Cheatsheet

dplyr	pandas
filter	query
mutate	assign
group_by	groupby
summarise	agg
select	filter
arrange	sort_values

R pipe example

```
# pipe operator
df %>%
  select(country, year, lifeExp) %>%
  filter(country == "Spain" | country == "Portugal") %>%
  mutate(days_to_live = lifeExp * 365.25) %>%
  group_by(year) %>%
  summarise(avg_days_to_live = mean(days_to_live)) %>%
  arrange(desc(year)) %>%
  head(5)
```

```
## # A tibble: 5 x 2
##    year avg_days_to_live
##   <int>            <dbl>
## 1  2007           29044.
## 2  2002           28685.
## 3  1997           28259.
## 4  1992           27838.
## 5  1987           27569.
```

Python method chaining example

```
# open parenthesis, allows inline comments
(df
  .filter(["country", "year", "lifeExp"])
  .query('country == "Spain" or country == "Portugal"')
  .assign(days_to_live = df["lifeExp"] * 365.25)
  .groupby(["year"])
  .agg(avg_days_to_live = ("days_to_live", "mean"))
  .reset_index() # essentially ungroup()
  .sort_values(by="avg_days_to_live", ascending=False)
  .head(5))
```

```
##     year  avg_days_to_live
## 11  2007      29044.497375
## 10  2002      28684.908750
## 9   1997      28259.392500
## 8   1992      27837.528750
## 7   1987      27569.070000
```