

R scraping and how process in the amazon for the data gathering.

```
library('rvest')
library('polite')
```

```
polite::use_manners(save_as = 'polite_scrape.R')
```

```
## v Setting active project to "/cloud/project".
```

```
urlr <- "https://www.amazon.com/"
```

```
amazon <- read_html(urlr)
```

```
session2 <- bow(urlr,
                user_agent = "Student's Demo Educational")
session2
```

```
## <polite session> https://www.amazon.com/
##   User-agent: Student's Demo Educational
##   robots.txt: 138 rules are defined for 5 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
page2 <- scrape(session2)
```

```
num_products = 31
```

creating a empty data frame for store the data later.

```
data <- data.frame()
```

url for categories

```
shirt_cat<- "https://www.amazon.com/s?k=shirt&i=fashion-mens-intl-ship&crd=6IQRNOUJ0LB&sprefix=shirt%
pants_cat <- "https://www.amazon.com/s?k=pants&i=fashion-mens-intl-ship&crd=9U0VNEZTF2CR&sprefix=pants%
shoe_cat<- "https://www.amazon.com/s?k=shoes&i=fashion-mens-intl-ship&crd=ADB2HOWLHCPK&sprefix=sho%2Cf
head_phone <- "https://www.amazon.com/s?k=headphone&i=fashion-mens-intl-ship&crd=25P9FL9QS4YNZ&sprefix=
medkit_cat<- "https://www.amazon.com/s?k=medkit&i=fashion-mens-intl-ship&crd=1HF70Z2EVLHQY&sprefix=medk
```