R Notebook

#1.10  
attach(Orange)  
names(Orange)

## [1] "Tree" "age" "circumference"

1.10 discussion: The three variables are “Tree”, “age”, and “circumference.”

#1.11  
mean(age)

## [1] 922.1429

#1.12  
max(circumference)

## [1] 214

detach(Orange)

#2.4.1  
rep("a",5)

## [1] "a" "a" "a" "a" "a"

#2.4.1  
seq(1,1100, by=2)

## [1] 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29  
## [16] 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59  
## [31] 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89  
## [46] 91 93 95 97 99 101 103 105 107 109 111 113 115 117 119  
## [61] 121 123 125 127 129 131 133 135 137 139 141 143 145 147 149  
## [76] 151 153 155 157 159 161 163 165 167 169 171 173 175 177 179  
## [91] 181 183 185 187 189 191 193 195 197 199 201 203 205 207 209  
## [106] 211 213 215 217 219 221 223 225 227 229 231 233 235 237 239  
## [121] 241 243 245 247 249 251 253 255 257 259 261 263 265 267 269  
## [136] 271 273 275 277 279 281 283 285 287 289 291 293 295 297 299  
## [151] 301 303 305 307 309 311 313 315 317 319 321 323 325 327 329  
## [166] 331 333 335 337 339 341 343 345 347 349 351 353 355 357 359  
## [181] 361 363 365 367 369 371 373 375 377 379 381 383 385 387 389  
## [196] 391 393 395 397 399 401 403 405 407 409 411 413 415 417 419  
## [211] 421 423 425 427 429 431 433 435 437 439 441 443 445 447 449  
## [226] 451 453 455 457 459 461 463 465 467 469 471 473 475 477 479  
## [241] 481 483 485 487 489 491 493 495 497 499 501 503 505 507 509  
## [256] 511 513 515 517 519 521 523 525 527 529 531 533 535 537 539  
## [271] 541 543 545 547 549 551 553 555 557 559 561 563 565 567 569  
## [286] 571 573 575 577 579 581 583 585 587 589 591 593 595 597 599  
## [301] 601 603 605 607 609 611 613 615 617 619 621 623 625 627 629  
## [316] 631 633 635 637 639 641 643 645 647 649 651 653 655 657 659  
## [331] 661 663 665 667 669 671 673 675 677 679 681 683 685 687 689  
## [346] 691 693 695 697 699 701 703 705 707 709 711 713 715 717 719  
## [361] 721 723 725 727 729 731 733 735 737 739 741 743 745 747 749  
## [376] 751 753 755 757 759 761 763 765 767 769 771 773 775 777 779  
## [391] 781 783 785 787 789 791 793 795 797 799 801 803 805 807 809  
## [406] 811 813 815 817 819 821 823 825 827 829 831 833 835 837 839  
## [421] 841 843 845 847 849 851 853 855 857 859 861 863 865 867 869  
## [436] 871 873 875 877 879 881 883 885 887 889 891 893 895 897 899  
## [451] 901 903 905 907 909 911 913 915 917 919 921 923 925 927 929  
## [466] 931 933 935 937 939 941 943 945 947 949 951 953 955 957 959  
## [481] 961 963 965 967 969 971 973 975 977 979 981 983 985 987 989  
## [496] 991 993 995 997 999 1001 1003 1005 1007 1009 1011 1013 1015 1017 1019  
## [511] 1021 1023 1025 1027 1029 1031 1033 1035 1037 1039 1041 1043 1045 1047 1049  
## [526] 1051 1053 1055 1057 1059 1061 1063 1065 1067 1069 1071 1073 1075 1077 1079  
## [541] 1081 1083 1085 1087 1089 1091 1093 1095 1097 1099

#2.4.3  
rep(c(1, 2, 3), times=c(3, 3, 3))

## [1] 1 1 1 2 2 2 3 3 3

#2.4.4  
rep(c(1, 2, 3), times=c(3, 2, 1))

## [1] 1 1 1 2 2 3

#2.4.5  
upanddown<-c(seq(1,5, 1), seq(4,1,-1))  
upanddown

## [1] 1 2 3 4 5 4 3 2 1

#2.20  
sales<- c(79, 74, 161, 127, 133, 210, 99, 143, 249, 249, 368, 302)  
names(sales) <- c('JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC')  
sales31 <-sales[c('JAN', 'MAR', 'MAY', 'JUL', 'AUG', 'OCT', 'DEC')]  
saleselse <-sales[c('FEB', 'APR', 'JUN', 'SEP', 'NOV')]  
mean(sales31)

## [1] 166.5714

mean(saleselse)

## [1] 205.6

2.20 discussion: The average sales for months that did not have 31 days was higher (205.6) than in the months that did have 31 days (166.5).

#2.21  
mlbsal<-c(0.57, 0.89, 1.08, 1.12, 1.18, 1.07, 1.17, 1.38, 1.44, 1.72)  
names(mlbsal) <- c(1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999)  
diff(mlbsal)

## 1991 1992 1993 1994 1995 1996 1997 1998 1999   
## 0.32 0.19 0.04 0.06 -0.11 0.10 0.21 0.06 0.28

pctmlbsal <-(diff(mlbsal)/mlbsal)\*100

## Warning in diff(mlbsal)/mlbsal: longer object length is not a multiple of  
## shorter object length

pctmlbsal[-10]

## 1990 1991 1992 1993 1994 1995 1996 1997   
## 56.140351 21.348315 3.703704 5.357143 -9.322034 9.345794 17.948718 4.347826   
## 1998   
## 19.444444

biggest<-pctmlbsal[1]  
biggest

## 1990   
## 56.14035

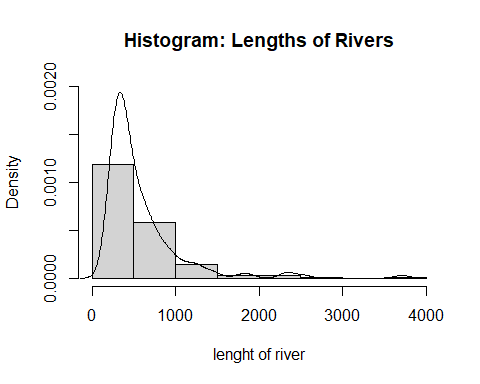
2.21 Discussion: There are several years in which the amount dropped from the previous year. This includes the following: 1992, 1993, 1995, and 1998.

The year with the biggest percent increase is 1990 at 56.1%

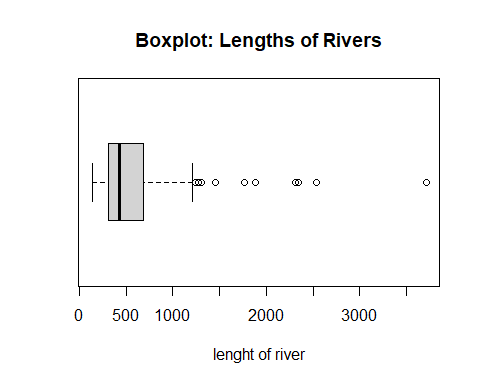
#2.23  
f<-function(x){  
 double<-x\*\*x  
 mean(double)-(mean(x)\*mean(x))  
}  
f(1:10)

## [1] 1040507101

#histoggram and boxplot for rivers data  
hist(rivers, freq=F, breaks=c(0, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000), xlab = "lenght of river", main= "Histogram: Lengths of Rivers", ylim=c(0,.0020))  
lines(density(rivers))



boxplot(rivers, horizontal = T, xlab = "lenght of river", main= "Boxplot: Lengths of Rivers")



Histogram and box plot discussion: The histogram shows that the data is highly skewed to the right. It is also showing that there are no river lengths between 3000-3500 miles long which suggests that the data at approximately 3750 miles long is an outlier. The boxplot confirms the outlier theory. There are several outliers beginning around 2500 miles.

#2.42.1  
plt500<-length(rivers[rivers<500])/length(rivers)\*100

plt500

## [1] 58.15603

#2.42.2  
pltmean<-length(rivers[rivers<mean(rivers)])/length(rivers)\*100

pltmean

## [1] 66.66667

#2.42.3  
quantile(rivers)[4]

## 75%   
## 680

#2.44  
datarivers<-c(mean(rivers), median(rivers), mean(rivers, trim=.25))  
names(datarivers) <- c("mean", "median", "trimmed")  
datarivers

## mean median trimmed   
## 591.1844 425.0000 449.9155

2.44: There quite a bit of a difference between the mean and median. This could suggest a skew or outliers in the data. The trimmed data is closer to the median. This would suggest that some, but not all, of the outliers have been removed.

#2.47  
z.rivers <- scale(rivers)  
mean(z.rivers)

## [1] -5.006707e-17

sd(z.rivers)

## [1] 1

2.47: The z-score is essentially 0 at -5.006707e-17, and the standard deviation =1.

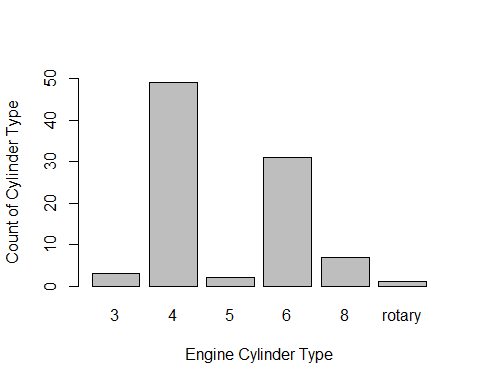
#2.62  
library(MASS)  
attach(Cars93)  
summary(Cylinders)

## 3 4 5 6 8 rotary   
## 3 49 2 31 7 1

2.62: The summary function, when used on factored data, will give the count of each category for the variable.

#2.64

car=table(Cylinders)  
barplot(car, xlab="Engine Cylinder Type", ylab="Count of Cylinder Type", ylim=c(0,50))



detach(Cars93)