INTRODUCTION TO R

Factors

Categorical Variables

- Limited number of different values

- Belong to category

- In R: `factor`

Create factor   `factor()`

```
> blood <- c("B", "AB", "O", "A", "O", "O", "A", "B")
> blood
[1] "B"  "AB" "O"  "A"  "O"  "O"  "A"  "B"


> blood_factor <- factor(blood)
> blood_factor
[1] B  AB O  A  O  O  A  B
Levels: A AB B O


> str(blood_factor)
 Factor w/ 4 levels "A","AB","B","O": 3 2 4 1 4 4 1 3
```

R sorts levels alphabetically!

Order levels differently

```
> blood_factor2 <- factor(blood,
                          levels = c("O", "A", "B", "AB"))

> blood_factor2
[1] B  AB O  A  O  O  A  B
Levels: O A B AB

> str(blood_factor2)
 Factor w/ 4 levels "O","A","B","AB": 3 4 1 2 1 1 2 3

> str(blood_factor)
 Factor w/ 4 levels "A","AB","B","O": 3 2 4 1 4 4 1 3
```

Rename factor levels

```
> blood <- c("B", "AB", "O", "A", "O", "O", "A", "B")
> blood_factor <- factor(blood)

> levels(blood_factor) <- c("BT_A", "BT_AB", "BT_B", "BT_O")
> blood_factor
[1] BT_B  BT_AB BT_O  BT_A  BT_O  BT_O  BT_A  BT_B
Levels: BT_A BT_AB BT_B BT_O

> factor(blood, labels = c("BT_A", "BT_AB", "BT_B", "BT_O"))
[1] BT_B  BT_AB BT_O  BT_A  BT_O  BT_O  BT_A  BT_B
Levels: BT_A BT_AB BT_B BT_O
```

Rename factor levels

```
> blood <- c("B", "AB", "O", "A", "O", "O", "A", "B")
> blood_factor <- factor(blood)


> factor(blood,
         levels = c("O", "A", "B", "AB"),
         labels = c("BT_O", "BT_A", "BT_B", "BT_AB"))
[1] BT_B  BT_AB BT_O  BT_A  BT_O  BT_O  BT_A  BT_B
Levels: BT_O BT_A BT_B BT_AB
```

Nominal versus Ordinal

```
> blood <- c("B", "AB", "O", "A", "O", "O", "A", "B")
> blood_factor <- factor(blood)
> blood_factor[1] < blood_factor[2]
[1] NA
Warning message:
In Ops.factor(blood_factor[1], blood_factor[2]) :
  '<' not meaningful for factors

> tshirt <- c("M", "L", "S", "S", "L", "M", "L", "M")

> tshirt_factor <- factor(tshirt, ordered = TRUE,
                          levels = c("S", "M", "L"))
> tshirt_factor
[1] M L S S L M L M
Levels: S < M < L
```

Ordered factor

```
> tshirt <- c("M", "L", "S", "S", "L", "M", "L", "M")

> tshirt_factor <- factor(tshirt, ordered = TRUE,
                          levels = c("S", "M", "L"))

> tshirt_factor
[1] M L S S L M L M
Levels: S < M < L

> tshirt_factor[1] < tshirt_factor[2]
[1] TRUE
```

Wrap-up

- Factors for categorical variables

- Factors are integer vectors

- Change factor levels:
  `levels()` function or `labels` argument

- Ordered factors: `ordered = TRUE`