The results below are generated from an R script.

```
# • Cuantos campos y observaciones tiene el dataframe.  Utilizar "head" y "dim".
head(airquality) # -> Hay 6 campos: Ozone Solar.R Wind Temp Month Day.

##      Ozone  Solar.R Wind Temp Month Day
## 1 41.00000 190.0000  7.4   67     5   1
## 2 36.00000 118.0000  8.0   72     5   2
## 3 12.00000 149.0000 12.6   74     5   3
## 4 18.00000 313.0000 11.5   62     5   4
## 5 59.11538 181.2963 14.3   56     5   5
## 6 28.00000 181.2963 14.9   66     5   6

dim(airquality)  # -> 153 observaciones con 6 campos.

## [1] 153   6

# • Evaluar el dataframe con la instrucción "summary".
#     o ¿Tiene observaciones con elementos nulos (NA)?
#     o ¿A que meses corresponden las observaciones?
summary(airquality)

##      Ozone           Solar.R           Wind             Temp           Month
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000
##  1st Qu.: 21.00   1st Qu.:118.5   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000
##  Median : 45.00   Median :199.0   Median : 9.700   Median :79.00   Median :7.000
##  Mean   : 46.24   Mean   :185.8   Mean   : 9.958   Mean   :77.88   Mean   :6.993
##  3rd Qu.: 59.12   3rd Qu.:257.5   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000
##                   NA's   :3
##       Day
##  Min.   : 1.0
##  1st Qu.: 8.0
##  Median :16.0
##  Mean   :15.8
##  3rd Qu.:23.0
##  Max.   :31.0
##

# Hay valores NA en Ozone (37) y en Solar Radiation (7).
# Los meses durante los que se realizaron las observaciones son del 5 al 9 (es decir de mayo a septiembr

# • Temperatura máxima del viento en el mes de mayo.
max(airquality[airquality$Month == 5,]$Temp) # <- 81

## [1] 81

# • Media del ozono en el mes de Julio.

mean(airquality[airquality$Month == 7,]$Ozone,na.rm = TRUE) # -> 59.11538

## [1] 59.11538

media=mean(airquality[airquality$Month == 7,]$Ozone,na.rm = TRUE) # -> 59.11538

# Transformar al valor de la media los NA.
```

```r
airquality$Ozone[is.na(airquality$Ozone)] <- media
# Estudiar el efecto de esta asignación sobre la desviación típica

mean(airquality[airquality$Month == 7,]$Ozone)
```

```
## [1] 59.11538
```

```r
# • Mes donde la temperatura fue mayor.
airquality[max(airquality$Temp),]$Month # -> Agosto (8)
```

```
## [1] 8
```

```r
# • Mes donde la temperatura y el ozono fue mayor.
length(airquality[airquality$Temp > 90 & airquality$Ozone < 100,"Month"]) # -> 13
```

```
## [1] 13
```

```r
# • Haciendo un estudio de los datos, ¿Qué podemos concluir?
# ¿Existe alguna relación entre las variables Ozono, Temperatura y Radiación Solar?
# Se recomienda hacer la media mes a mes de cada variable.

mean(airquality$Ozone[airquality$Month == 5])
```

```
## [1] 29.34119
```

```r
mean(airquality$Ozone[airquality$Month == 6])
```

```
## [1] 50.2141
```

```r
mean(airquality$Ozone[airquality$Month == 7])
```

```
## [1] 59.11538
```

```r
mean(airquality$Ozone[airquality$Month == 8])
```

```
## [1] 59.82506
```

```r
mean(airquality$Ozone[airquality$Month == 9])
```

```
## [1] 32.37051
```

```r
mean(airquality$Temp[airquality$Month == 5])
```

```
## [1] 65.54839
```

```r
mean(airquality$Temp[airquality$Month == 6])
```

```
## [1] 79.1
```

```r
mean(airquality$Temp[airquality$Month == 7])
```

```
## [1] 83.90323
```

```r
mean(airquality$Temp[airquality$Month == 8])
```

```
## [1] 83.96774
```

```r
mean(airquality$Temp[airquality$Month == 9])
```

```
## [1] 76.9

media_Solar.R_5=mean(airquality$Solar.R[airquality$Month == 5],na.rm = TRUE)
sqrt(var(airquality$Solar.R[airquality$Month == 5],na.rm=TRUE))

## [1] 107.1295

media_Solar.R_6=mean(airquality$Solar.R[airquality$Month == 6],na.rm = TRUE)
media_Solar.R_7=mean(airquality$Solar.R[airquality$Month == 7],na.rm = TRUE)
media_Solar.R_8=mean(airquality$Solar.R[airquality$Month == 8],na.rm = TRUE)
media_Solar.R_9=mean(airquality$Solar.R[airquality$Month == 9],na.rm = TRUE)

# Transformar los NA.
table(is.na(airquality$Solar.R))

##
## FALSE  TRUE
##   150     3

indices5=which(is.na(airquality$Solar.R[airquality$Month == 5]))
# ojo, estamos cambiando todos los datos sin haber salvado la anterior versión
# del dataframe
airquality$Solar.R[airquality$Month == 5][indices5]=media_Solar.R_5
mean(airquality$Solar.R[airquality$Month == 5])

## [1] 181.2963

sqrt(var(airquality$Solar.R[airquality$Month == 5]))

## [1] 107.1295

# Si se hace la media del ozono, temperatura y radiacion solar podemos observar como más o menos todos
# El mes de junio es el único que presenta algo de variación.
# Se podría concluir que todas las variables indicadas tiene relación entre ellas.
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.3.1 (2023-06-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.6 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/atlas/libblas.so.3.10.3
## LAPACK: /usr/lib/x86_64-linux-gnu/atlas/liblapack.so.3.10.3;  LAPACK version 3.9.0
##
## locale:
##  [1] LC_CTYPE=es_ES.UTF-8       LC_NUMERIC=C               LC_TIME=es_ES.UTF-8
##  [4] LC_COLLATE=es_ES.UTF-8     LC_MONETARY=es_ES.UTF-8    LC_MESSAGES=es_ES.UTF-8
##  [7] LC_PAPER=es_ES.UTF-8       LC_NAME=C                  LC_ADDRESS=C
## [10] LC_TELEPHONE=C             LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Madrid
## tzcode source: system (glibc)
##
```

```
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] knitr_1.44       factoextra_1.0.7 ggplot2_3.4.3    arules_1.7-6     Matrix_1.6-1.1
##
## loaded via a namespace (and not attached):
##  [1] gtable_0.3.4         xfun_0.40            recipes_1.0.8        ggrepel_0.9.3
##  [5] lattice_0.21-9       vctrs_0.6.3          tools_4.3.1          generics_0.1.3
##  [9] stats4_4.3.1         parallel_4.3.1       tibble_3.2.1         fansi_1.0.5
## [13] highr_0.10           pkgconfig_2.0.3      ModelMetrics_1.2.2.2 data.table_1.14.8
## [17] lifecycle_1.0.3      farver_2.1.1         compiler_4.3.1       stringr_1.5.0
## [21] munsell_0.5.0        codetools_0.2-19     DALEX_2.4.3          htmltools_0.5.6.1
## [25] class_7.3-22         yaml_2.3.7           prodlim_2023.08.28   pillar_1.9.0
## [29] MASS_7.3-60          gower_1.0.1          iterators_1.0.14     rpart_4.1.19
## [33] foreach_1.5.2        nlme_3.1-163         parallelly_1.36.0    lava_1.7.2.1
## [37] tidyselect_1.2.0     digest_0.6.33        stringi_1.7.12       future_1.33.0
## [41] dplyr_1.1.3          reshape2_1.4.4       purrr_1.0.2          listenv_0.9.0
## [45] labeling_0.4.3       splines_4.3.1        cowplot_1.1.1        fastmap_1.1.1
## [49] grid_4.3.1           colorspace_2.1-0     cli_3.6.1            magrittr_2.0.3
## [53] survival_3.5-7       utf8_1.2.3           future.apply_1.11.0  withr_2.5.1
## [57] scales_1.2.1         xgboost_1.7.5.1      lubridate_1.9.3      timechange_0.2.0
## [61] rmarkdown_2.25       globals_0.16.2       nnet_7.3-19          timeDate_4022.108
## [65] evaluate_0.22        hardhat_1.3.0        caret_6.0-94         rlang_1.1.1
## [69] Rcpp_1.0.11          glue_1.6.2           pROC_1.18.4          ipred_0.9-14
## [73] rstudioapi_0.15.0    jsonlite_1.8.7       R6_2.5.1             plyr_1.8.9
```

```
Sys.time()
```

```
## [1] "2023-10-31 22:29:26 CET"
```