R For Data Science Cheat Sheet

Creating a data.table

```r
# set.seed(45L)
DT <- data.table(V1=c(11,21), V2=LETTERS[1:3], V3=LETTERS[4:6], V4=c(1:22), V5=c(1:12))
```

Set a data.table and call it DT

Subsetting Rows Using i

```r
> DT[i=3:5,]
> DT[i=5:1,]
> DT[(V2=="A")]
> DT[(V2 in c("A","C")]]
```

Select 3rd to 5th row
Select 5th to 1st row
Select all rows where the key column (V2) has value "A"
Select all rows where the key column (V2) has value "A" or "C"

Manipulating on Columns in j

```r
> DT[,V2]
> DT[,sum(V1)]
> DT[,sum(V1,na.rm=T)]
> DT[,c(Aggregate=sum(V1), V3+sd(V3))] # Apply a function

The same as the above, with new names
```

Select column V2 and compute std. dev. of V3, which returns a single value and gets recycled. Print column V2 and plot V3

Adding/Updating Columns by Reference in i Using j

```r
> DT[,lapply(.SD,sum),by=V2]
> DT[,print(.SD),by=V2]

Return the result of .SD as a list, grouped by all possible values of V2
Return the result of .SD, grouped by all possible values of V2, and ordered on V1
```

Set key on V2

```r
> setkey(DT, V2)
```

Set name of column V2

```r
DT[,c("V1","V2"):=list(round(exp(V1),2), LETTERS[4:6])]
```

Create a vector as a column

```
DT[,c("V1","V2"):=NULL]
```

Delete columns specified in the variable Cols.chosen

Calculating row summaries (by=V1)

```r
> DT[,.(V4.Sum=sum(V4)),by=V1]

Return the result of V4 as a column, grouped by V1
```

Data.table

data.table is an R package that provides a high-performance version of base R’s data.frame with syntax and feature enhancements for ease of use, convenience and programming speed.

Load the package:

```r
> library(data.table)
```

General form: DT[i, j, by]

```
> DT[i, j, by]
```

“Take DT, subset rows using i, then calculate j grouped by by”

Adding/Updating Columns By Reference in i Using j

```r
> DT[, V1:=round(exp(V1),2)]

Calculate sum of column V1
Return the result of what is after := by calling DT
```

Indexing And Keys

```r
> setkey(DT, V2)
```

Set name of column V2

```r
> DT[.(2,c("V1","V2"),]
```

Create a vector as a column

```
```

Advanced Data Table Operations

```r
> DT[,.(V4.Sum=sum(V4)),by=V1]
```

Calculate sum of column V4, grouped by V1

```r
> DT[,print(.SD),by=V2]
```

Return the result of .SD as a list, grouped by all possible values of V2

```r
> DT[,lapply(.SD,sum),by=V2]
```

Calculate sum of columns in .SD, grouped by V2

```r
> DT[,lapply(.SD,sum),by=V2]
```

Calculate sum of v3 and v4 in .SD, grouped by V2

Chaining

```r
> DT <- DT[,V4,by=V1]
```

Select that group of which the sum is >0

```r
> DT[.<(V4.Sum,V4)].Sum]
```

Calculate sum of V4, grouped by V1, ordered on V1

```
```

Set() - Family

```r
> DT[.<(V4.Sum,V4)].Sum]
```

Select that group of which the sum is >0 (chaining)

```
```

Documentation is available on the DataCamp website: [DataCamp](https://www.datacamp.com)