770 5770 317 325 F4549 46 F ME 5511.9 888 Yes Maine 91.865*

*5771 5771 65 72 F6064 64 F ME 5523.9 889 Yes Maine 92.065*

*5772 5772 234 237 F2024 24 F ME 5538.7 892 Yes Maine 92.312*

*5773 5773 388 394 F4044 41 F ME 5540.4 892 Yes Maine 92.340*

*5776 5776 66 72 F6064 61 F ME 5565.3 896 Yes Maine 92.755*

*5777 5777 222 229 F5054 53 F ME 5567.6 897 Yes Maine 92.793*

*5779 5779 223 229 F5054 54 F ME 5592.0 900 Yes Maine 93.200*

*5784 5784 224 229 F5054 52 F ME 5617.9 905 Yes Maine 93.632*

*5786 5786 353 357 M4549 45 M ME 5661.7 912 Yes Maine 94.362*

*5787 5787 462 471 F3034 33 F ME 5665.6 912 Yes Maine 94.427*

*5788 5788 390 394 F4044 41 F ME 5681.5 915 Yes Maine 94.692*

*5789 5789 421 426 F3539 37 F ME 5682.2 915 Yes Maine 94.703*

*5790 5790 135 142 F5559 58 F ME 5701.5 918 Yes Maine 95.025*

*5791 5791 318 325 F4549 47 F ME 5704.2 918 Yes Maine 95.070*

*5792 5792 319 325 F4549 49 F ME 5712.7 920 Yes Maine 95.212*

*5793 5793 354 357 M4549 48 M ME 5712.9 920 Yes Maine 95.215*

*5794 5794 67 72 F6064 61 F ME 5721.6 921 Yes Maine 95.360*

*5796 5796 137 142 F5559 57 F ME 5740.5 924 Yes Maine 95.675*

*5797 5797 463 471 F3034 32 F ME 5761.6 928 Yes Maine 96.027*

*5798 5798 464 471 F3034 30 F ME 5761.8 928 Yes Maine 96.030*

*5799 5799 391 394 F4044 44 F ME 5772.0 929 Yes Maine 96.200*

*5801 5801 170 174 F1519 15 F ME 5784.0 931 Yes Maine 96.400*

*5804 5804 25 29 F6569 65 F ME 5802.6 934 Yes Maine 96.710*

*5805 5805 467 471 F3034 31 F ME 5808.8 935 Yes Maine 96.813*

*5806 5806 422 426 F3539 36 F ME 5808.9 935 Yes Maine 96.815*

*[ reached getOption("max.print") -- omitted 51 rows ]*

*> (b.away.time.outliers = subset(b.away, ((b.away$Time..minutes. < (b.away.time.q1 - 1.5 \* b.away.time.iqr) ) | (b.away$Time..minutes. > (b.away.time.q3 + 1.5 \* b.away.time.iqr)))))*

*Place Division.Place Division.Entrants Division Age Sex State.Country Time..seconds. Mile.pace..seconds. From.USA Maine Time..minutes.*

*5734 5734 456 471 F3034 33 F VA 5317.0 856 Yes Away 88.617*

*5746 5746 387 394 F4044 43 F CA 5391.7 868 Yes Away 89.862*

*5748 5748 460 471 F3034 32 F MD 5392.9 868 Yes Away 89.882*

*5749 5749 141 144 M6064 63 M MA 5393.3 868 Yes Away 89.888*

*5750 5750 57 57 M0114 13 M TX 5397.0 869 Yes Away 89.950*

*5751 5751 313 325 F4549 46 F TX 5397.4 869 Yes Away 89.957*

*5753 5753 12 13 M7579 77 M NH 5405.8 870 Yes Away 90.097*

*5755 5755 234 235 M5559 57 M MI 5426.9 874 Yes Away 90.448*

*5756 5756 132 142 F5559 57 F MI 5427.4 874 Yes Away 90.457*

*5757 5757 3 3 F7579 76 F VA 5435.3 875 Yes Away 90.588*

*5765 5765 2 5 F8099 86 F MO 5485.6 883 Yes Away 91.427*

*5766 5766 316 325 F4549 45 F DE 5493.6 885 Yes Away 91.560*

*5774 5774 168 174 F1519 16 F VT 5540.7 892 Yes Away 92.345*

*5775 5775 169 174 F1519 19 F VT 5541.0 892 Yes Away 92.350*

*5778 5778 134 142 F5559 56 F MD 5586.9 900 Yes Away 93.115*

*5780 5780 389 394 F4044 41 F MA 5596.5 901 Yes Away 93.275*

*5781 5781 11 14 F7074 72 F MA 5609.3 903 Yes Away 93.488*

*5782 5782 23 23 M7074 72 M MA 5609.6 903 Yes Away 93.493*

*5783 5783 315 316 M3034 32 M MA 5614.4 904 Yes Away 93.573*

*5785 5785 461 471 F3034 31 F VT 5645.9 909 Yes Away 94.098*

*5795 5795 136 142 F5559 57 F CO 5738.4 924 Yes Away 95.640*

*5800 5800 61 62 M6569 67 M NY 5779.8 931 Yes Away 96.330*

*5802 5802 465 471 F3034 32 F NH 5793.1 933 Yes Away 96.552*

*5803 5803 466 471 F3034 32 F NH 5793.6 933 Yes Away 96.560*

*5808 5808 225 229 F5054 52 F VA 5817.4 937 Yes Away 96.957*

*5812 5812 320 325 F4549 47 F MD 5904.5 951 Yes Away 98.408*

*5816 5816 138 142 F5559 57 F MA 5918.8 953 Yes Away 98.647*

*5818 5818 139 142 F5559 58 F NH 5933.0 955 Yes Away 98.883*

*5823 5823 13 14 F7074 74 F CA 5969.8 961 Yes Away 99.497*

*5827 5827 393 394 F4044 42 F VT 6033.0 971 Yes Away 100.550*

*5828 5828 355 357 M4549 48 M VT 6033.4 971 Yes Away 100.557*

*5829 5829 322 325 F4549 46 F VT 6035.2 972 Yes Away 100.587*

*5831 5831 335 335 M3539 36 M MA 6084.5 980 Yes Away 101.408*

*5855 5855 470 471 F3034 33 F MA 6527.3 1051 Yes Away 108.788*

*5857 5857 316 316 M3034 30 M NY 6527.5 1051 Yes Away 108.792*

*5858 5858 412 412 M4044 44 M CA 6545.5 1054 Yes Away 109.092*

*5864 5864 228 229 F5054 50 F FL 6640.7 1069 Yes Away 110.678*

*5865 5865 29 29 F6569 65 F CA 6688.4 1077 Yes Away 111.473*

*5866 5866 142 142 F5559 57 F MA 6689.3 1077 Yes Away 111.488*

*5871 5871 143 144 M6064 60 M NH 7419.6 1195 Yes Away 123.660*

*5873 5873 471 471 F3034 34 F NH 7488.4 1206 Yes Away 124.807*

*5874 5874 229 229 F5054 54 F CO 8022.6 1292 Yes Away 133.710*

*>*

*> #Question 1(d)*

*> # Creating subset of values on Column Sex*

*> b.male = subset(roadrace, roadrace$Sex=="M")*

*> b.female = subset(roadrace, roadrace$Sex=="F")*

*> boxplot(b.male$Age, b.female$Age, main="Age of Male and Female")*

*>*

*> # Five Number summary*

*> (b.male.age.summary = summary(b.male$Age))*

*Min. 1st Qu. Median Mean 3rd Qu. Max.*

*9.00 30.00 41.00 40.45 51.00 83.00*

*> (b.male.age.q1 = b.male.age.summary[2])*

*1st Qu.*

*30*

*> (b.male.age.q3 = b.male.age.summary[5])*

*3rd Qu.*

*51*

*>*

*> (b.female.age.summary = summary(b.female$Age))*

*Min. 1st Qu. Median Mean 3rd Qu. Max.*

*7.00 28.00 36.00 37.24 46.00 86.00*

*> (b.female.age.q1 = b.female.age.summary[2])*

*1st Qu.*

*28*

*> (b.female.age.q3 = b.female.age.summary[5])*

*3rd Qu.*

*46*

*>*

*> # Mean*

*> (b.male.age.mean = mean(b.male$Age))*

*[1] 40.4468*

*> (b.female.age.mean = mean(b.female$Age))*

*[1] 37.23653*

*>*

*> # Standard Deviation*

*> (b.male.age.sd = sd(b.male$Age))*

*[1] 13.99289*

*> (b.female.age.sd = sd(b.female$Age))*

*[1] 12.26925*

*>*

*> # Range*

*> (b.male.age.range = range(b.male$Age))*

*[1] 9 83*

*> (b.female.age.range = range(b.female$Age))*

*[1] 7 86*

*>*

*> # Median*

*> (b.male.age.median = median(b.male$Age))*

*[1] 41*

*> (b.female.age.median = median(b.female$Age))*

*[1] 36*

*>*

*> # Interquartile range*

*> (b.male.age.iqr = IQR(b.male$Age))*

*[1] 21*

*> (b.female.age.iqr = IQR(b.female$Age))*

*[1] 18*

*>*

*> (b.male.age.outliers = subset(b.male, (b.male$Age > (b.male.age.q3 + 1.5\*b.male.age.iqr)) | (b.male$Age < (b.male.age.q1 - 1.5\*b.male.age.iqr))))*

*Place Division.Place Division.Entrants Division Age Sex State.Country Time..seconds. Mile.pace..seconds. From.USA Maine Time..minutes.*

*5710 5710 3 4 M8099 83 M ME 5229.4 842 Yes Maine 87.157*

*> (b.female.age.outliers = subset(b.female, (b.female$Age > (b.female.age.q3 + 1.5\*b.female.age.iqr)) | (b.female$Age < (b.female.age.q1 - 1.5\*b.female.age.iqr))))*

*Place Division.Place Division.Entrants Division Age Sex State.Country Time..seconds. Mile.pace..seconds. From.USA Maine Time..minutes.*

*3423 3423 1 14 F7074 74 F ME 3546.8 571 Yes Maine 59.113*

*4662 4662 3 14 F7074 74 F WI 3992.5 643 Yes Away 66.542*

*5455 5455 1 3 F7579 76 F CANADA 4604.7 742 No Away 76.745*

*5617 5617 1 5 F8099 82 F FL 4951.9 797 Yes Away 82.532*

*5714 5714 2 3 F7579 76 F MA 5256.4 846 Yes Away 87.607*

*5757 5757 3 3 F7579 76 F VA 5435.3 875 Yes Away 90.588*

*5765 5765 2 5 F8099 86 F MO 5485.6 883 Yes Away 91.427*

*5807 5807 3 5 F8099 83 F ME 5817.1 937 Yes Maine 96.952*

*5820 5820 4 5 F8099 83 F ME 5945.1 957 Yes Maine 99.085*

*5823 5823 13 14 F7074 74 F CA 5969.8 961 Yes Away 99.497*

*5851 5851 5 5 F8099 81 F ME 6449.9 1038 Yes Maine 107.498*

*>*

*> # Question 2*

*> # Reading the csv into c*

*> c = read.csv("Documents/Fall18/Stats/R\_Mini\_Projects/MP2/motorcycle.csv")*

*> summary(c)*

*County Fatal.Motorcycle.Accidents*

*ABBEVILLE: 1 Min. : 0.00*

*AIKEN : 1 1st Qu.: 6.00*

*ALLENDALE: 1 Median :13.50*

*ANDERSON : 1 Mean :17.02*

*BAMBERG : 1 3rd Qu.:23.00*

*BARNWELL : 1 Max. :60.00*

*(Other) :42*

*>*

*> # Storing just the number of accidents*

*> c.noOfAcc = c$Fatal.Motorcycle.Accidents*

*>*

*> # Boxplot for the number of accidents in various counties*

*> boxplot(c.noOfAcc, main="Number of Accidents")*

*>*

*> # Summary of the distribution*

*> (c.noOfAcc.summary =summary(c.noOfAcc))*

*Min. 1st Qu. Median Mean 3rd Qu. Max.*

*0.00 6.00 13.50 17.02 23.00 60.00*

*>*

*> # Storing q1, q3 and iqr for finding out the outliers*

*> (c.noOfAcc.q1 = c.noOfAcc.summary[2])*

*1st Qu.*

*6*

*> (c.noOfAcc.q3 = c.noOfAcc.summary[5])*

*3rd Qu.*

*23*

*> (c.noOfAcc.iqr = IQR(c.noOfAcc))*

*[1] 17*

*>*

*> # Finding the outliers*

*> (c.noOfAcc.county.outliers = subset(c, ((c.noOfAcc < (c.noOfAcc.q1 - 1.5 \* c.noOfAcc.iqr))|(c.noOfAcc > (c.noOfAcc.q3 + 1.5 \* c.noOfAcc.iqr)))))*

*County Fatal.Motorcycle.Accidents*

*23 GREENVILLE 51*

*26 HORRY 60*