

# Package ‘GWASr’

December 13, 2021

**Type** Package

**Title** Plotting Summary Statistics In GGPlot

**Version** 0.1.0

**Author** c(  
 person("`Franjo", "`Ivankovic", email = "`njofrica@ufl.edu", role = c("`aut", "`cre")),  
 person("`Mingjing", "`Chen", email = "`mingjingchen@ufl.edu", role = c("`aut")),  
 person("`Xiaoxiao", "`Wen", email = "`xiaoxiao.wen@ufl.edu", role = c("`aut")))

**Maintainer** Franjo Ivankovic <njofrica@ufl.edu>

**Description** GGWASr is a GGPlot2 expansion pack of functions which are capable of plotting the genome-wide association study summary statistics using ggplot package, as well as running some basic analyses with the data.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Imports** Rcpp, dplyr (>= 1.0.7), ggplot2 (>= 3.3.5)

**LinkingTo** Rcpp

**RoxygenNote** 7.1.2

**Depends** R (>= 2.10)

**Suggests** rmarkdown, knitr, testthat (>= 3.0.0)

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**NeedsCompilation** yes

## R topics documented:

ADHDMeta . . . . .	2
ADHDMetaP . . . . .	2
GLambda . . . . .	3
plot_manhattan . . . . .	3
plot_qq . . . . .	5
<b>Index</b>	7

ADHDMeta

*ADHD Meta analysis dataset***Description**

A dataset of pruned summary statistics from a large ADHD GWAS study. Dataset consists of 239961 observations and 5 variables.

**Usage**

ADHDMeta

**Format**

A data frame with 239961 rows and 5 variables:

**CHR** chromosome ID (chr)

**SNP** SNP ID (chr)

**BP** base-pair genomic location (num)

**OR** Odds-ratio from GWAS (num)

**P** P values from GWAS (num)

**Source**

Demontis, D., Walters, R.K., Martin, J. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51, 63–75 (2019). <https://doi.org/10.1038/s41588-018-0269-7>

ADHDMetaP

*ADHD Meta P analysis dataset***Description**

A dataset P values from summary statistics from a large ADHD GWAS study. Dataset consists of 8094094 observations and 2 variables.

**Usage**

ADHDMetaP

**Format**

A data frame with 8094094 rows and 2 variables:

**SNP** SNP ID (chr)

**P** P values from GWAS (num)

**Source**

Demontis, D., Walters, R.K., Martin, J. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51, 63–75 (2019). <https://doi.org/10.1038/s41588-018-0269-7>

---

GLambda	<i>GLambda</i>
---------	----------------

---

**Description**

GLambda (Description)

**Usage**

GLambda(x)

**Arguments**

x                      A numeric vector of P values

**Details**

This function calculates genomic constant lambda from observed p values. Values are calcuated as median of p-values' chi-test statistic, divided by 0.4549364. This result indicate deflation (< 1) or inflation (> 1) of the GWAS p values. If pvalues are extremely inflated, then adjustments to model are necessary.

**Value**

numeric value

**Author(s)**

Franjo, Mingjing, and Xiaoxiao

**Examples**

GLambda(ADHDMetaP\$P)

---

plot_manhattan	<i>Manhattan Plot</i>
----------------	-----------------------

---

**Description**

*plot\_manhattan* takes chromosome, base pair, and p-value data and creates a ggplot based Manhattan plot.

**Usage**

```
plot_manhattan(
  data = data,
  P = "P",
  CHR = "CHR",
  BP = "BP",
  maxP = NULL,
  xlab = NULL,
  ylab = NULL,
  main = NULL,
  colrs = NULL,
  sigA = NULL,
  sigAcol = NULL,
  sigB = NULL,
  sigBcol = NULL,
  ylims = NULL,
  ybreaks = NULL
)
```

**Arguments**

data	Data frame which contains GWAS summary statistics (REQUIRED)
P	A column name which contains P values, default is "P", column should be numeric.
CHR	A column name which contains chromosome ID, default is "CHR", column should be numeric or character.
BP	A column name which contains base pair location, default is "BP", column should be numeric.
maxP	A maximum P-value cutoff (sometimes, it's not a good use of memory to plot millions of p values above 0.1), default is 1.
xlab	Optional x-axis label, default is NULL.
ylab	Optional y-axis label, default is "-log10(P)"
main	Optional plot title, default is "Manhattan plot"
colrs	Optional character vector of colors, default is set by package.
sigA	Optional adjustment to stringent p-value reference, default is 3e-8
sigAcol	Option string to adjust the color of sigA line
sigB	Optional adjustment to stringent p-value reference, default is 1e-5
sigBcol	Option string to adjust the color of sigB line
ylims	Optional adjustments to y limits, default adjusts it to data
ybreaks	Optional adjustments to y breaks, default adds whole numbers in increments of 1

**Details**

Basic functionality: *plot\_manhattan* takes chromosome and basepair address information, as well as p values from the GWAS summary statistics. Subsequently, data are additionally processed to generate Manhattan plot where chromosome and basepair information are used to generate x-axis, and the  $-\log_{10}(P)$  values are plotted on the y axis.

**Value**

ggplot object

**Author(s)**

Franjo, Mingjing, and Xiaoxiao

**Examples**

```
plot_manhattan(ADHDMeta)
```

---

plot_qq	<i>QQ Plot</i>
---------	----------------

---

**Description**

*plot\_qq* takes GWAS summary statistics p-values and creates a ggplot based QQ plot.

**Usage**

```
plot_qq(
  data = data,
  showN = TRUE,
  showGL = TRUE,
  P = "P",
  xlab = NULL,
  ylab = NULL,
  main = NULL,
  pcol = NULL,
  lcol = NULL
)
```

**Arguments**

data	Data frame which contains GWAS summary statistics (REQUIRED)
showN	Shows total number of SNPs on the plot, default is TRUE
showGL	Shows genomic constant lambda on the plot, default is TRUE
P	A column name which contains P values, default is "P", column should be numeric.
xlab	Optional x-axis label, default is NULL.
ylab	Optional y-axis label, default is "-log10(P)"
main	Optional plot title, default is "Manhattan plot"
pcol	String indicating point colors, default is set by the package.
lcol	String indicating reference line color, default is set by the package.

**Details**

Basic functionality: *plot\_qq* first takes the p-values from specified data frame and then based on sorted p-values creates expectations. Subsequently the data are pruned by observed p-values in a bin-wise manner before getting plotted. This pruning is done to reduce the likelihood of crashing as usual GWAS meta-analyses have large number of datapoints (millions).

**Value**

ggplot object

**Author(s)**

Franjo, Mingjing, and Xiaoxiao

**Examples**

```
plot_qq(ADHDMetaP)
```

# Index

## \* **datasets**

ADHDMeta, [2](#)

ADHDMetaP, [2](#)

ADHDMeta, [2](#)

ADHDMetaP, [2](#)

GLambda, [3](#)

plot\_manhattan, [3](#)

plot\_qq, [5](#)