Organizing Data

Code

Hide

```
# organising the data

# i)creating string variable
name<-c("Ben",  "Martin",  "Andy",  "Paul",  "Graham",  "Carina",  "Karina",
"Doug", "Mark", "Zoe")


# ii)creating Date Variables

husband<-c("1973-06-21", "1970-07-16", "1949-10-08", "1969-05-24")
wife<-c("1984-11-12", "1973-08-02", "1948-11-11", "1983-07-23")

# agegap<-husband-wife

agegap <- as.Date(husband) - as.Date(wife)
agegap
```

```
Time differences in days
[1] -4162 -1113   331 -5173
```

Hide

```


# iii)creating coding variables/factor

job<-c(1,1,1,1,1,2,2,2,2,2)
job
```

```
 [1] 1 1 1 1 1 2 2 2 2 2
```

Hide

```
# instead we could use
jobs <- c(rep(1,5),rep(2,5))
jobs
```

```
 [1] 1 1 1 1 1 2 2 2 2 2
```

Hide

```
# converting jobs into factor
jobs <- factor(jobs, levels=c(1:2), labels = c('lecturers','students'))
jobs
```

```
 [1] lecturers lecturers lecturers lecturers lecturers students  students  s
tudents  students  students
Levels: lecturers students
```

Hide

```
# another way
job<-gl(2, 5, labels = c("Lecturer", "Student"))
levels(job)
```

```
[1] "Lecturer" "Student"
```

Hide

```
# iv)creating   numeric variable
friends<-c(5,2,0,4,1,10,12,15,12,17)
friends
```

```
 [1]  5  2  0  4  1 10 12 15 12 17
```

Hide

```
income<-c(20000,40000,35000,22000,50000,5000,100,3000,10000,10)
income
```

```
 [1] 20000 40000 35000 22000 50000  5000   100  3000 10000    10
```

Hide

```
alcohol<-c(10,15,20,5,30,25,20,16,17,18)
alcohol
```

```
 [1] 10 15 20  5 30 25 20 16 17 18
```

Hide

```
neurotic<-c(10,17,14,13,21,7,13,9,14,13)
neurotic
```

```
 [1] 10 17 14 13 21  7 13  9 14 13
```

Hide

```
# creating final dataframe

lecturer_data = data.frame(name, friends, income, alcohol, neurotic,job)
lecturer_data
```

name <chr>	friends <dbl>	income <dbl>	alcohol <dbl>	neurotic <dbl>	job <fctr>
Ben	5	20000	10	10	Lecturer
Martin	2	40000	15	17	Lecturer
Andy	0	35000	20	14	Lecturer
Paul	4	22000	5	13	Lecturer
Graham	1	50000	30	21	Lecturer
Carina	10	5000	25	7	Student
Karina	12	100	20	13	Student
Doug	15	3000	16	9	Student
Mark	12	10000	17	14	Student
Zoe	17	10	18	13	Student

1-10 of 10 rows

Hide

```
# Missing Data, use NA
neurotic<-c(10,17,NA,13,21,7,13,9,14,NA)
neurotic
```

```
 [1] 10 17 NA 13 21  7 13  9 14 NA
```

Hide

```
# when data is missing
k = mean(neurotic,na.rm = TRUE)
k
```

```
[1] 13
```

Hide

```
rm(list = ls())
```