The University of British Columbia

Faculty of Sciences

DATA 101

1. Write out the decimal representation of the following binary number: 0.101_2. (2 points)

2. Write out the binary representation of the following decimal number: 0.101_{10}. (2 points)

3. Consider the following programs, designed to calculate $x^4 - y^4$, for $x = 1000000$ and $y = 999999$: (3 points)

   Program A:

   ```
   x <- 1000000
   y <- 999999
   A <- x^4 - y^4
   ```

   Program B:

   ```
   x <- 1000000
   y <- 999999
   B <- (x^2+y^2)*(x+y)*(x-y)
   ```

   Which of `A` and `B` is most accurate? Explain briefly.

4. Suppose the vector `primes` contains the first 10000 prime numbers. (4 points)

   (a) Write down the R code which, when executed, would print the last 100 elements of this vector into the R console window.

   (b) Write down the R code which, when executed, would add up the first 9000 prime numbers.

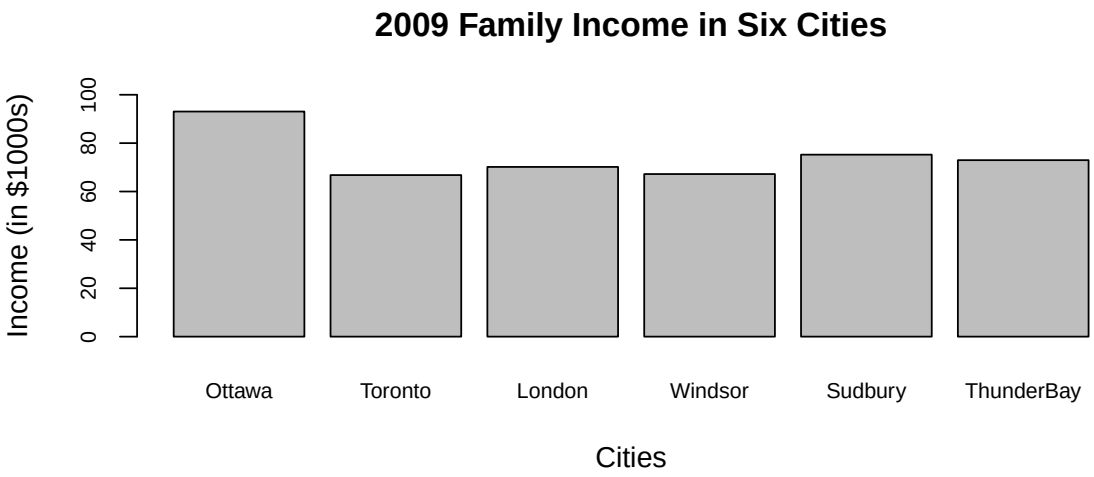
5. Suppose $x = 100$ and $y = 99$. Find the **true** value of (2 points)

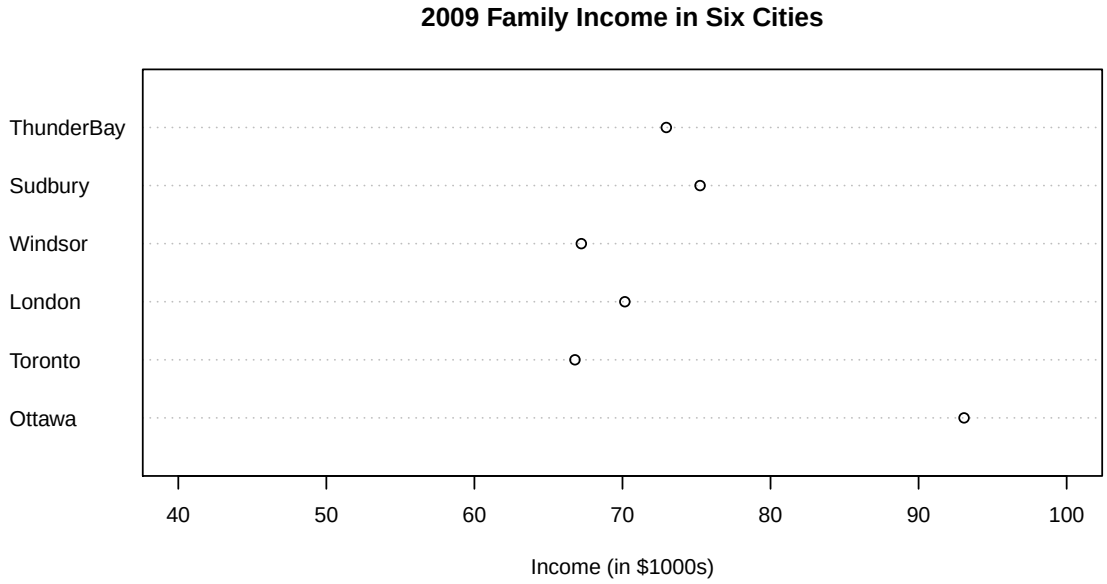
$$x^{16}\left[\frac{(x^8 - y^8)}{(196059601)(19801)(199)} - 1\right].$$

6. Tom started to work in IBM on July 5, 1999. David started on December 12, 2003. How many days did Tom start earlier than David? Using R to solve this question. (4 points)

7. Enter the following numeric data on 2009 family income (in $1000s) for 6 Ontario cities into a 1 row matrix called `income`. Assign the city names to `colnames(income)`. Construct a bar plot and dot plot for this data set as shown below. (8 points)

```
Ottawa Toronto London Windsor Sudbury ThunderBay
93.07     66.79     70.16     67.22       75.24       72.96
```





8. Consider the closing price data in `EuStockMarkets`, a data set that is built in to R.

   (a) Closing prices often have trends which can be removed by taking successive differences using the `diff()` function, yielding what are called "returns". Create a vector called `returns1` which contains the successive differences of the first column of `EuStockMarkets`. (3 points)

   (b) Plot a histogram of the data in `returns1`, using the `scott` breaks and then using the `fd` breaks. Which procedure gives smaller binwidths? (3 points)

   (c) Based on the histograms that you have plotted, would you say that the returns symmetrically distributed or highly skewed to the right or to the left? (1 points)

Solution

(a) 0.625_{10}.

(b) $0.101_{10} = 0.1_{10} + (0.1_{10})^3$ where

$$0.1_{10} = 0.0\overline{0011}_2$$

so that

$$(0.1_{10})^3 = 0.000\overline{000000011011}_2.$$

Adding these together gives the result:

$$0.000110011100111_2...$$

(c) Method B is more accurate because it does not require subtraction of really large numbers (which can result in very large rounding errors).

(d) i. 3

ii. $(2, 4, 6)^(3, 2, 1) = (8, 16, 6)$.

(e) i. `primes[9901:10000]`

ii. `sum(primes[1:9000])`

(f) Suppose $x = 100$ and $y = 99$. Find the **true** value of

$$x^{16}\left[\frac{(x^8 - y^8)}{(196059601)(19801)(199)} - 1\right].$$

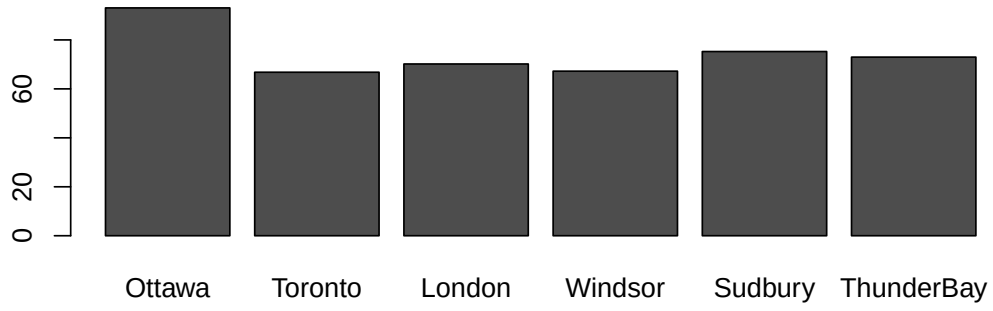
```
> x <- 100
> y <- 99
> x^16*((x^4+y^4)*(x^2+y^2)*(x+y)*(x-y)/(196059601*19801*199)-1)
[1] 0
```

(g) Tom started to work for IBM on July 5, 1999. David started on December 12, 2003. How many days did Tom start earlier than David? Using R to solve this question.

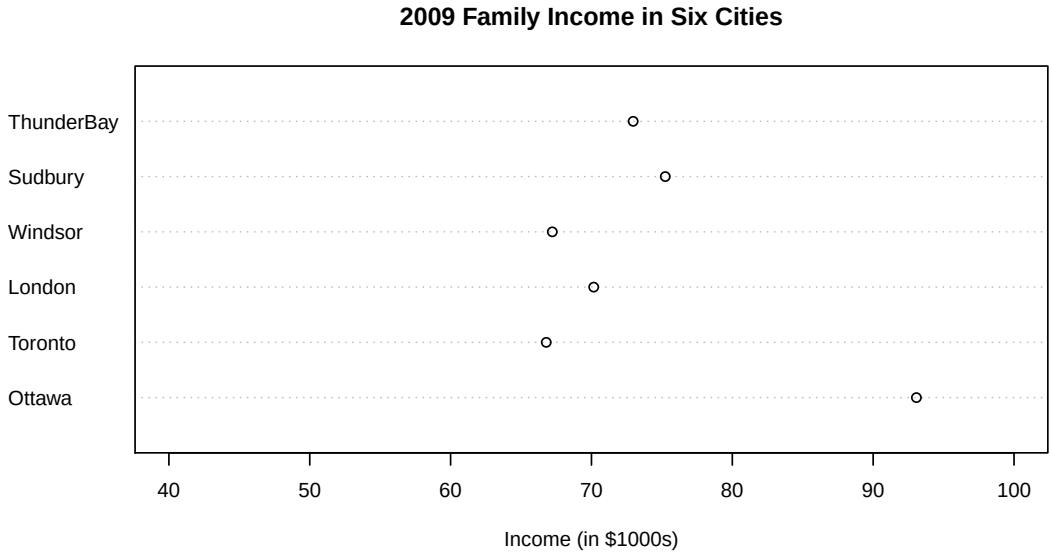
(h) Enter the following numeric data on 2009 family income (in $1000s) for 6 Ontario cities into a 1 row matrix called `income`. Assign the city names to `colnames(income)`. Construct a bar plot for this data set as shown below. (3 points)

```
Ottawa Toronto London Windsor Sudbury ThunderBay
93.07     66.79     70.16     67.22       75.24       72.96
```

```
income <- matrix(c(93.07, 66.79, 70.16, 67.22, 75.24, 72.96), nrow=1)
colnames(income) <- c("Ottawa", "Toronto", "London", "Windsor", "Sudbury", "Thun
barplot(income)
```




```
income <- c(93.07, 66.79, 70.16, 67.22, 75.24, 72.96)
#income <- matrix(c(93.07, 66.79, 70.16, 67.22, 75.24, 72.96), nrow=6)
names(income) <- c("Ottawa", "Toronto", "London", "Windsor", "Sudbury", "Thunderl
dotchart(income, xlab="Income (in $1000s)",main="2009 Family Income in Six Cities
```



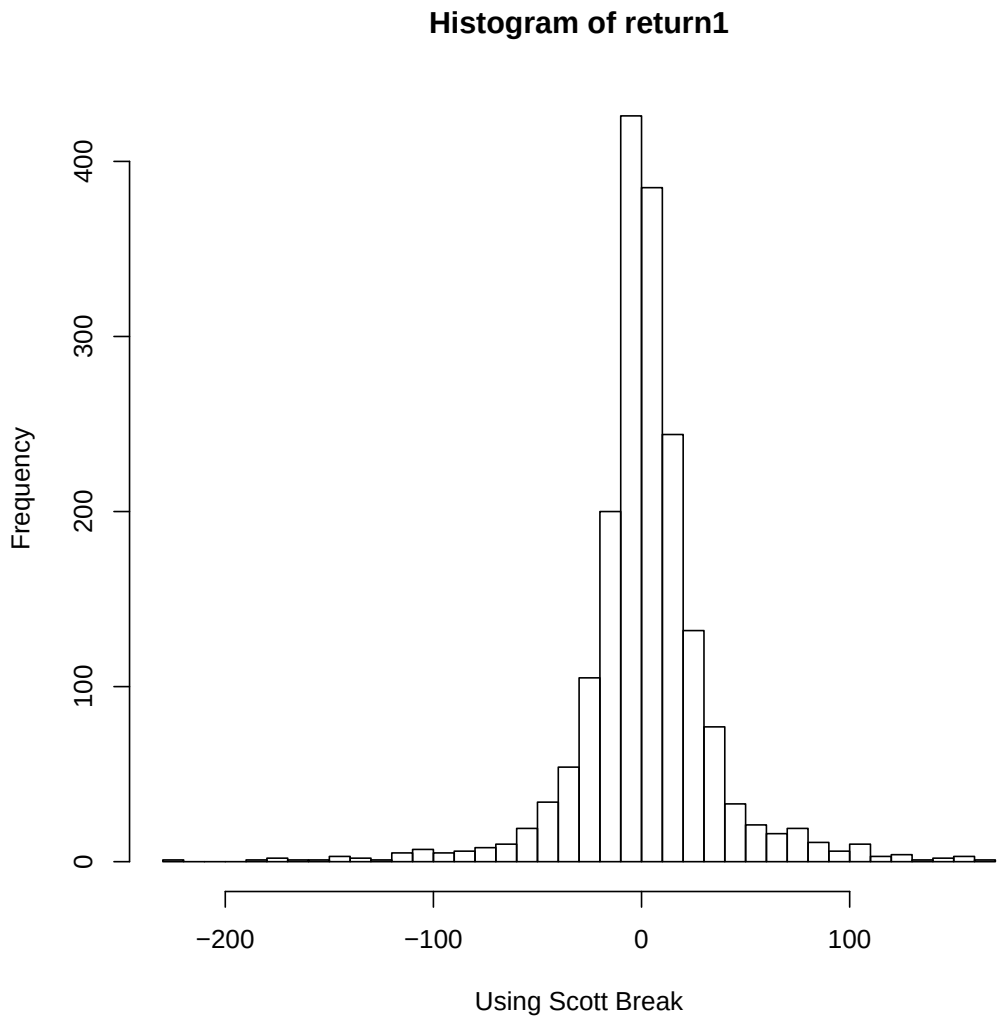
(i) Consider the closing price data in `EuStockMarkets`, a data set that is built in to R.

i. Closing prices often have trends which can be removed by taking successive differences using the `diff()` function, yielding what are called "returns". Create a vector called `returns1` which contains the successive differences of the first column of `EuStockMarkets`. (3 points)

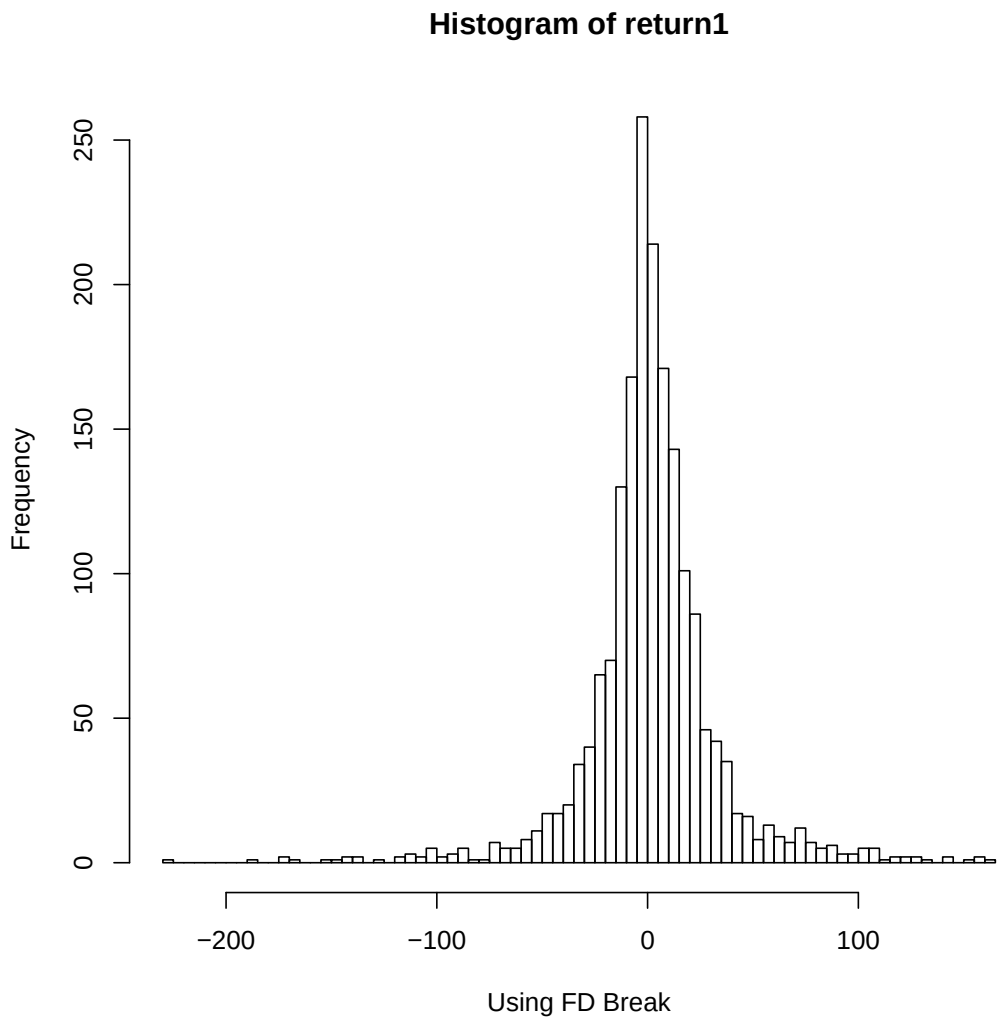
```
return1 <- EuStockMarkets[,1][-1]-EuStockMarkets[,1][-length(EuStockMarkets[,
```

ii. Plot a histogram of the data in `returns1`, using the `scott` breaks and then using the `fd` breaks. Which procedure gives smaller binwidths? (3 points)

```
#hist(return1)
hist(return1, breaks="Scott", xlab="Using Scott Break")
```




```
hist(return1, breaks="fd", xlab="Using FD Break")
```



`Using fd breaks gives smaller binwidth`.

iii. Based on the histograms that you have plotted, would you say that the returns symmetrically distributed or highly skewed to the right or to the left? (1 points)
`The returns is symmetrically distributed`.