

---

# 인문학 텍스트 마이닝

---

# **Data Handling(Input & Output)**

# Data Handling(Input & Output)

---

- 경로 확인 및 지정

```
> getwd()
```

```
[1] "/Users/Seongmin_M/Desktop/Data"
```

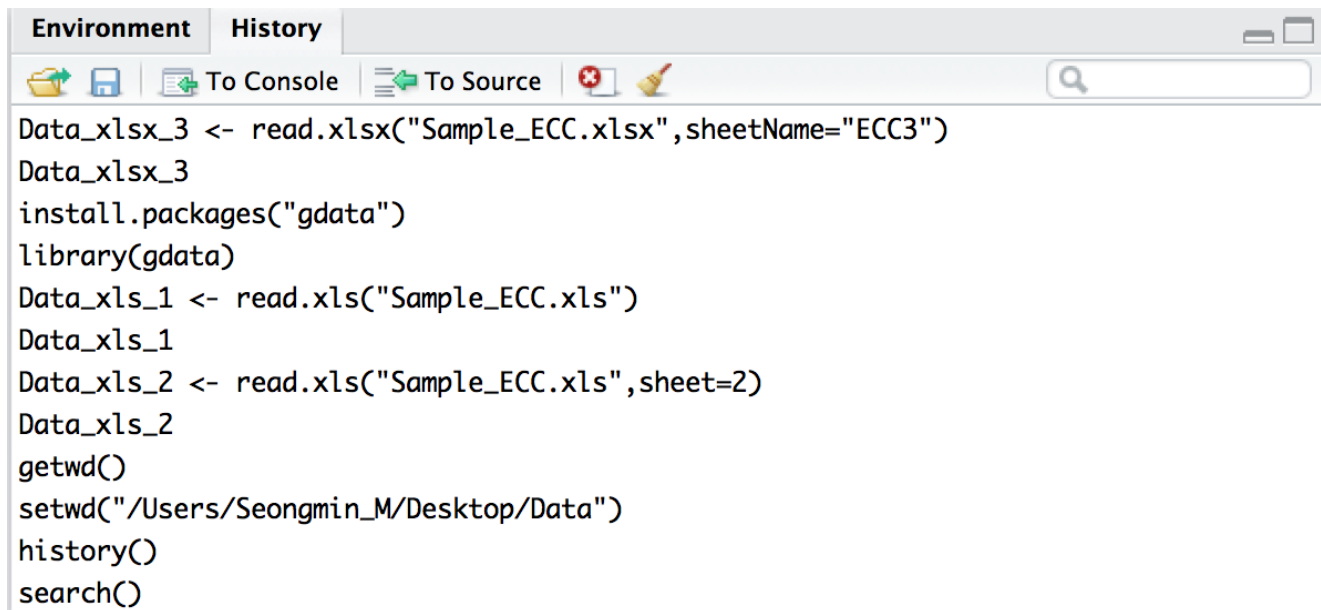
```
>
```

```
> setwd("/Users/Seongmin_M/Desktop/Data")
```

- 작업 내역 확인

```
> history()
```

```
>
```



```
Environment History
To Console To Source
Data_xlsx_3 <- read.xlsx("Sample_ECC.xlsx",sheetName="ECC3")
Data_xlsx_3
install.packages("gdata")
library(gdata)
Data_xls_1 <- read.xls("Sample_ECC.xls")
Data_xls_1
Data_xls_2 <- read.xls("Sample_ECC.xls",sheet=2)
Data_xls_2
getwd()
setwd("/Users/Seongmin_M/Desktop/Data")
history()
search()
```

# Data Handling(Input & Output)

---

- 설치된 패키지 확인

> search()

```
[1] ".GlobalEnv"          "package:gdata"          "package:xlsx"
[4] "package:xlsxjars"    "package:XML"            "package:devtools"
[7] "package:RColorBrewer" "package:ROAuth"         "package:twitter"
[10] "package:RJSONIO"     "package:RCurl"          "package:bitops"
[13] "package:tm"           "package:NLP"            "package:KoNLP"
[16] "package:Sejong"      "package:tau"            "package:hash"
[19] "package:stringr"     "package:rJava"          "tools:rstudio"
[22] "package:stats"       "package:graphics"       "package:grDevices"
[25] "package:utils"       "package:datasets"       "package:methods"
[28] "Autoloads"           "package:base"
```

- 데이터 생성

```
> Sample_data = rbind(
+   c("Anakin", "male", "Tatooine", "41.9BBY", "yes"),
+   c("Amidala", "female", "Naboo", "46BBY", "no"),
+   c("Luke", "male", "Tatooine", "19BBY", "yes"),
+   c("Leia", "female", "Alderaan", "19BBY", "no"),
+   c("Obi-Wan", "male", "Stewjon", "57BBY", "yes"),
+   c("Han", "male", "Corellia", "29BBY", "no"),
+   c("Palpatine", "male", "Naboo", "82BBY", "no"),
+   c("R2-D2", "unknown", "Naboo", "33BBY", "no"))
```

# Data Handling(Input & Output)

---

- DataFrame으로 형태 변환

```
> Sample_df = data.frame(Sample_data)
>
```

- 열 이름 지정

```
> names(Sample_df) = c("Name", "Gender", "Homeworld", "Born", "Jedi")
>
> Sample_df
```

	Name	Gender	Homeworld	Born	Jedi
1	Anakin	male	Tatooine	41.9BBY	yes
2	Amidala	female	Naboo	46BBY	no
3	Luke	male	Tatooine	19BBY	yes
4	Leia	female	Alderaan	19BBY	no
5	Obi-Wan	male	Stewjon	57BBY	yes
6	Han	male	Corellia	29BBY	no
7	Palpatine	male	Naboo	82BBY	no
8	R2-D2	unknown	Naboo	33BBY	no

# Data Handling(Input & Output)

---

- 행 이름 지정

```
> row.names(Sample_df) = c("#1", "#2", "#3", "#4", "#5", "#6", "#7", "#8")
```

```
>
```

```
> Sample_df
```

	Name	Gender	Homeworld	Born	Jedi
#1	Anakin	male	Tatooine	41.9BBY	yes
#2	Amidala	female	Naboo	46BBY	no
#3	Luke	male	Tatooine	19BBY	yes
#4	Leia	female	Alderaan	19BBY	no
#5	Obi-Wan	male	Stewjon	57BBY	yes
#6	Han	male	Coreellia	29BBY	no
#7	Palpatine	male	Naboo	82BBY	no
#8	R2-D2	unknown	Naboo	33BBY	no

- 데이터 속성 확인

```
> str(Sample_df)
```

```
'data.frame': 8 obs. of 5 variables:
```

```
$ Name : Factor w/ 8 levels "Amidala","Anakin",...: 2 1 5 4 6 3 7 8
```

```
$ Gender : Factor w/ 3 levels "female","male",...: 2 1 2 1 2 2 2 3
```

```
$ Homeworld: Factor w/ 5 levels "Alderaan","Coreellia",...: 5 3 5 1 4 2 3 3
```

```
$ Born : Factor w/ 7 levels "19BBY","29BBY",...: 4 5 1 1 6 2 7 3
```

```
$ Jedi : Factor w/ 2 levels "no","yes": 2 1 2 1 2 1 1 1
```

# Data Handling(Input & Output)

---

- 상위 6개 데이터 확인

```
> head(Sample_df)
```

	Name	Gender	Homeworld	Born	Jedi
#1	Anakin	male	Tatooine	41.9BBY	yes
#2	Amidala	female	Naboo	46BBY	no
#3	Luke	male	Tatooine	19BBY	yes
#4	Leia	female	Alderaan	19BBY	no
#5	Obi-Wan	male	Stewjon	57BBY	yes
#6	Han	male	Corellia	29BBY	no

- 저장된 데이터 확인

```
> ls()
```

```
[1] "data"          "Data_xls_1"    "Data_xls_2"    "Data_xlsx_2"   "Data_xlsx_3"  
[6] "Sample_csv_1" "Sample_csv_2" "Sample_data"   "Sample_df"     "Sample_txt_1"
```

- 데이터 지정하여 삭제하기

```
> rm(data)
```

```
> ls()
```

```
[1] "Data_xls_1"    "Data_xls_2"    "Data_xlsx_2"   "Data_xlsx_3"   "Sample_csv_1"  
[6] "Sample_csv_2" "Sample_data"   "Sample_df"     "Sample_txt_1"
```

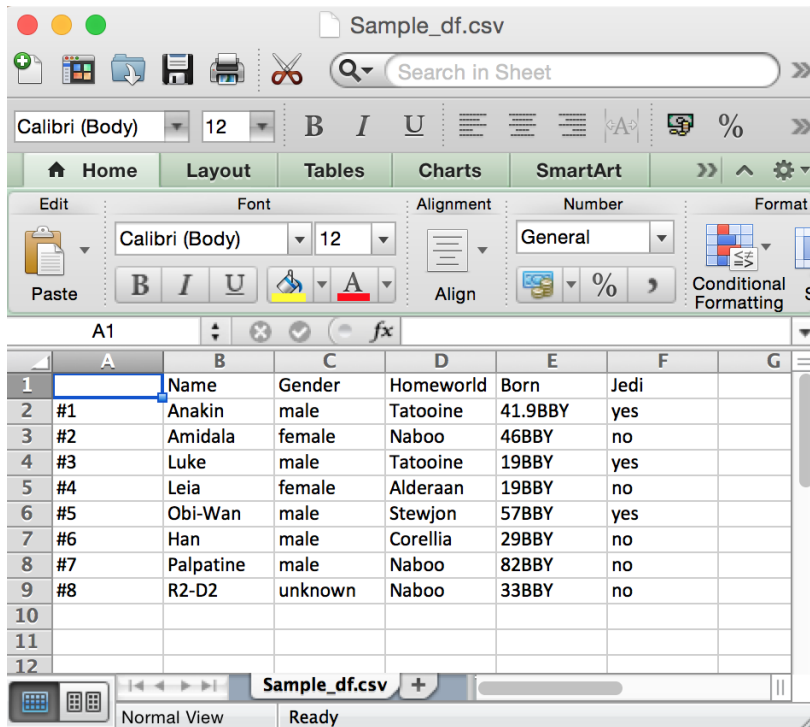
# Data Handling(Input & Output)

- 전체 데이터 삭제하기

```
> rm(list=ls())  
> ls()  
character(0)
```

- 데이터 CSV형태로 내보내기(Output)

```
> write.csv(Sample_df,file="Sample_df.csv")
```



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
1		Name	Gender	Homeworld	Born	Jedi	
2	#1	Anakin	male	Tatooine	41.9BBY	yes	
3	#2	Amidala	female	Naboo	46BBY	no	
4	#3	Luke	male	Tatooine	19BBY	yes	
5	#4	Leia	female	Alderaan	19BBY	no	
6	#5	Obi-Wan	male	Stewjon	57BBY	yes	
7	#6	Han	male	Corellia	29BBY	no	
8	#7	Palpatine	male	Naboo	82BBY	no	
9	#8	R2-D2	unknown	Naboo	33BBY	no	
10							
11							
12							



# Data Handling(Input & Output)

- CSV형태의 데이터 읽어 들이기(Input)

```
> Sample_csv_1 <- read.csv("Sample_df.csv",head=T)
>
> Sample_csv_1
  X      Name Gender Homeworld   Born Jedi
1 #1   Anakin  male   Tatooine 41.9BBY yes
2 #2  Amidala female    Naboo  46BBY  no
3 #3    Luke   male   Tatooine 19BBY  yes
4 #4    Leia  female Alderaan 19BBY  no
5 #5  Obi-Wan  male   Stewjon 57BBY  yes
6 #6     Han   male   Corellia 29BBY  no
7 #7 Palpatine male    Naboo  82BBY  no
8 #8    R2-D2 unknown    Naboo 33BBY  no
```

- 원하는 열 데이터 추출하기

```
> Sample_csv_2 = Sample_csv_1[,2:6]
>
> Sample_csv_2
  Name Gender Homeworld   Born Jedi
1  Anakin  male   Tatooine 41.9BBY yes
2  Amidala female    Naboo  46BBY  no
3    Luke   male   Tatooine 19BBY  yes
4    Leia  female Alderaan 19BBY  no
5  Obi-Wan  male   Stewjon 57BBY  yes
6     Han   male   Corellia 29BBY  no
7 Palpatine male    Naboo  82BBY  no
8    R2-D2 unknown    Naboo 33BBY  no
```

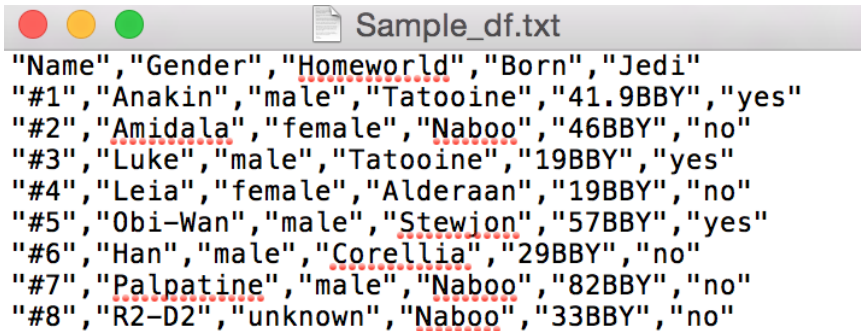
- 행이름 지정하기

```
> row.names(Sample_csv_2) = Sample_csv_1[,1]
> Sample_csv_2
  Name Gender Homeworld   Born Jedi
#1  Anakin  male   Tatooine 41.9BBY yes
#2  Amidala female    Naboo  46BBY  no
#3    Luke   male   Tatooine 19BBY  yes
#4    Leia  female Alderaan 19BBY  no
#5  Obi-Wan  male   Stewjon 57BBY  yes
#6     Han   male   Corellia 29BBY  no
#7 Palpatine male    Naboo  82BBY  no
#8    R2-D2 unknown    Naboo 33BBY  no
```

# Data Handling(Input & Output)

- 데이터 TXT형태로 내보내기(Output)

```
> write.table(Sample_df, file="Sample_df.txt", sep=",")
```



```
"Name", "Gender", "Homeworld", "Born", "Jedi"  
"#1", "Anakin", "male", "Tatooine", "41.9BBY", "yes"  
"#2", "Amidala", "female", "Naboo", "46BBY", "no"  
"#3", "Luke", "male", "Tatooine", "19BBY", "yes"  
"#4", "Leia", "female", "Alderaan", "19BBY", "no"  
"#5", "Obi-Wan", "male", "Stewjon", "57BBY", "yes"  
"#6", "Han", "male", "Coreellia", "29BBY", "no"  
"#7", "Palpatine", "male", "Naboo", "82BBY", "no"  
"#8", "R2-D2", "unknown", "Naboo", "33BBY", "no"
```

- TXT형태의 데이터 읽어 들이기(Input)

```
> Sample_txt_1 <- read.table("Sample_df.txt", header=TRUE, sep=",")
```

```
> str(Sample_txt_1)
```

```
'data.frame': 8 obs. of 5 variables:  
 $ Name      : Factor w/ 8 levels "Amidala","Anakin",...: 2 1 5 4 6 3 7 8  
 $ Gender    : Factor w/ 3 levels "female","male",...: 2 1 2 1 2 2 2 3  
 $ Homeworld: Factor w/ 5 levels "Alderaan","Coreellia",...: 5 3 5 1 4 2 3 3  
 $ Born      : Factor w/ 7 levels "19BBY","29BBY",...: 4 5 1 1 6 2 7 3  
 $ Jedi      : Factor w/ 2 levels "no","yes": 2 1 2 1 2 1 1 1
```

```
> Sample_txt_1
```

	Name	Gender	Homeworld	Born	Jedi
#1	Anakin	male	Tatooine	41.9BBY	yes
#2	Amidala	female	Naboo	46BBY	no
#3	Luke	male	Tatooine	19BBY	yes
#4	Leia	female	Alderaan	19BBY	no
#5	Obi-Wan	male	Stewjon	57BBY	yes
#6	Han	male	Coreellia	29BBY	no
#7	Palpatine	male	Naboo	82BBY	no
#8	R2-D2	unknown	Naboo	33BBY	no