

main

August 27, 2021

1 Plot and comparisons

```
[1]: library(tidyverse)
library(ggpubr)
```

```
Attaching packages: tidyverse
1.3.1
```

```
ggplot2 3.3.5    purrr  0.3.4
tibble  3.1.2    dplyr  1.0.7
tidyr   1.1.3    stringr 1.4.0
readr   1.4.0    forcats 0.5.1
```

```
Conflicts
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag()    masks stats::lag()
```

1.1 Functions

```
[2]: save_plot <- function(p, fn, w=7, h=6){
  for(ext in c(".pdf", ".png", ".svg")){
    ggsave(filename=paste0(fn,ext), plot=p, width=w, height=h)
  }
}

get_metrics <- function(fn, model, label){
  dt = data.table::fread(fn) %>% as.data.frame %>% mutate_if(is.character, as.
  ↪factor) %>%
    mutate_at("fold", as.character) %>%
    select(tissue, feature, fold, n_features, starts_with("test_score_r2"))
  ↪%>%
    pivot_longer(-c(tissue, feature, fold), names_to="metric",
  ↪values_to="score") %>%
    group_by(tissue, feature, metric) %>%
```

```

    summarise(Mean=mean(score), Median=median(score), Std=sd(score), .
    ↪groups = "keep") %>%
    filter(metric == "test_score_r2") %>% mutate("model"=model) %>%
    ↪mutate(Type = label)
    return(dt)
}

```

1.2 Genetic variation prediction for expression of ancestry DE genes and random genes

1.2.1 Load data

```

[3]: top100 = data.table::fread("../_m/degs_annotation.txt") %>%
    group_by(Tissue) %>% mutate(rank = row_number(adj.P.Val)) %>%
    filter(rank <= 100) %>% select(V1, ensemblID, gene_name, Tissue) %>%
    distinct %>% rename("Feature"="V1") %>% rename("tissue"="Tissue")

```

Annotate and merge data

```

[4]: dtu = data.table::fread(paste0("../_m/differential_analysis/
    ↪tissue_comparison/",
                                "ds_summary/_m/
    ↪diffSplicing_ancestry_FDR05_4regions.tsv")) %>%
    select(gene, Tissue) %>% distinct %>% rename("gene_name"="gene")

degs = data.table::fread("../_m/degs_annotation.txt") %>%
    select(V1, ensemblID, gene_name, Tissue) %>% distinct %>%
    rename("Feature"="V1") %>% inner_join(dtu, by=c("Tissue", "gene_name")) %>%
    rename("tissue"="Tissue") %>% mutate("DTU"="DTU")

random = data.table::fread("../_m/randomGenes_annotation.txt") %>%
    select(V1, ensemblID, gene_name, Tissue) %>% distinct %>%
    rename("Feature"="V1") %>% inner_join(dtu, by=c("Tissue", "gene_name")) %>%
    rename("tissue"="Tissue") %>% mutate("DTU"="DTU")

```

DE genes

```

[5]: rf = get_metrics("../de_genes/rf/summary_10Folds_allTissues.tsv", "Random_
    ↪Forest", "DE")
enet = get_metrics("../de_genes/enet/summary_10Folds_allTissues.tsv",
    ↪"Elastic Net", "DE")
de = bind_rows(rf, enet) %>% mutate(Feature=gsub("_", ".", feature)) %>%
    left_join(degs, by=c("tissue", "Feature")) %>% as.data.frame %>%
    mutate(New_Type = paste(Type, replace_na(DTU, ""))) %>%
    mutate_if(is.character, as.factor)
de %>% head
de$Type %>% unique

```

		tissue <fct>	feature <fct>	metric <fct>	Mean <dbl>	Median <dbl>	Std <dbl>
A data.frame: 6 × 13	1	Caudate	ENSG000000003249_13	test_score_r2	-0.002856899	0.03781765	0.198
	2	Caudate	ENSG000000003509_15	test_score_r2	-0.128148560	-0.06360597	0.150
	3	Caudate	ENSG000000004468_12	test_score_r2	-0.198605136	-0.09842729	0.188
	4	Caudate	ENSG000000004777_18	test_score_r2	-0.344985568	-0.25373680	0.482
	5	Caudate	ENSG000000005243_9	test_score_r2	-0.248017098	-0.21435654	0.290
	6	Caudate	ENSG000000005436_13	test_score_r2	0.126125017	0.13625009	0.160

DE Levels: 'DE'

Top 100 DE genes

```
[6]: de100 = bind_rows(rf, enet) %>% mutate(Feature=gsub("_", ".", feature)) %>%
      inner_join(top100, by=c("tissue", "Feature"))
de100 %>% dim
```

1. 764 2. 11

Random genes

```
[7]: rf = get_metrics("../random_genes/rf/summary_10Folds_allTissues.tsv",
                     "Random Forest", "Random")
enet = get_metrics("../random_genes/enet/summary_10Folds_allTissues.tsv",
                   "Elastic Net", "Random")
rand = bind_rows(rf, enet)%>% mutate(Feature=gsub("_", ".", feature)) %>%
      left_join(random, by=c("tissue", "Feature")) %>% as.data.frame %>%
      mutate(New_Type = paste(Type, replace_na(DTU, ""))) %>%
      mutate_if(is.character, as.factor)
rand %>% head
rand$Type %>% unique
```

		tissue	feature	metric	Mean	Median	Std
		<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>
A data.frame: 6 × 13	1	Caudate	ENSG000000001084_10	test_score_r2	-0.10568232	-0.09784934	0.1329
	2	Caudate	ENSG000000001630_15	test_score_r2	-0.18184880	-0.11087436	0.2012
	3	Caudate	ENSG000000002587_9	test_score_r2	-0.16901411	-0.08722611	0.2458
	4	Caudate	ENSG000000002933_7	test_score_r2	-0.07044221	-0.03925164	0.1650
	5	Caudate	ENSG000000003393_14	test_score_r2	-0.19319686	-0.14058706	0.2751
	6	Caudate	ENSG000000003400_14	test_score_r2	-0.11172415	-0.08874340	0.1617

Random Levels: 'Random'

Merge data

```
[8]: df = bind_rows(de, rand)
dim(df)
df %>% head(2)
df$Type %>% unique
```

1. 37468 2. 13

A data.frame: 2 × 13		tissue	feature	metric	Mean	Median	Std
		<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>
	1	Caudate	ENSG00000003249_13	test_score_r2	-0.002856899	0.03781765	0.198
	2	Caudate	ENSG00000003509_15	test_score_r2	-0.128148560	-0.06360597	0.150

1. DE 2. Random

Levels: 1. 'DE' 2. 'Random'

1.2.2 Summarize

```
[9]: df %>% group_by(tissue, Type, model) %>%
      summarise(Mean=mean(Median), Median=median(Median), .groups = "keep")
```

A grouped_df: 16 × 5	tissue	Type	model	Mean	Median
	<fct>	<fct>	<fct>	<dbl>	<dbl>
	Caudate	DE	Elastic Net	0.046824293	-0.01074297
	Caudate	DE	Random Forest	-0.003779483	-0.04764570
	Caudate	Random	Elastic Net	-0.031713741	-0.03899560
	Caudate	Random	Random Forest	-0.096066770	-0.09682914
	Dentate Gyrus	DE	Elastic Net	-0.042714190	-0.07105638
	Dentate Gyrus	DE	Random Forest	-0.109158980	-0.13368794
	Dentate Gyrus	Random	Elastic Net	-0.141836455	-0.12442962
	Dentate Gyrus	Random	Random Forest	-0.321247001	-0.31566815
	DLPFC	DE	Elastic Net	0.043382091	-0.01427053
	DLPFC	DE	Random Forest	-0.015966827	-0.05965098
	DLPFC	Random	Elastic Net	-0.049235705	-0.05156694
	DLPFC	Random	Random Forest	-0.130636227	-0.12972883
	Hippocampus	DE	Elastic Net	0.030628956	-0.01784020
	Hippocampus	DE	Random Forest	-0.024028727	-0.05773127
	Hippocampus	Random	Elastic Net	-0.038067283	-0.03998911
	Hippocampus	Random	Random Forest	-0.105771529	-0.10358206

```
[10]: df %>% group_by(tissue, Type, model) %>%
       summarise(Mean=mean(Median), Median=median(Median), .groups = "keep") %>%
       filter(model == "Elastic Net")
```

A grouped_df: 8 × 5	tissue	Type	model	Mean	Median
	<fct>	<fct>	<fct>	<dbl>	<dbl>
	Caudate	DE	Elastic Net	0.04682429	-0.01074297
	Caudate	Random	Elastic Net	-0.03171374	-0.03899560
	Dentate Gyrus	DE	Elastic Net	-0.04271419	-0.07105638
	Dentate Gyrus	Random	Elastic Net	-0.14183646	-0.12442962
	DLPFC	DE	Elastic Net	0.04338209	-0.01427053
	DLPFC	Random	Elastic Net	-0.04923571	-0.05156694
	Hippocampus	DE	Elastic Net	0.03062896	-0.01784020
	Hippocampus	Random	Elastic Net	-0.03806728	-0.03998911

```
[11]: df %>% group_by(tissue, Type, model) %>%
      summarise(Mean=mean(Median), Median=median(Median), .groups = "keep") %>%
      filter(model == "Random Forest")
```

	tissue <fct>	Type <fct>	model <fct>	Mean <dbl>	Median <dbl>
A grouped_df: 8 × 5	Caudate	DE	Random Forest	-0.003779483	-0.04764570
	Caudate	Random	Random Forest	-0.096066770	-0.09682914
	Dentate Gyrus	DE	Random Forest	-0.109158980	-0.13368794
	Dentate Gyrus	Random	Random Forest	-0.321247001	-0.31566815
	DLPFC	DE	Random Forest	-0.015966827	-0.05965098
	DLPFC	Random	Random Forest	-0.130636227	-0.12972883
	Hippocampus	DE	Random Forest	-0.024028727	-0.05773127
	Hippocampus	Random	Random Forest	-0.105771529	-0.10358206

```
[12]: df %>% filter(DTU == "DTU") %>% group_by(tissue, New_Type, model) %>%
      summarise(Mean=mean(Median), Median=median(Median), .groups = "keep")
```

	tissue <fct>	New_Type <fct>	model <fct>	Mean <dbl>	Median <dbl>
A grouped_df: 16 × 5	Caudate	DE DTU	Elastic Net	0.081422714	0.01311672
	Caudate	DE DTU	Random Forest	0.033633330	-0.02671328
	Caudate	Random DTU	Elastic Net	-0.042723875	-0.04345409
	Caudate	Random DTU	Random Forest	-0.110650265	-0.09868582
	Dentate Gyrus	DE DTU	Elastic Net	0.005700645	-0.07592577
	Dentate Gyrus	DE DTU	Random Forest	-0.035912422	-0.06890107
	Dentate Gyrus	Random DTU	Elastic Net	-0.124886491	-0.09891128
	Dentate Gyrus	Random DTU	Random Forest	-0.263556534	-0.22179058
	DLPFC	DE DTU	Elastic Net	0.113791408	0.03734998
	DLPFC	DE DTU	Random Forest	0.062114999	-0.01159725
	DLPFC	Random DTU	Elastic Net	-0.032211432	-0.04758998
	DLPFC	Random DTU	Random Forest	-0.104759586	-0.12445701
	Hippocampus	DE DTU	Elastic Net	0.087471515	0.01387694
	Hippocampus	DE DTU	Random Forest	0.036761580	-0.02414809
	Hippocampus	Random DTU	Elastic Net	-0.027494227	-0.03365646
	Hippocampus	Random DTU	Random Forest	-0.086672716	-0.09647761

```
[13]: df %>% filter(DTU == "DTU") %>% group_by(tissue, New_Type, model) %>%
      summarise(Mean=mean(Median), Median=median(Median), .groups = "keep") %>%
      filter(model == "Elastic Net")
```

	tissue <fct>	New_Type <fct>	model <fct>	Mean <dbl>	Median <dbl>
A grouped_df: 8 × 5	Caudate	DE DTU	Elastic Net	0.081422714	0.01311672
	Caudate	Random DTU	Elastic Net	-0.042723875	-0.04345409
	Dentate Gyrus	DE DTU	Elastic Net	0.005700645	-0.07592577
	Dentate Gyrus	Random DTU	Elastic Net	-0.124886491	-0.09891128
	DLPFC	DE DTU	Elastic Net	0.113791408	0.03734998
	DLPFC	Random DTU	Elastic Net	-0.032211432	-0.04758998
	Hippocampus	DE DTU	Elastic Net	0.087471515	0.01387694
	Hippocampus	Random DTU	Elastic Net	-0.027494227	-0.03365646

```
[14]: de100 %>% group_by(tissue, model) %>%
      summarise(Mean=mean(Median), .groups = "keep") %>% as.data.frame %>%
      pivot_wider(names_from="model", values_from="Mean")
```

	tissue <chr>	Elastic Net <dbl>	Random Forest <dbl>
A tibble: 4 × 3	Caudate	0.2283147	0.19525370
	Dentate Gyrus	0.1031959	0.08820506
	DLPFC	0.2610151	0.21962943
	Hippocampus	0.2452457	0.19171992

```
[15]: de100 %>% group_by(tissue, model) %>%
      summarise(Median=median(Median), .groups = "keep") %>% as.data.frame %>%
      pivot_wider(names_from="model", values_from="Median")
```

	tissue <chr>	Elastic Net <dbl>	Random Forest <dbl>
A tibble: 4 × 3	Caudate	0.21095822	0.1600187
	Dentate Gyrus	0.08515138	0.0644140
	DLPFC	0.23666764	0.1924569
	Hippocampus	0.19743658	0.1490395

Test if DE genes are significant more predictive than random genes

```
[16]: for(tissue in c("Caudate", "Dentate Gyrus", "DLPFC", "Hippocampus")){
  xx = de %>% filter(tissue == tissue)
  yy = rand %>% filter(tissue == tissue)
  tt = t.test(xx$Median, yy$Median, alternative = "greater")$p.value
  print(tt)
}
```

```
[1] 0
[1] 0
[1] 0
[1] 0
```

```
[17]: df %>% group_by(tissue) %>%
      do(fit = broom::tidy(lm(Median ~ Type, data=..))) %>%
      unnest(fit) %>% filter(term != '(Intercept)')
```

```
mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	tissue <fct>	term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>	p.bonf <dbl>
A tibble: 4 × 7	Caudate	TypeRandom	-0.08541266	0.002645699	-32.28359	3.751268e-219	1.5005
	Dentate Gyrus	TypeRandom	-0.15560514	0.007755310	-20.06434	2.680821e-84	1.0723
	DLPFC	TypeRandom	-0.10364360	0.002945424	-35.18801	5.020572e-257	2.0082
	Hippocampus	TypeRandom	-0.07521952	0.002425112	-31.01693	4.660740e-203	1.8642

```
[18]: df %>% filter(Type == "DE") %>% group_by(tissue) %>%
  do(fit = broom::tidy(lm(Median ~ New_Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	tissue <fct>	term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>	p.b <dbl>
A tibble: 4 × 7	Caudate	New_TypeDE DTU	0.04130844	0.006818805	6.058018	1.463818e-09	5.8
	Dentate Gyrus	New_TypeDE DTU	0.06459084	0.026013145	2.483008	1.313335e-02	5.2
	DLPFC	New_TypeDE DTU	0.08152278	0.008870552	9.190271	5.468858e-20	2.1
	Hippocampus	New_TypeDE DTU	0.06360128	0.008007479	7.942734	2.356165e-15	9.4

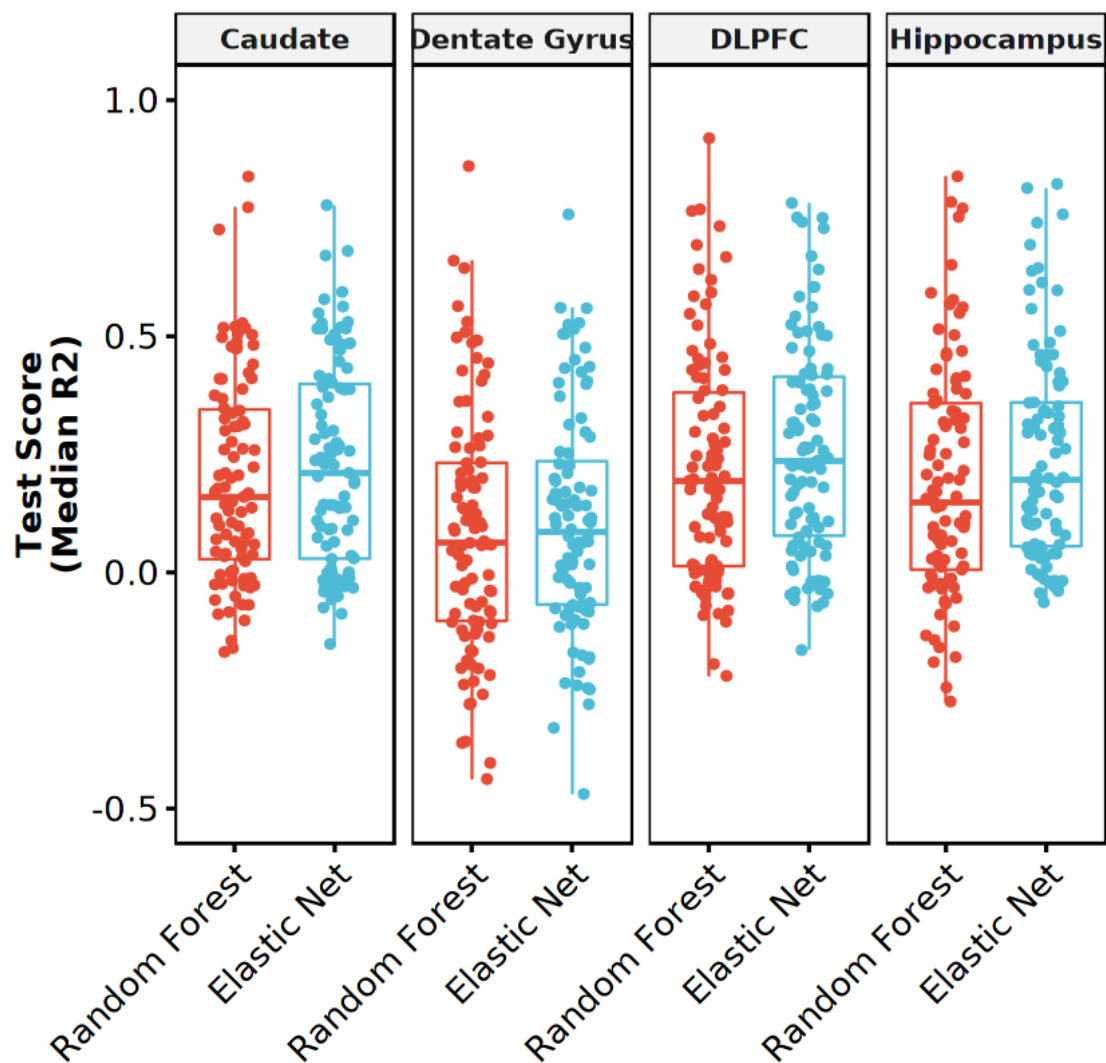
```
[19]: df %>% filter(Type == "Random") %>% group_by(tissue) %>%
  do(fit = broom::tidy(lm(Median ~ New_Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	tissue <fct>	term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>
A tibble: 4 × 7	Caudate	New_TypeRandom DTU	-0.01356067	0.005795044	-2.340045	0.01931459
	Dentate Gyrus	New_TypeRandom DTU	0.03810179	0.033865163	1.125103	0.26071874
	DLPFC	New_TypeRandom DTU	0.02228672	0.007857250	2.836452	0.00457848
	Hippocampus	New_TypeRandom DTU	0.01540211	0.006287738	2.449547	0.01433259

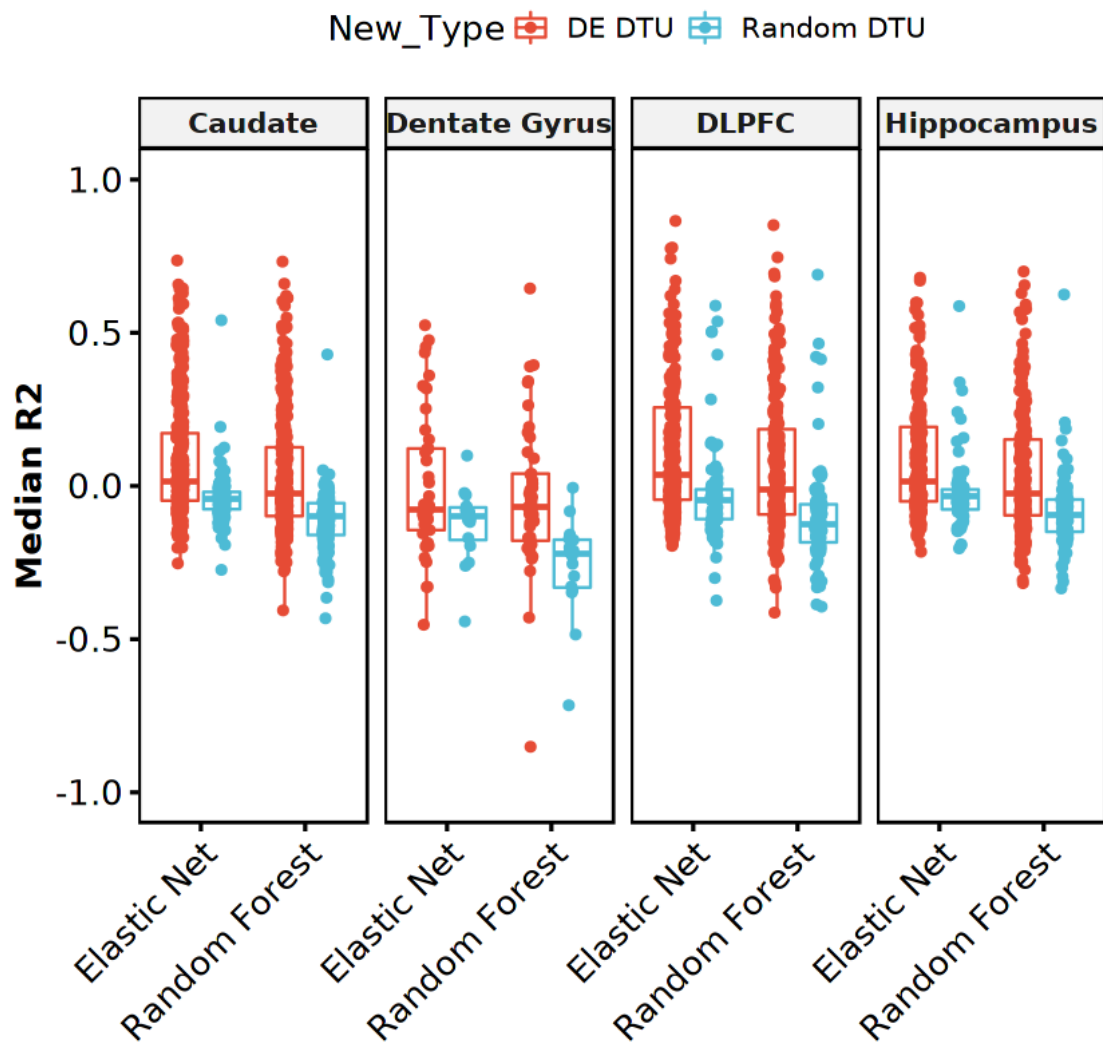
1.2.3 Plot

Boxplots

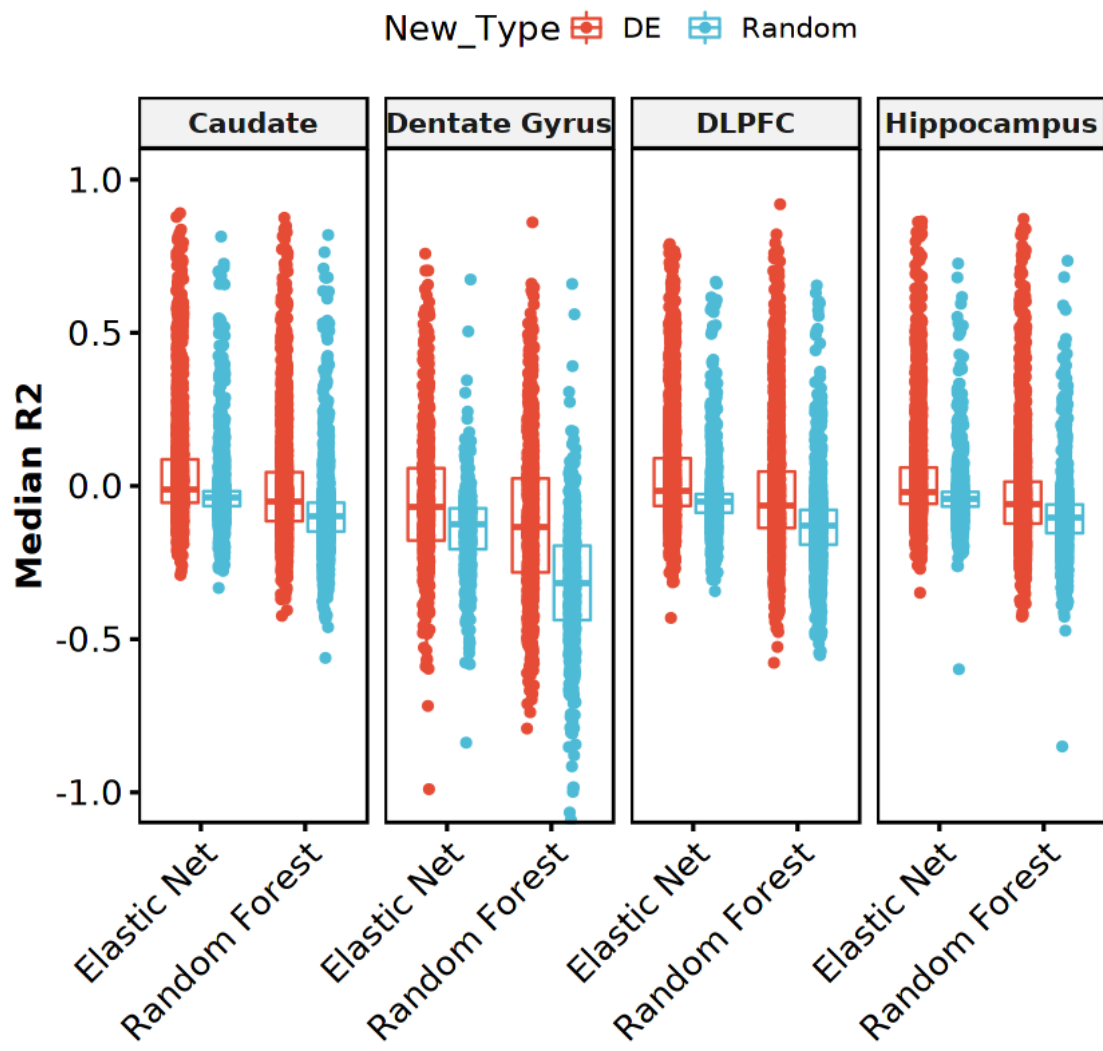
```
[20]: bxp = de100 %>%
  ggboxplot(x="model", y="Median", color="model", add="jitter",
    facet.by="tissue", legend="none", palette="npg",
    ylim=c(-0.5, 1), ylab="Test Score\n(Median R2)",
    xlab="", ncol=4, panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_top100", 9, 4)
bxp
```



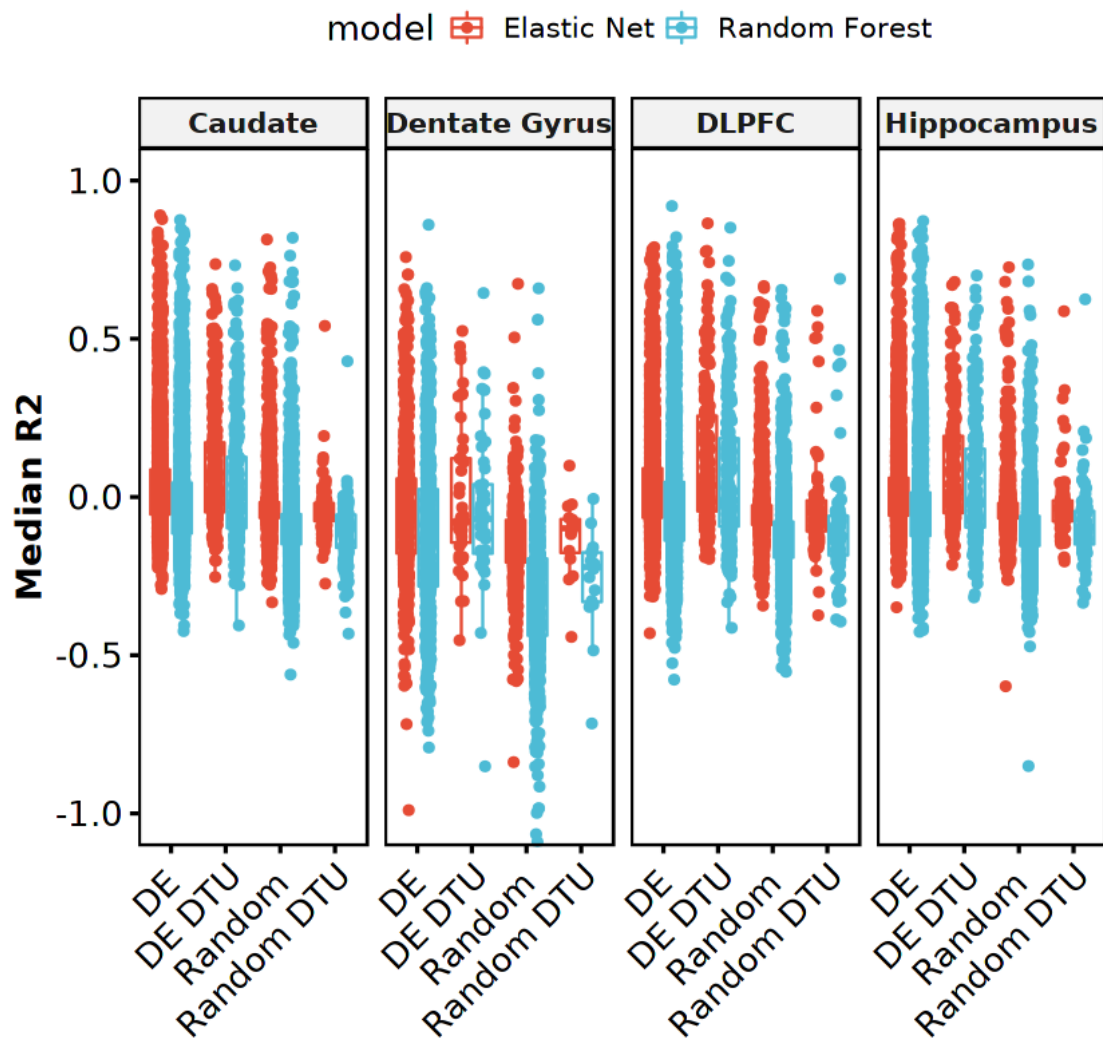
```
[21]: bxp = df %>% filter(DTU == "DTU") %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_byType_dtu", 10, 5)
bxp
```

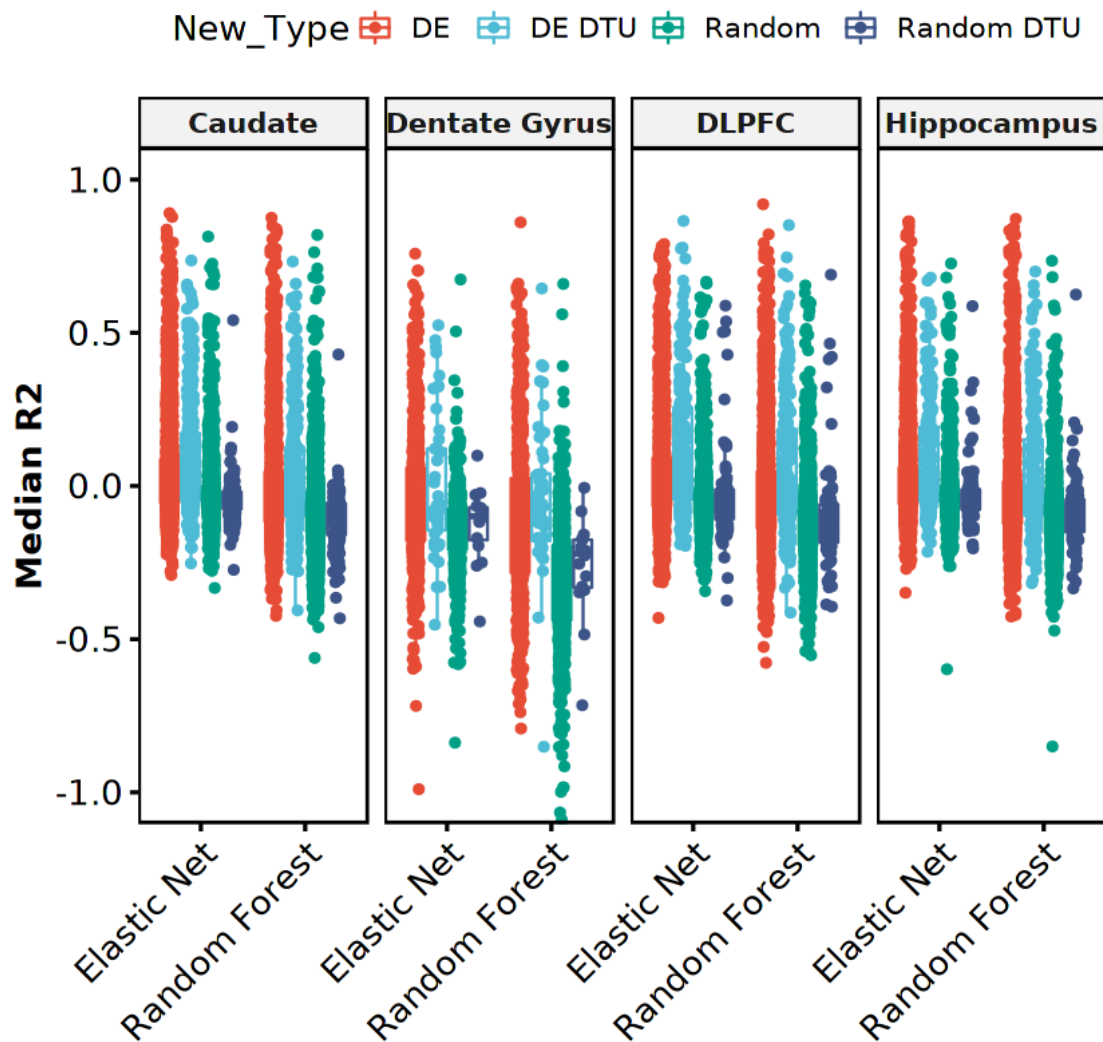
```
[22]: bxp = df %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_byType_de", 10, 5)
bxp
```



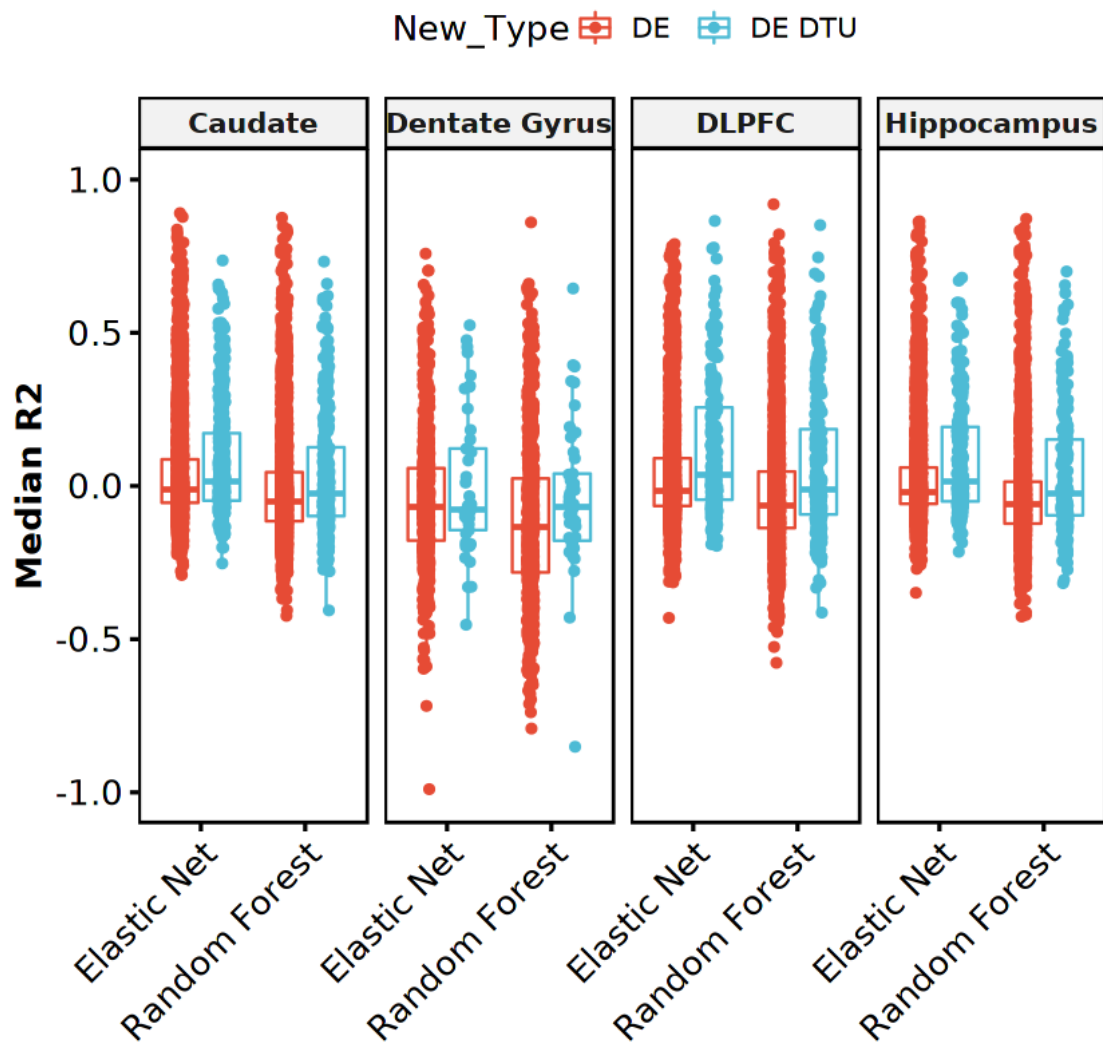
```
[23]: bxp = df %>% ggboxplot(x="New_Type", y="Median", color="model", add="jitter",
  facet.by="tissue", palette="npg", ylim=c(-1, 1),
  ylab="Median R2", xlab="", ncol=4,
  panel.labs.font=list(face='bold'),
  ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_byModel", 10, 5)
bxp
```



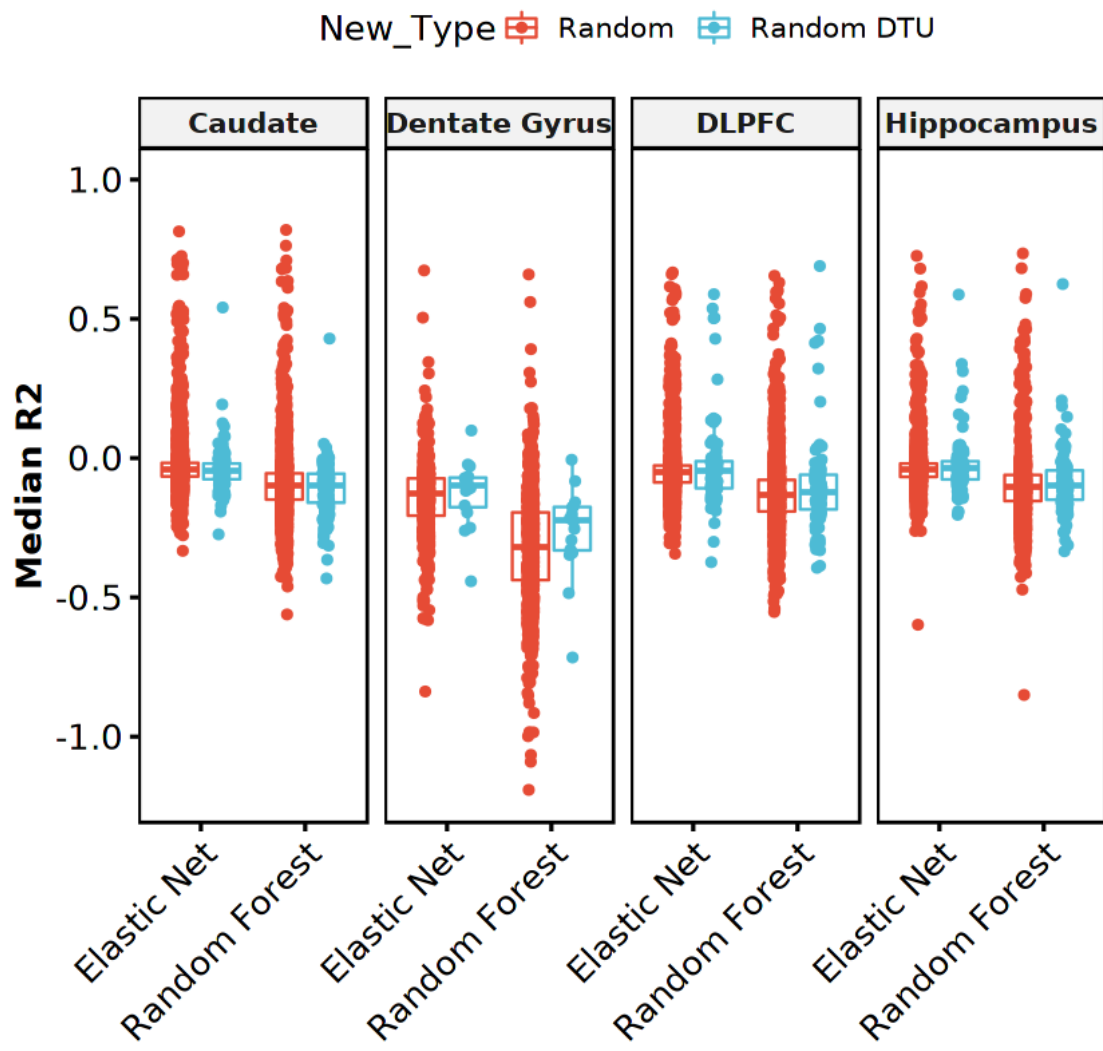
```
[24]: bxp = df %>% ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
  facet.by="tissue", palette="npg", ylim=c(-1, 1),
  ylab="Median R2", xlab="", ncol=4,
  panel.labs.font=list(face='bold'),
  ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_byType", 10, 5)
bxp
```



```
[25]: bxp = df %>% filter(Type == "DE") %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_deg", 10, 5)
bxp
```

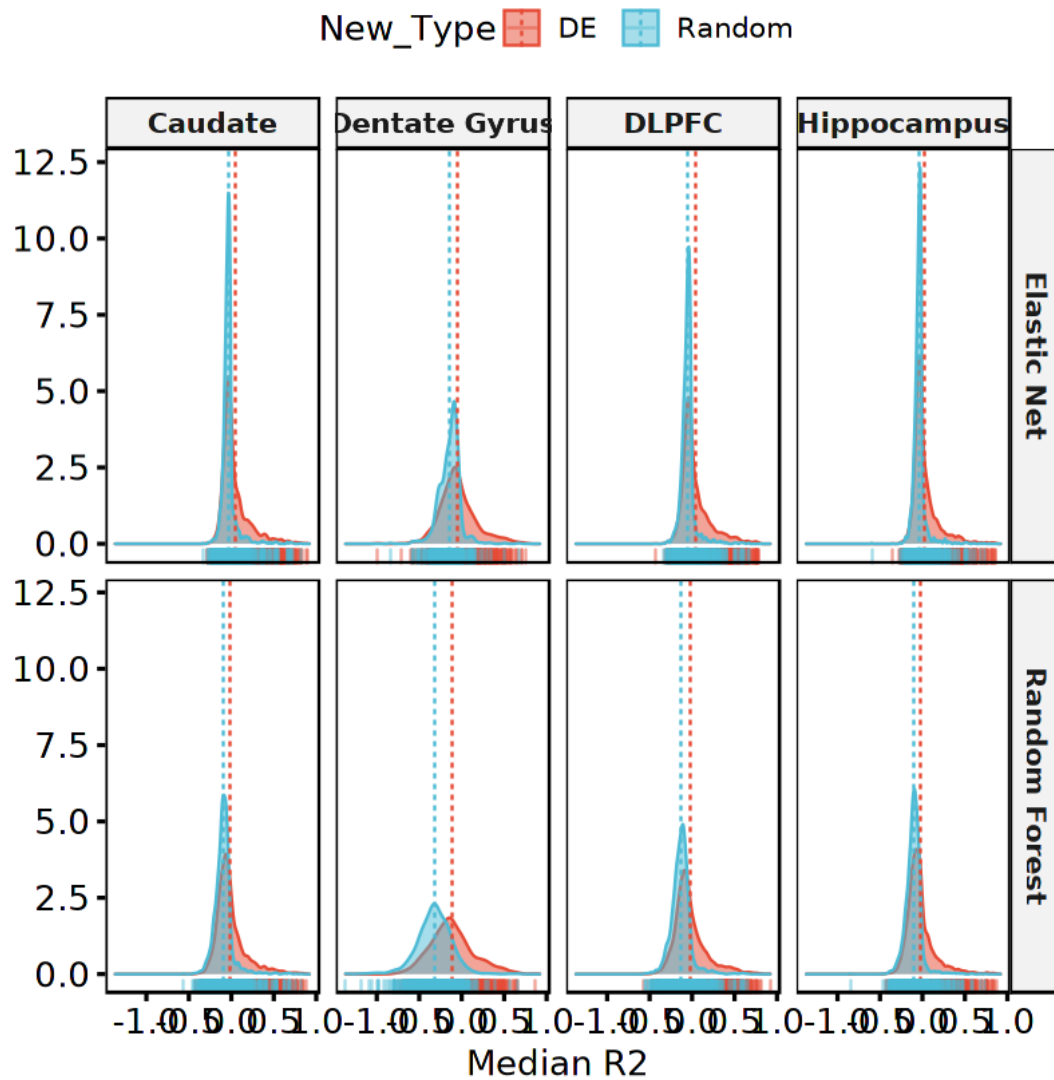


```
[26]: bxp = df %>% filter(Type == "Random") %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1.2, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "summary_boxplots_r2_random", 10, 5)
bxp
```

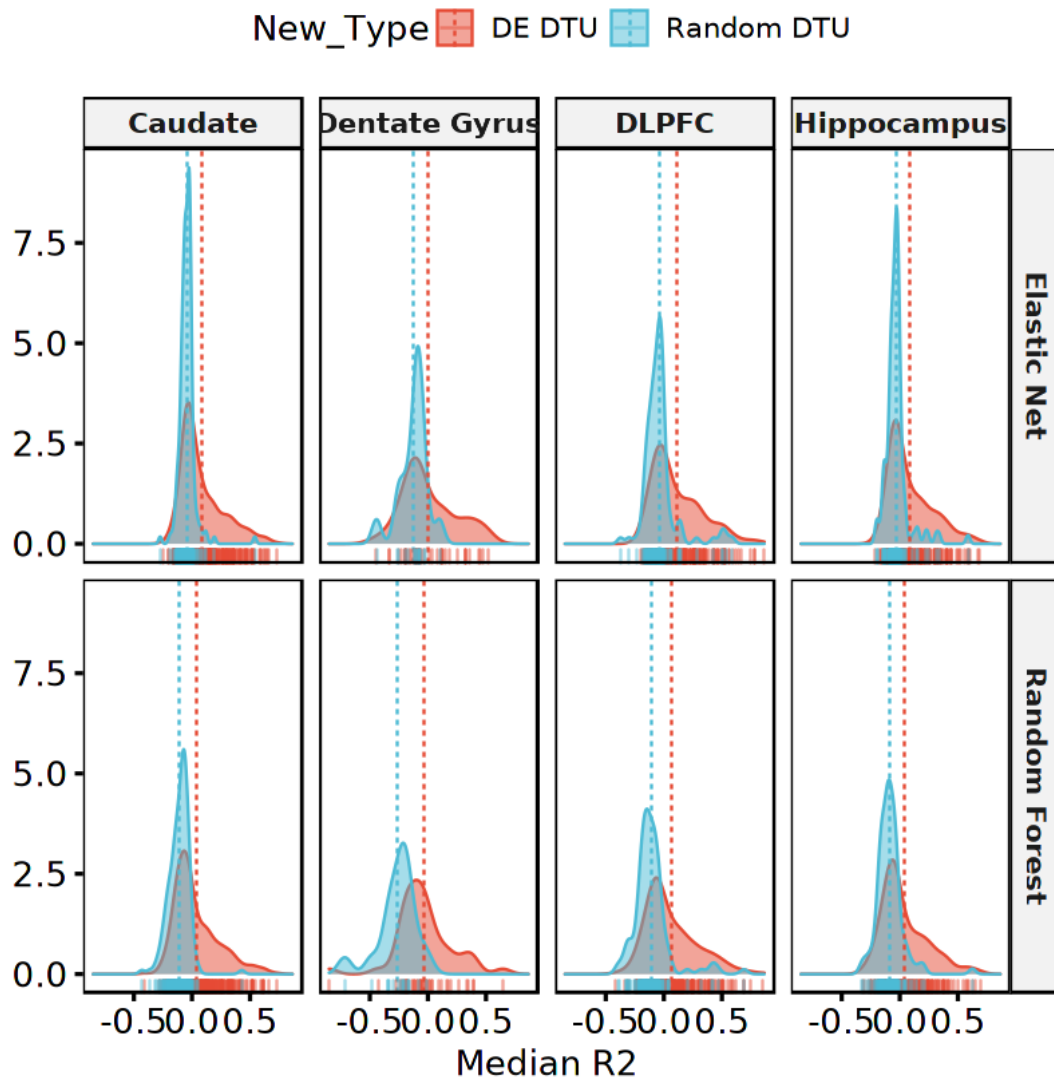


Distribution

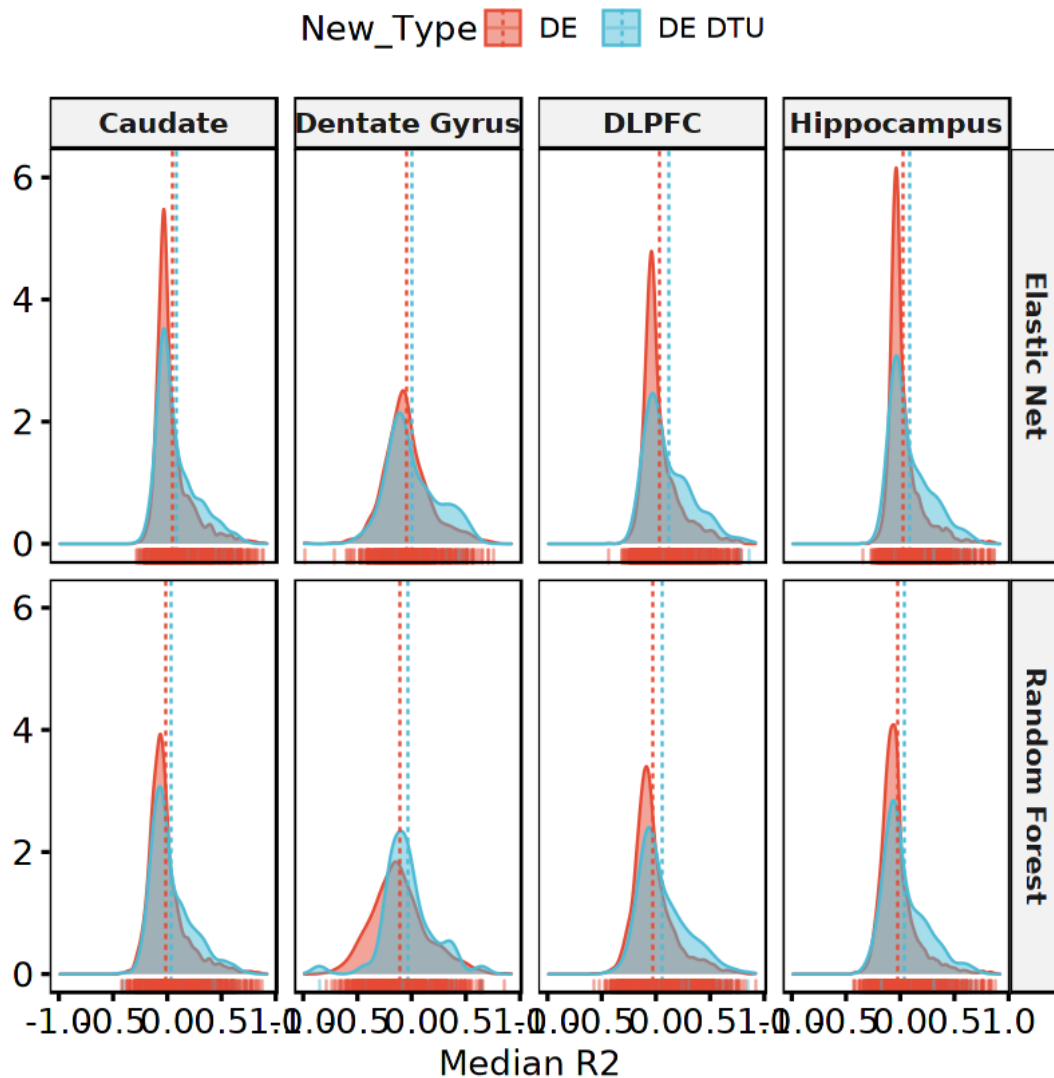
```
[27]: freq = df %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggdensity(x = "Median", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("model", "tissue"), xlab="Median R2",
    panel.labs.font=list(face='bold'), ylab="",
    ggtheme=theme_pubr(base_size=15, border=TRUE))
save_plot(freq, "summary_distribution_r2_de", 10, 5)
freq
```



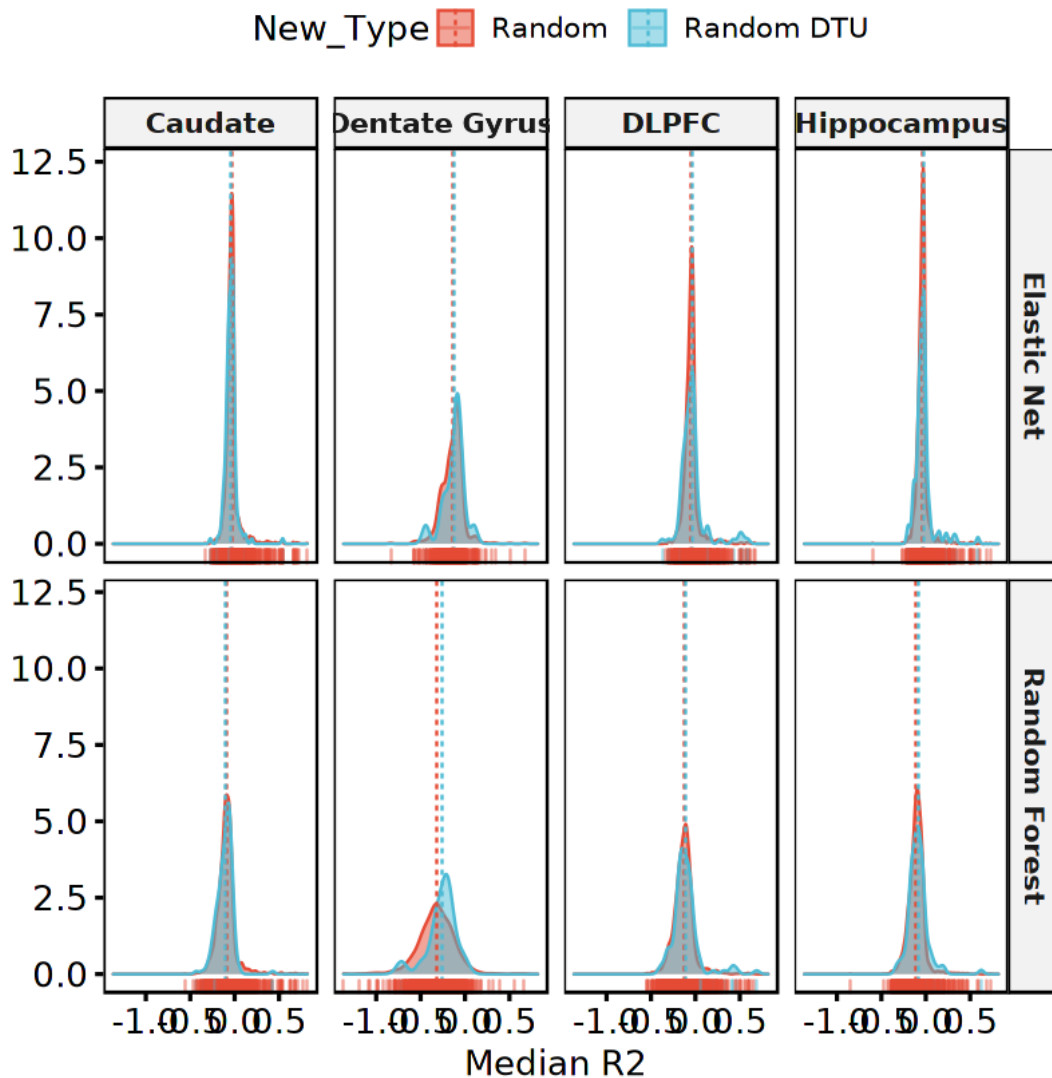
```
[28]: freq = df %>% filter(DTU == "DTU") %>%
  ggdensity(x = "Median", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("model", "tissue"), xlab="Median R2",
    panel.labs.font=list(face='bold'), ylab="",
    ggtheme=theme_pubr(base_size=15, border=TRUE))
save_plot(freq, "summary_distribution_r2_dtu", 10, 5)
freq
```



```
[29]: freq = df %>% filter(Type == "DE") %>%
  ggdensity(x = "Median", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("model", "tissue"), xlab="Median R2",
    panel.labs.font=list(face='bold'), ylab="",
    ggtheme=theme_pubr(base_size=15, border=TRUE))
save_plot(freq, "summary_distribution_r2_degs", 10, 5)
freq
```

```
[30]: freq = df %>% filter(Type == "Random") %>%
  ggdensity(x = "Median", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("model", "tissue"), xlab="Median R2",
    panel.labs.font=list(face='bold'), ylab="",
    ggtheme=theme_pubr(base_size=15, border=TRUE))
save_plot(freq, "summary_distribution_r2_random", 10, 5)
freq
```



1.3 Explained variance with partial r2

1.3.1 Load data

```
[31]: de2 = data.table::fread("../de_genes/partial_r2/rf_partial_r2_metrics.tsv")
      ↪ %>%
      mutate(Type = "DE") %>% left_join(degs, by=c("Tissue"="tissue",
      ↪ "Geneid"="Feature")) %>%
      as.data.frame %>% mutate(New_Type = paste(Type, replace_na(DTU, ""))) %>%
      mutate_if(is.character, as.factor)
de2 %>% head(2)
```

		Geneid	N_Features	Partial_R2	Full_R2	Reduced_R2	Tissue
		<fct>	<int>	<dbl>	<dbl>	<dbl>	<fct>
A data.frame: 2 × 11	1	ENSG00000003249.13	33	0.24367297	171.4880	226.7379	Caudate
	2	ENSG00000003509.15	2	0.01313991	225.4272	228.4288	Caudate

```
[32]: de100_v2 = data.table::fread("../de_genes/partial_r2/rf_partial_r2_metrics.
  ↪tsv") %>%
  mutate(Type = "DE") %>% inner_join(top100, by=c("Tissue"="tissue",
  ↪"Geneid"="Feature"))
de100_v2 %>% dim
de100_v2 %>% head(2)
```

1. 382 2. 9

		Geneid	N_Features	Partial_R2	Full_R2	Reduced_R2	Tissue	Ty
		<chr>	<int>	<dbl>	<dbl>	<dbl>	<chr>	<c
A data.table: 2 × 9		ENSG000000014824.13	3	0.09679049	167.7481	185.7245	Caudate	DE
		ENSG000000034053.14	31	0.13785014	169.7030	196.8370	Caudate	DE

```
[33]: rand2 = data.table::fread("../random_genes/partial_r2/rf_partial_r2_metrics.
  ↪tsv") %>%
  mutate(Type = "Random") %>% left_join(random, by=c("Tissue"="tissue",
  ↪"Geneid"="Feature")) %>%
  as.data.frame %>% mutate(New_Type = paste(Type, replace_na(DTU, ""))) %>%
  mutate_if(is.character, as.factor)
rand2 %>% head(2)
```

		Geneid	N_Features	Partial_R2	Full_R2	Reduced_R2	Tissue
		<fct>	<int>	<dbl>	<dbl>	<dbl>	<fct>
A data.frame: 2 × 11	1	ENSG000000001084.10	18	0.173657975	196.5602	237.8679	Caudate
	2	ENSG000000001630.15	2	0.006767595	233.4418	235.0324	Caudate

```
[34]: df2 = bind_rows(de2, rand2)
df2 %>% head(2)
```

		Geneid	N_Features	Partial_R2	Full_R2	Reduced_R2	Tissue
		<fct>	<int>	<dbl>	<dbl>	<dbl>	<fct>
A data.frame: 2 × 11	1	ENSG000000003249.13	33	0.24367297	171.4880	226.7379	Caudate
	2	ENSG000000003509.15	2	0.01313991	225.4272	228.4288	Caudate

1.3.2 Summary

```
[35]: df2 %>% group_by(Tissue, Type) %>%
  summarise(Mean=mean(Partial_R2), Median=median(Partial_R2),
  Std=sd(Partial_R2), .groups = "keep")
```

	Tissue <fct>	Type <fct>	Mean <dbl>	Median <dbl>	Std <dbl>
A grouped_df: 8 × 5	Caudate	DE	0.12115161	0.06459714	0.14915843
	Caudate	Random	0.05462609	0.01699683	0.09628326
	Dentate Gyrus	DE	0.20111520	0.14450007	0.18599017
	Dentate Gyrus	Random	0.12287941	0.07247299	0.14869899
	DLPFC	DE	0.12918919	0.07083068	0.15498073
	DLPFC	Random	0.05820650	0.02022916	0.09648807
	Hippocampus	DE	0.10386637	0.04787038	0.14271472
	Hippocampus	Random	0.04587896	0.01430467	0.08136296

```
[36]: de100_v2 %>% group_by(Tissue) %>%
      summarise(Mean=mean(Partial_R2), Median=median(Partial_R2),
                  Std=sd(Partial_R2), .groups = "keep")
```

	Tissue <chr>	Mean <dbl>	Median <dbl>	Std <dbl>
A grouped_df: 4 × 4	Caudate	0.2041160	0.1399830	0.1837352
	Dentate Gyrus	0.2727508	0.2280635	0.2120016
	DLPFC	0.2456374	0.2091248	0.1972761
	Hippocampus	0.2156581	0.1593486	0.1900711

```
[37]: df2 %>% group_by(Tissue, New_Type) %>%
      summarise(Mean=mean(Partial_R2), Median=median(Partial_R2),
                  Std=sd(Partial_R2), .groups = "keep")
```

	Tissue <fct>	New_Type <fct>	Mean <dbl>	Median <dbl>	Std <dbl>
A grouped_df: 16 × 5	Caudate	DE	0.11714984	0.06256508	0.14628216
	Caudate	DE DTU	0.14799147	0.08160512	0.16489588
	Caudate	Random	0.05484748	0.01691896	0.09715149
	Caudate	Random DTU	0.05096409	0.02011857	0.08074849
	Dentate Gyrus	DE	0.19965350	0.14300394	0.18680034
	Dentate Gyrus	DE DTU	0.22582812	0.20626251	0.17185173
	Dentate Gyrus	Random	0.12350021	0.07247299	0.14955208
	Dentate Gyrus	Random DTU	0.09354702	0.05169564	0.09920611
	DLPFC	DE	0.12382915	0.06713538	0.14951782
	DLPFC	DE DTU	0.18373550	0.11131387	0.19450281
	DLPFC	Random	0.05714081	0.02007225	0.09325817
	DLPFC	Random DTU	0.08482760	0.02604147	0.15528999
	Hippocampus	DE	0.09969056	0.04681746	0.13786980
	Hippocampus	DE DTU	0.15465271	0.07376830	0.18509903
	Hippocampus	Random	0.04519483	0.01393831	0.08053527
	Hippocampus	Random DTU	0.06308005	0.02792730	0.09885973

1.3.3 Significance

```
[38]: ## Test if DE genes are significant more predictive than random genes
for(tissue in c("Caudate", "Dentate Gyrus", "DLPFC", "Hippocampus")){
  xx = de2 %>% filter(Tissue == tissue)
  yy = rand2 %>% filter(Tissue == tissue)
  tt = t.test(xx$Partial_R2, yy$Partial_R2, alternative = "greater")$p.value
  print(tt)
}
```

```
[1] 8.96649e-86
[1] 1.187035e-19
[1] 1.037743e-85
[1] 7.833267e-76
```

```
[39]: df2 %>% group_by(Tissue) %>%
  do(fit = broom::tidy(lm(Partial_R2 ~ Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	Tissue <fct>	term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>	p.bonf <dbl>
A tibble: 4 × 7	Caudate	TypeRandom	-0.06652552	0.003348234	-19.868842	5.677769e-85	2.2711
	Dentate Gyrus	TypeRandom	-0.07823578	0.008574660	-9.124068	2.200512e-19	8.8020
	DLPFC	TypeRandom	-0.07098269	0.003557188	-19.954719	1.797423e-85	7.1896
	Hippocampus	TypeRandom	-0.05798741	0.003122240	-18.572373	9.228185e-75	3.6912

```
[40]: df2 %>% filter(Type == "DE") %>% group_by(Tissue) %>%
  do(fit = broom::tidy(lm(Partial_R2 ~ New_Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	Tissue <fct>	term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>	p <dbl>
A tibble: 4 × 7	Caudate	New_TypeDE DTU	0.03084163	0.008271409	3.7287033	1.961946e-04	7
	Dentate Gyrus	New_TypeDE DTU	0.02617462	0.029193712	0.8965843	3.702217e-01	1
	DLPFC	New_TypeDE DTU	0.05990636	0.010465602	5.7241197	1.156213e-08	4
	Hippocampus	New_TypeDE DTU	0.05496215	0.010050855	5.4684055	4.935523e-08	1

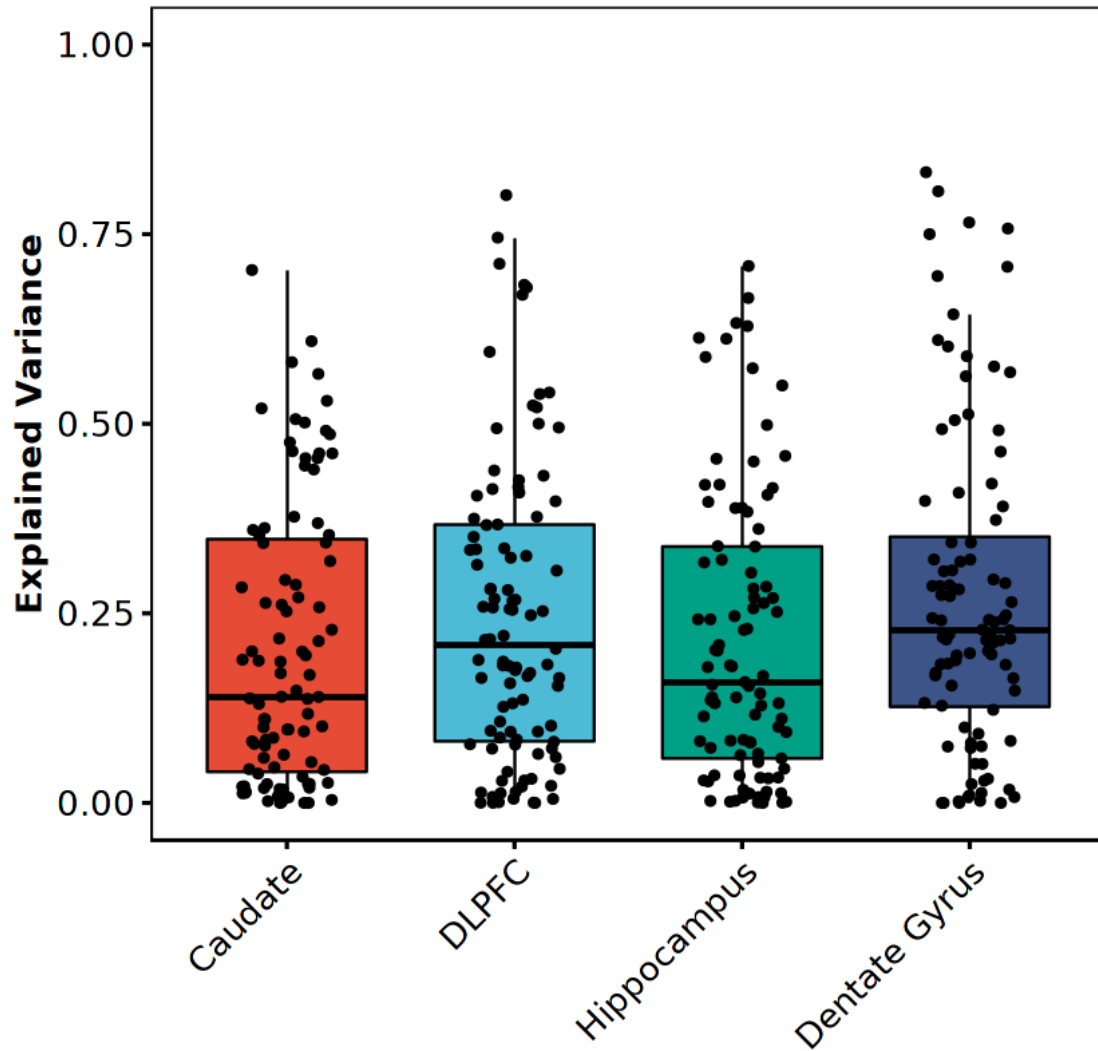
```
[41]: df2 %>% filter(Type == "Random") %>% group_by(Tissue) %>%
  do(fit = broom::tidy(lm(Partial_R2 ~ New_Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	Tissue <fct>	term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>
A tibble: 4 × 7	Caudate	New_TypeRandom DTU	-0.003883392	0.007864253	-0.4938030	0.621484
	Dentate Gyrus	New_TypeRandom DTU	-0.029953185	0.037574957	-0.7971582	0.425604
	DLPFC	New_TypeRandom DTU	0.027686784	0.009778135	2.8314994	0.004668
	Hippocampus	New_TypeRandom DTU	0.017885218	0.008090846	2.2105498	0.027149

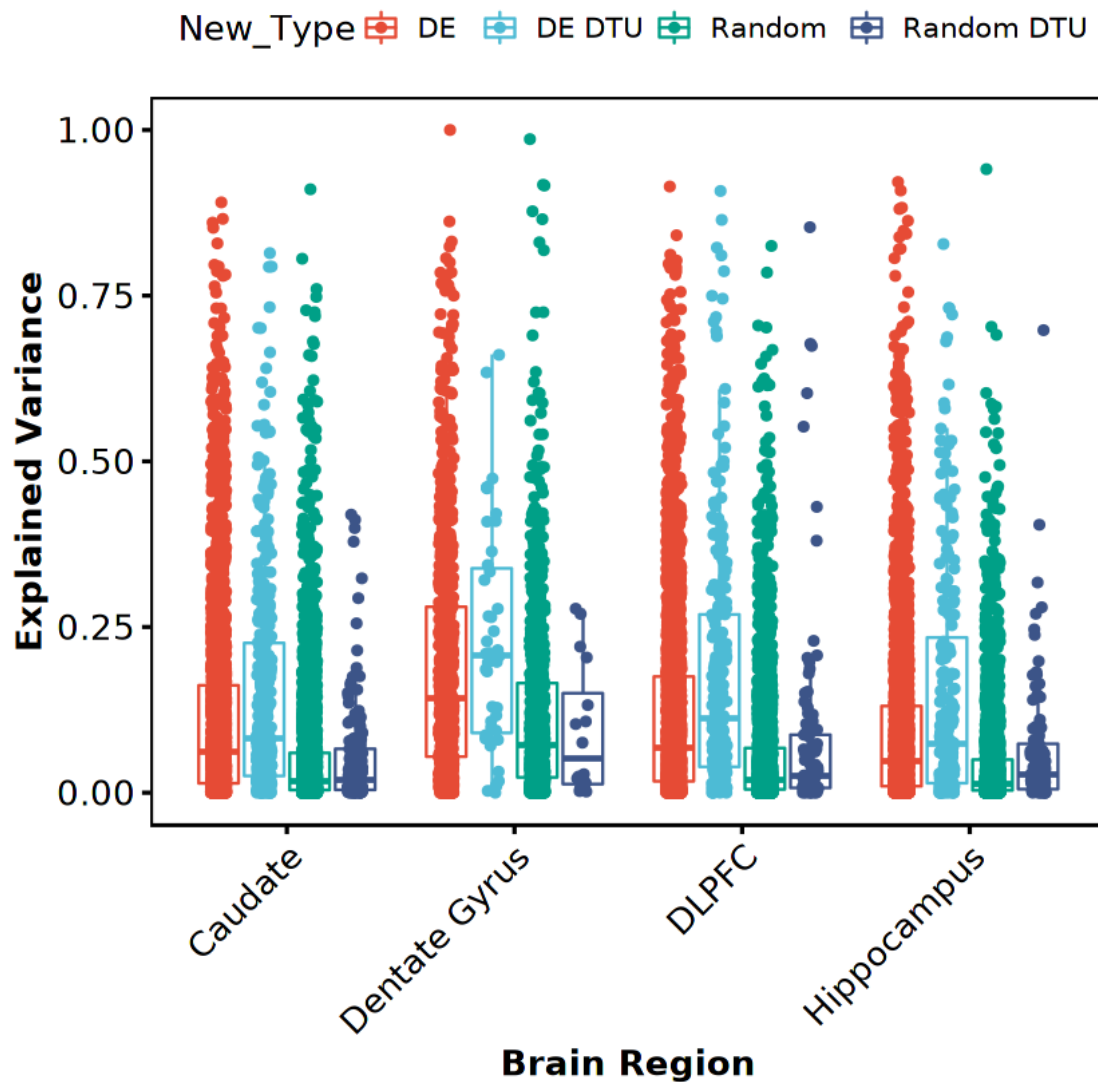
1.3.4 Plot

Boxplots

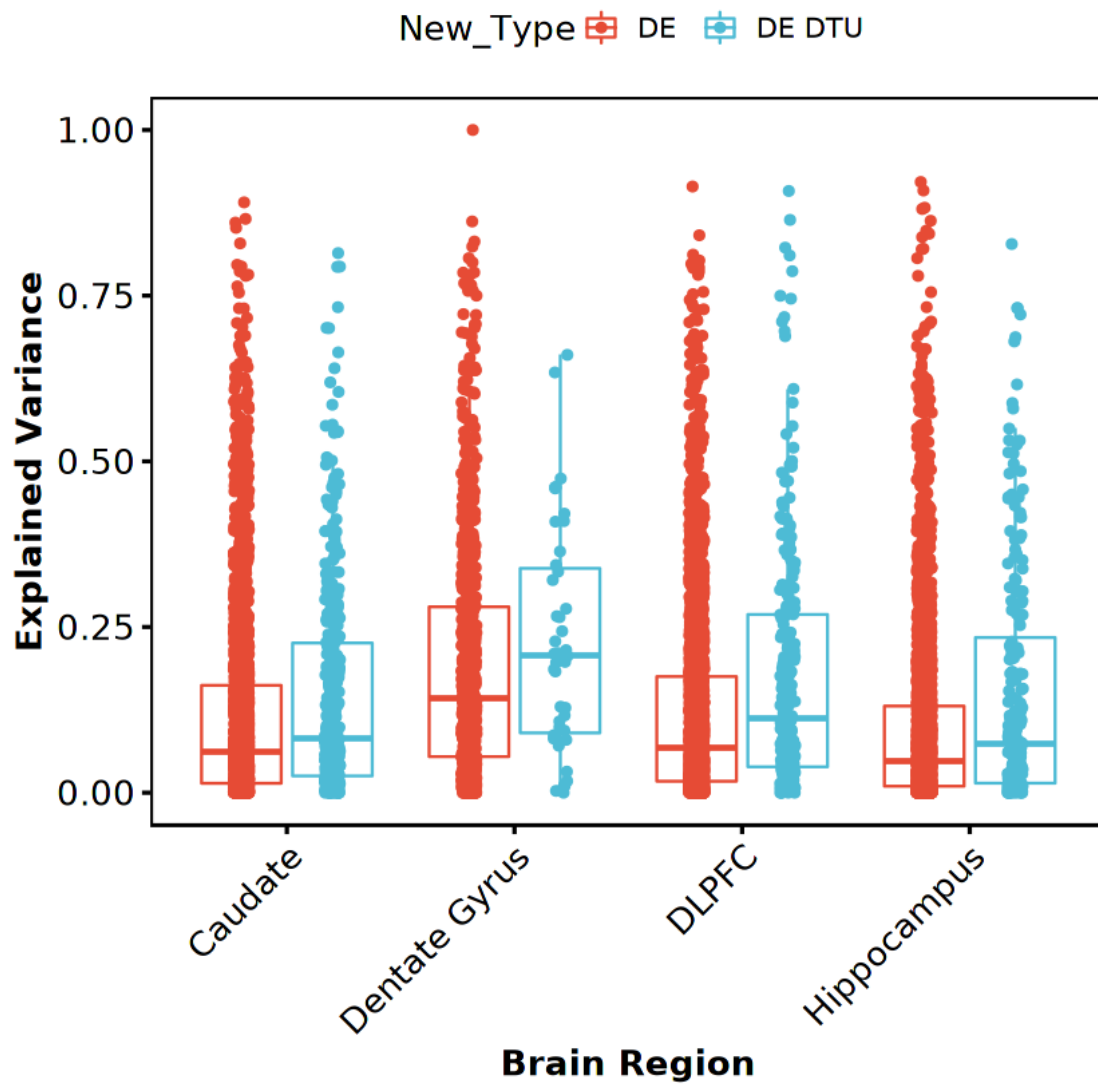
```
[42]: bxp = de100_v2 %>%
  ggboxplot(x="Tissue", y="Partial_R2", fill="Tissue", add="jitter",
    palette="npg", legend="none", ylim=c(0, 1), xlab="", ncol=4,
    ylab="Explained Variance", panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "explained_variance_boxplots_top100", 6, 5)
bxp
```



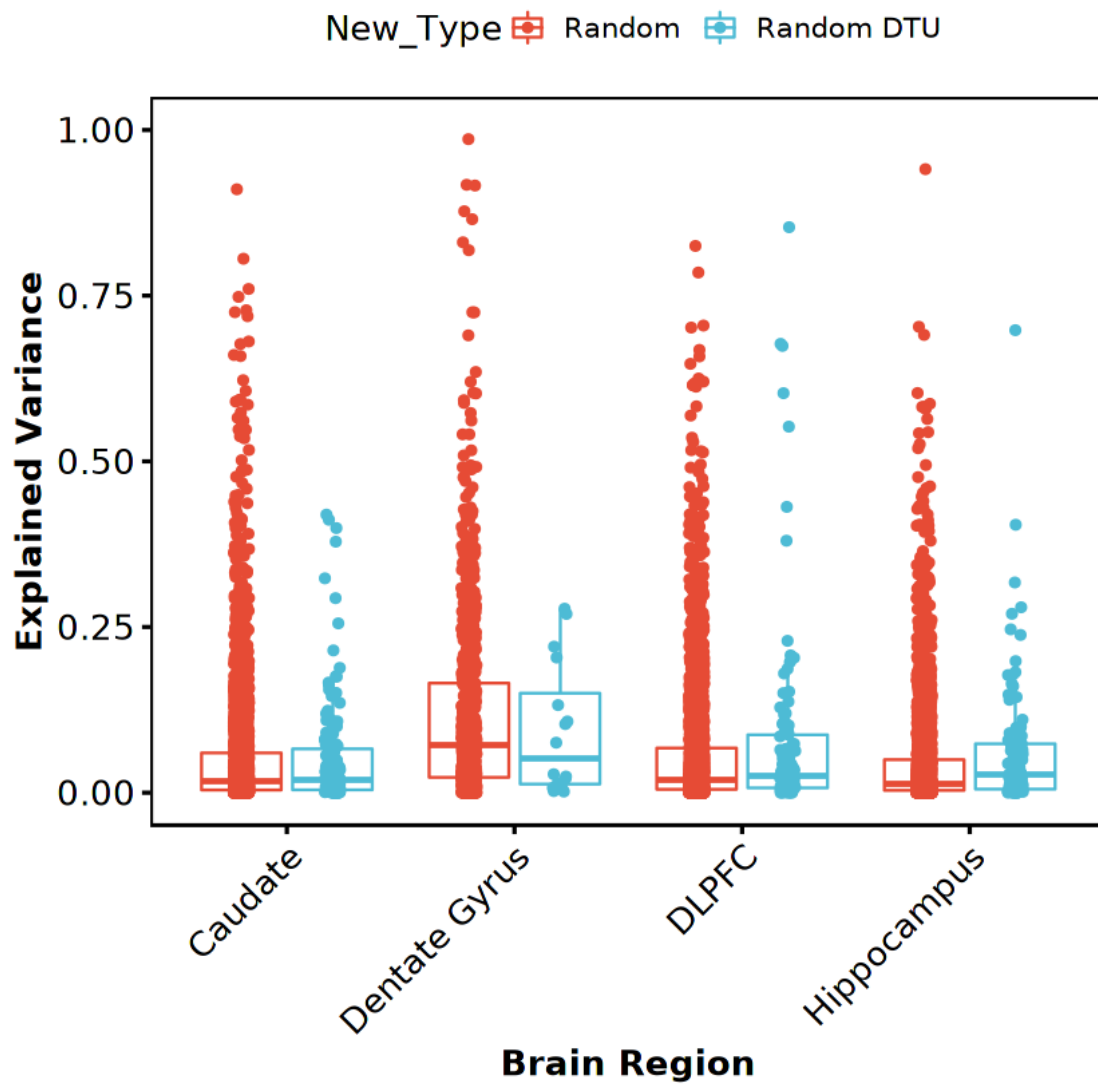
```
[43]: bxp = df2 %>% #filter(Type == "DE") %>%
      ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
               palette="npg", ylim=c(0, 1), ylab="Explained Variance",
               xlab="Brain Region", ncol=4,
               ggtheme=theme_pubr(base_size=15, border=TRUE)) +
      rotate_x_text(45) + font("xy.title", face="bold")
      save_plot(bxp, "explained_variance_boxplot_all", 7, 5.5)
      bxp
```



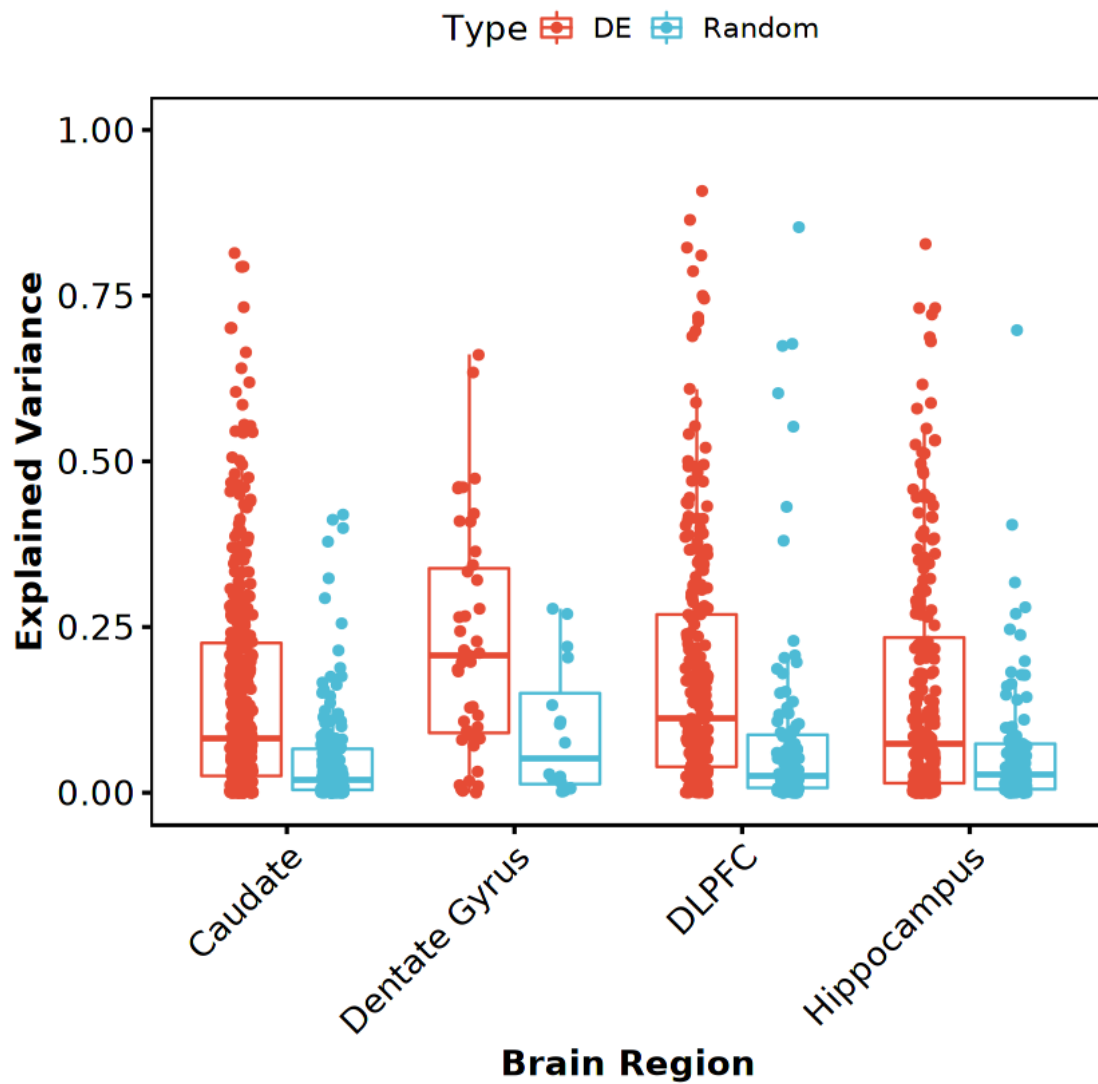
```
[44]: bxp = df2 %>% filter(Type == "DE") %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "explained_variance_boxplot_de", 6, 5.5)
bxp
```

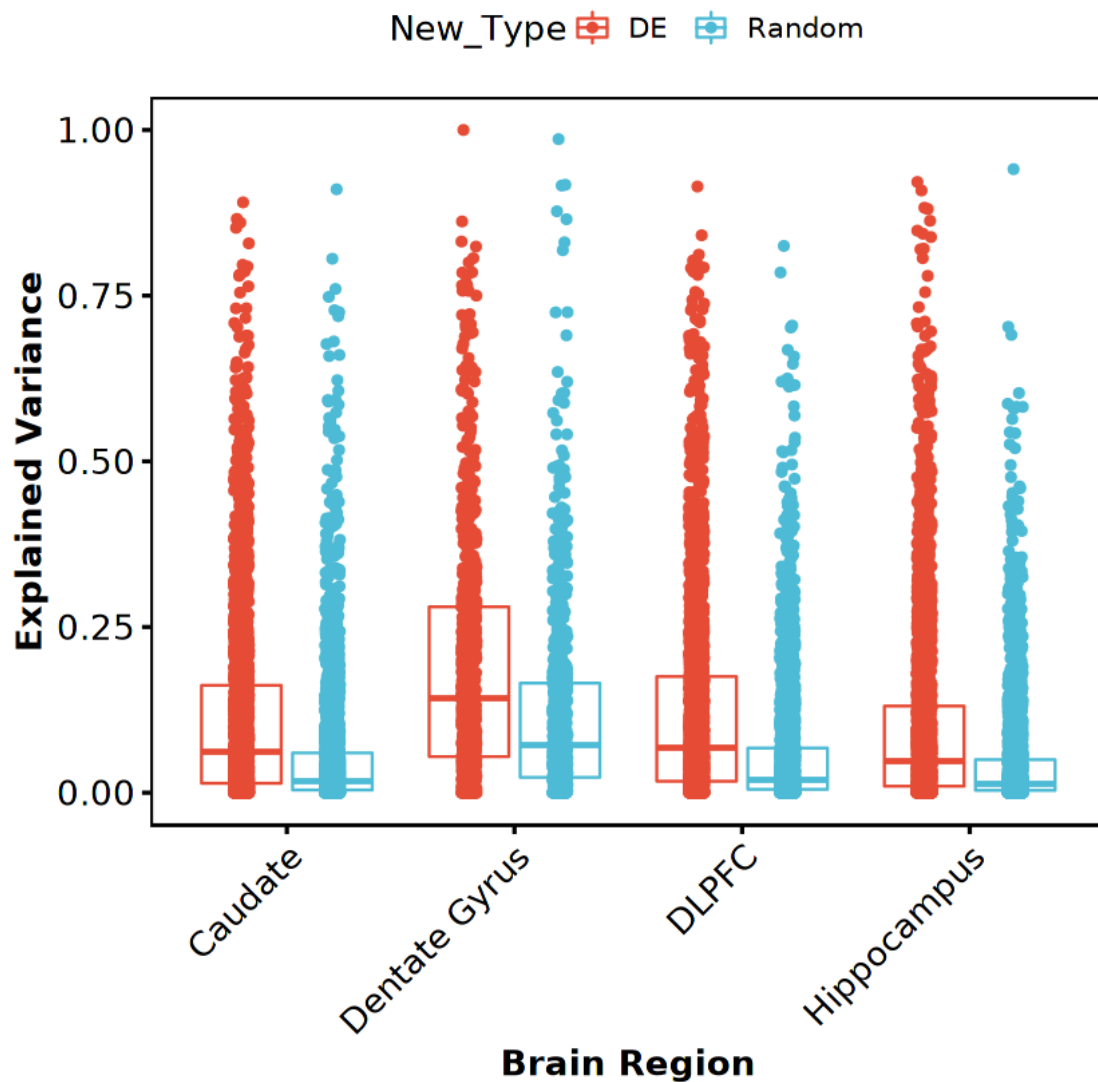
```
[45]: bxp = df2 %>% filter(Type == "Random") %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "explained_variance_boxplot_random", 6, 5.5)
bxp
```



```
[46]: bxp = df2 %>% filter(DTU == "DTU") %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "explained_variance_boxplot_dtu", 6, 5.5)
bxp
```

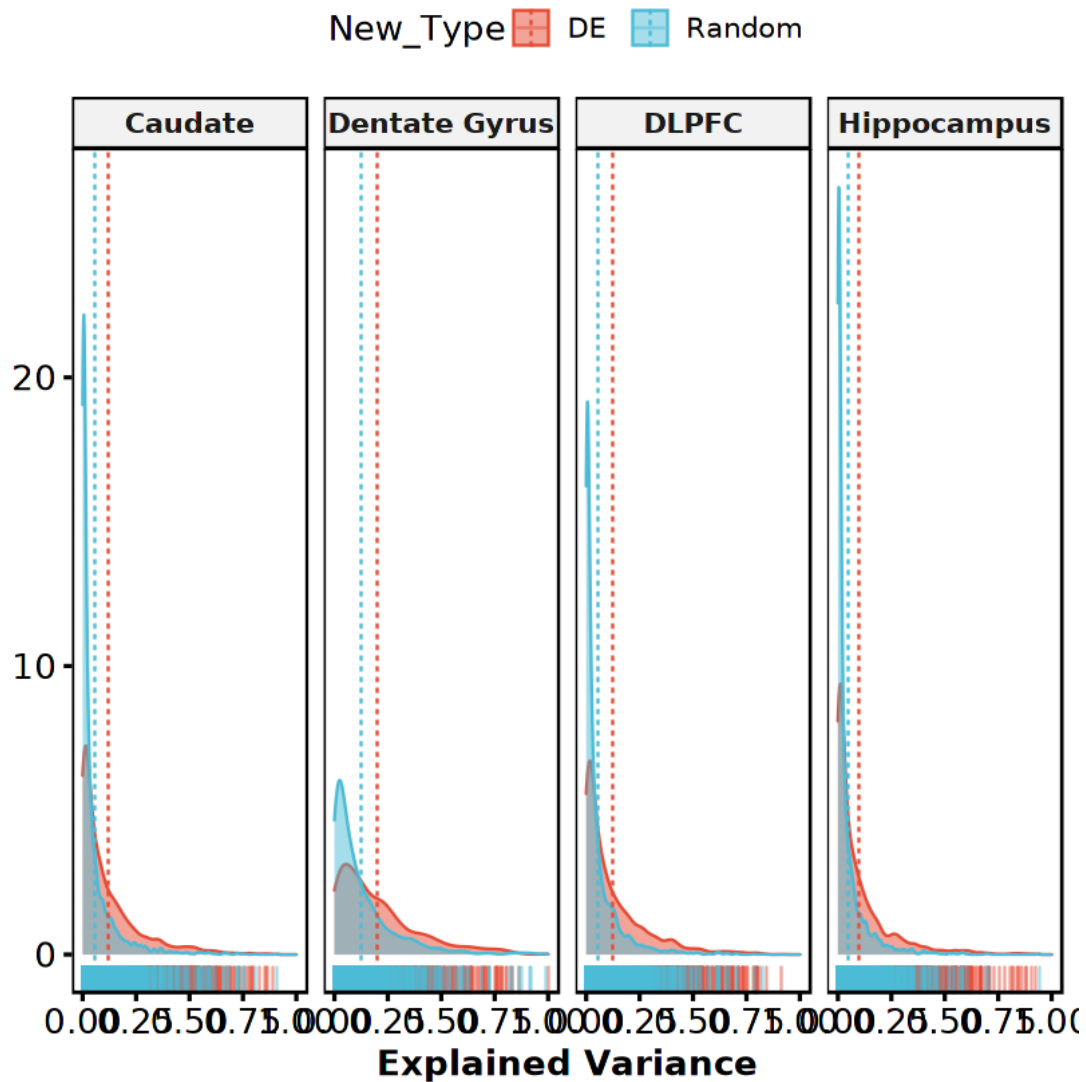


```
[47]: bxp = df2 %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title", face="bold")
save_plot(bxp, "explained_variance_boxplot", 6, 5.5)
bxp
```

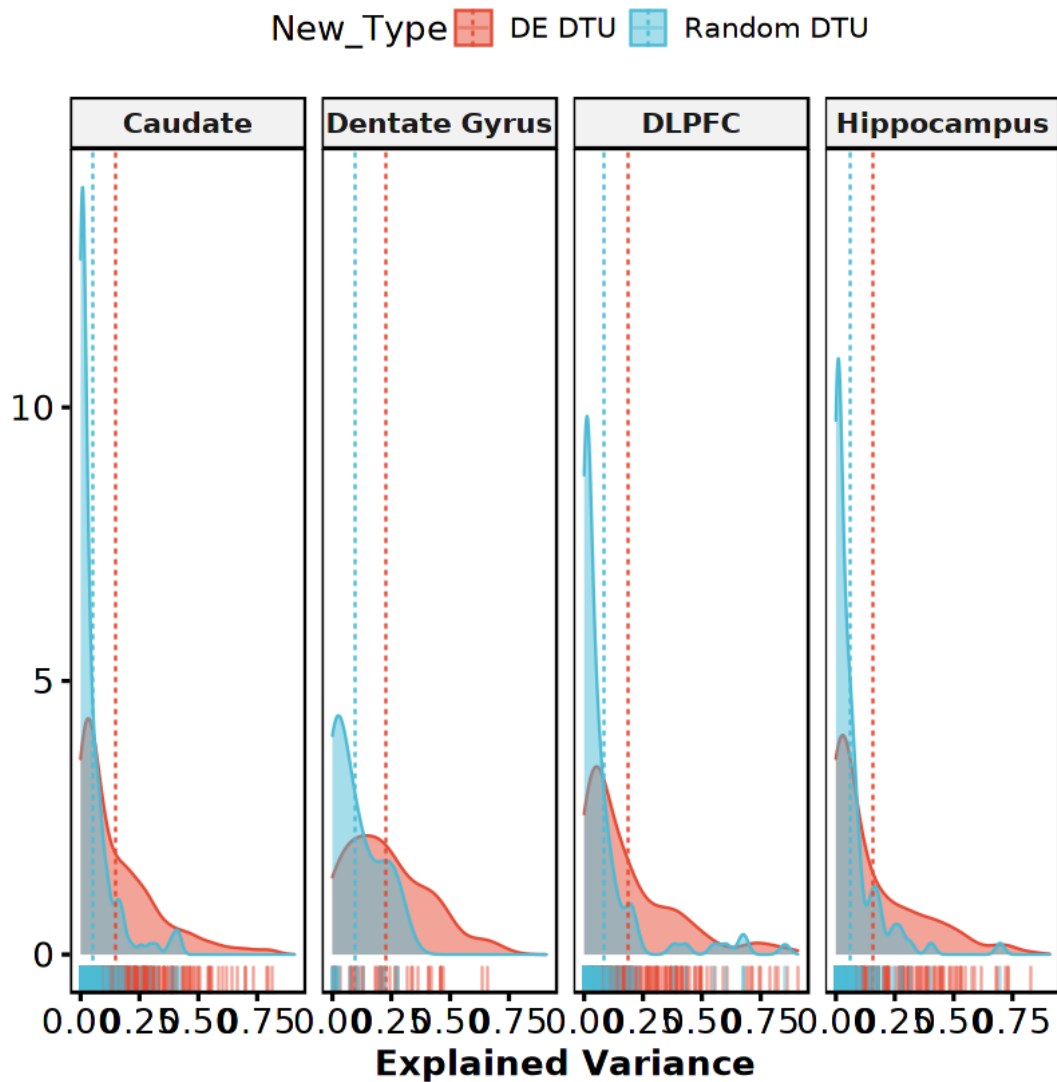


Distribution

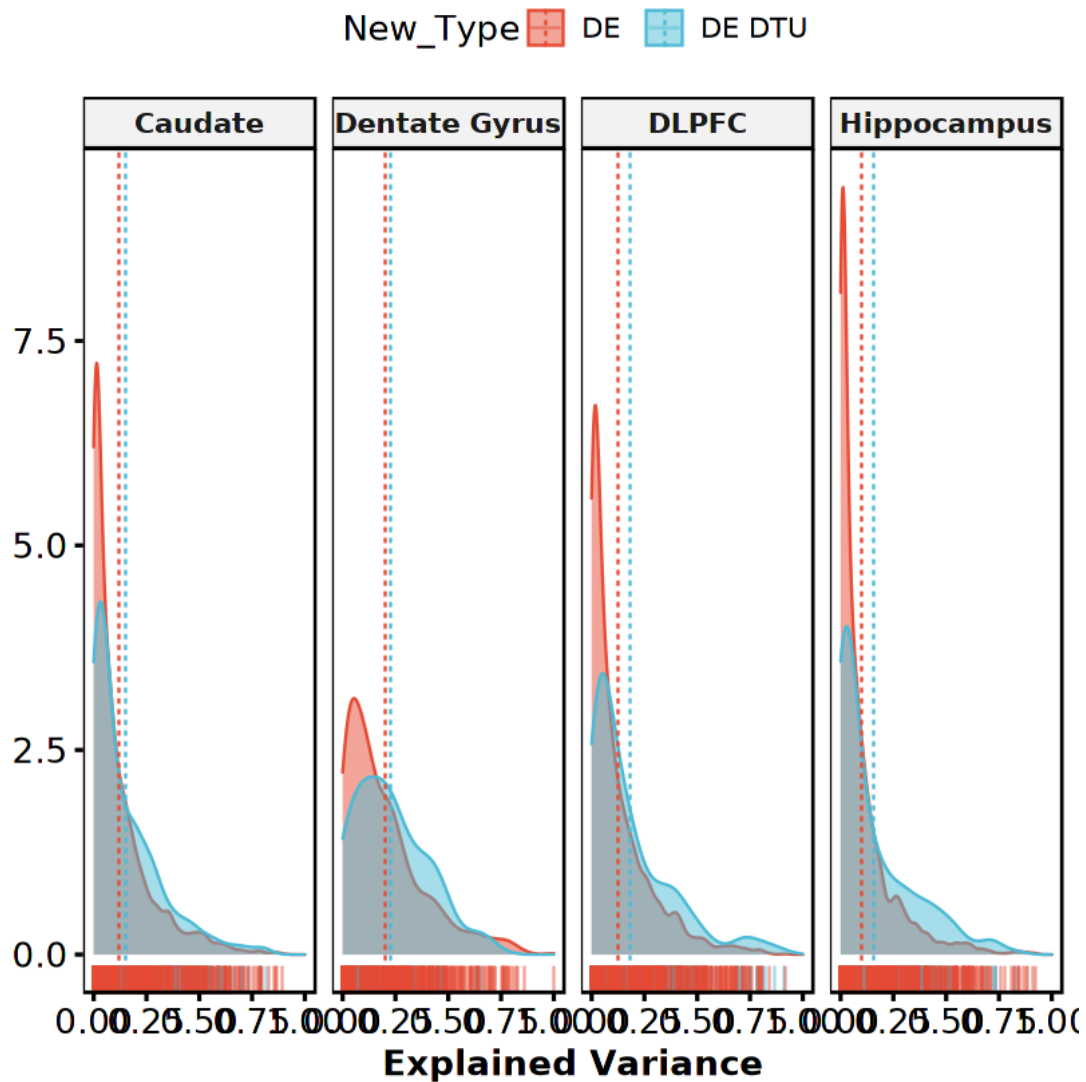
```
[48]: freq = df2 %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggdensity(x = "Partial_R2", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("Tissue"), xlab="Explained Variance",
    panel.labs.font=list(face='bold'), ylab="", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  font("xy.title", face="bold")
save_plot(freq, "explained_variance_distribution", 10, 4)
freq
```



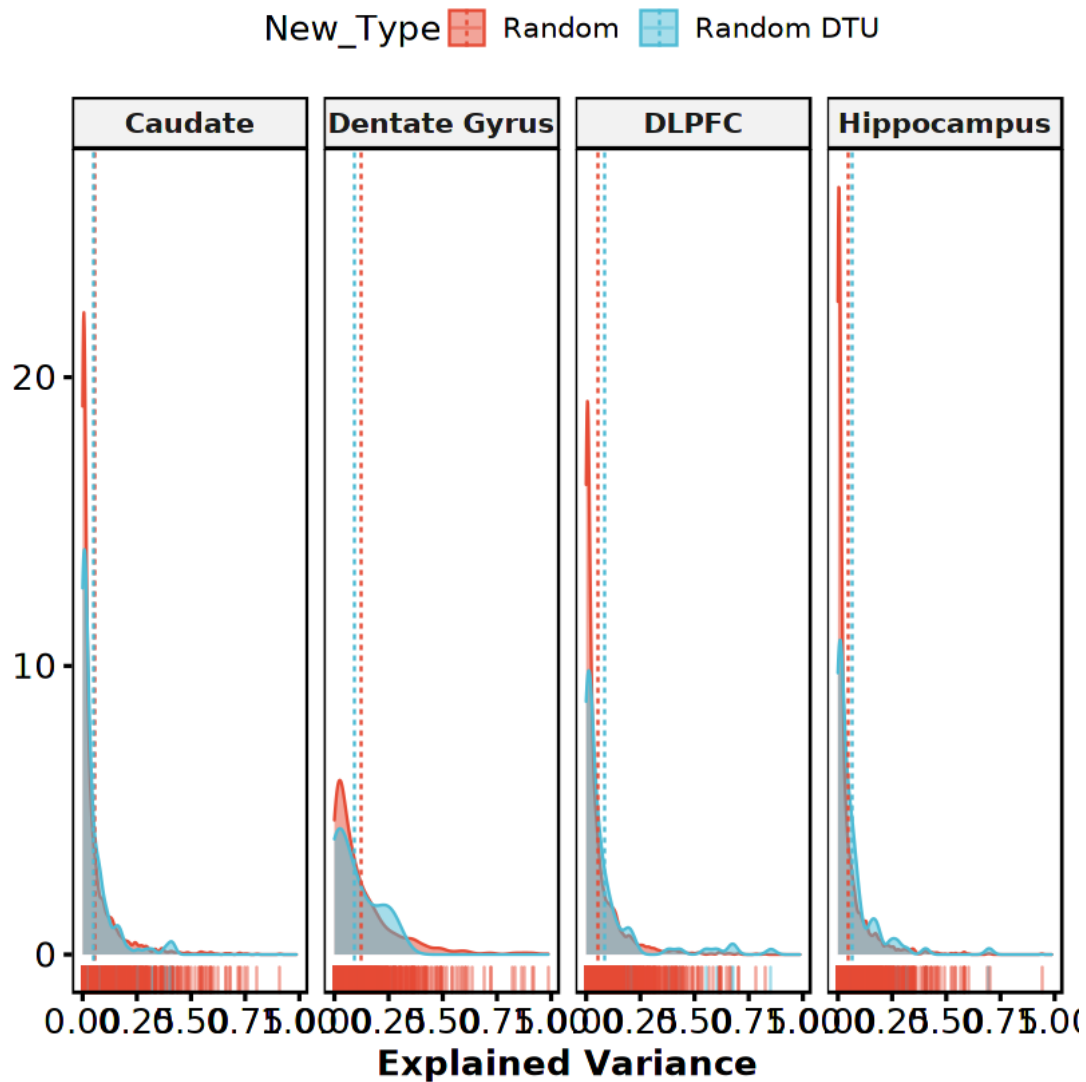
```
[49]: freq = df2 %>% filter(DTU == "DTU") %>%
  ggdensity(x = "Partial_R2", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("Tissue"), xlab="Explained Variance",
    panel.labs.font=list(face='bold'), ylab="", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  font("xy.title", face="bold")
save_plot(freq, "explained_variance_distribution_dtu", 10, 4)
freq
```



```
[50]: freq = df2 %>% filter(Type == "DE") %>%
  ggdensity(x = "Partial_R2", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("Tissue"), xlab="Explained Variance",
    panel.labs.font=list(face='bold'), ylab="", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  font("xy.title", face="bold")
save_plot(freq, "explained_variance_distribution_de", 10, 4)
freq
```



```
[51]: freq = df2 %>% filter(Type == "Random") %>%
  ggdensity(x = "Partial_R2", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("Tissue"), xlab="Explained Variance",
    panel.labs.font=list(face='bold'), ylab="", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  font("xy.title", face="bold")
save_plot(freq, "explained_variance_distribution_random", 10, 4)
freq
```



1.4 Reproducibility Information

```
[52]: Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-08-27 14:08:40 EDT"
```

```
user system elapsed
122.283 2.132 100.347
```

```
Session info
setting value
```



```

version R version 4.0.3 (2020-10-10)
os      Arch Linux
system  x86_64, linux-gnu
ui      X11
language (EN)
collate en_US.UTF-8
ctype   en_US.UTF-8
tz      America/New_York
date    2021-08-27

```

Packages

package	* version	date	lib	source
abind	1.4-5	2016-07-21	[1]	CRAN (R 4.0.2)
assertthat	0.2.1	2019-03-21	[1]	CRAN (R 4.0.2)
backports	1.2.1	2020-12-09	[1]	CRAN (R 4.0.2)
base64enc	0.1-3	2015-07-28	[1]	CRAN (R 4.0.2)
broom	0.7.8	2021-06-24	[1]	CRAN (R 4.0.3)
Cairo	1.5-12.2	2020-07-07	[1]	CRAN (R 4.0.2)
car	3.0-11	2021-06-27	[1]	CRAN (R 4.0.3)
carData	3.0-4	2020-05-22	[1]	CRAN (R 4.0.2)
cellranger	1.1.0	2016-07-27	[1]	CRAN (R 4.0.2)
cli	3.0.0	2021-06-30	[1]	CRAN (R 4.0.3)
colorspace	2.0-2	2021-06-24	[1]	CRAN (R 4.0.3)
crayon	1.4.1	2021-02-08	[1]	CRAN (R 4.0.3)
curl	4.3.2	2021-06-23	[1]	CRAN (R 4.0.3)
data.table	1.14.0	2021-02-21	[1]	CRAN (R 4.0.3)
DBI	1.1.1	2021-01-15	[1]	CRAN (R 4.0.2)
dbplyr	2.1.1	2021-04-06	[1]	CRAN (R 4.0.3)
digest	0.6.27	2020-10-24	[1]	CRAN (R 4.0.2)
dplyr	* 1.0.7	2021-06-18	[1]	CRAN (R 4.0.3)
ellipsis	0.3.2	2021-04-29	[1]	CRAN (R 4.0.3)
evaluate	0.14	2019-05-28	[1]	CRAN (R 4.0.2)
fansi	0.5.0	2021-05-25	[1]	CRAN (R 4.0.3)
farver	2.1.0	2021-02-28	[1]	CRAN (R 4.0.3)
forcats	* 0.5.1	2021-01-27	[1]	CRAN (R 4.0.2)
foreign	0.8-80	2020-05-24	[2]	CRAN (R 4.0.3)
fs	1.5.0	2020-07-31	[1]	CRAN (R 4.0.2)
generics	0.1.0	2020-10-31	[1]	CRAN (R 4.0.2)
ggplot2	* 3.3.5	2021-06-25	[1]	CRAN (R 4.0.3)
ggpubr	* 0.4.0	2020-06-27	[1]	CRAN (R 4.0.2)
ggsci	2.9	2018-05-14	[1]	CRAN (R 4.0.2)
ggsignif	0.6.2	2021-06-14	[1]	CRAN (R 4.0.3)
glue	1.4.2	2020-08-27	[1]	CRAN (R 4.0.2)
gtable	0.3.0	2019-03-25	[1]	CRAN (R 4.0.2)
haven	2.4.1	2021-04-23	[1]	CRAN (R 4.0.3)
hms	1.1.0	2021-05-17	[1]	CRAN (R 4.0.3)
htmltools	0.5.1.1	2021-01-22	[1]	CRAN (R 4.0.2)
httr	1.4.2	2020-07-20	[1]	CRAN (R 4.0.2)

IRdisplay	1.0	2021-01-20	[1]	CRAN	(R 4.0.2)
IRkernel	1.2	2021-05-11	[1]	CRAN	(R 4.0.3)
jsonlite	1.7.2	2020-12-09	[1]	CRAN	(R 4.0.2)
labeling	0.4.2	2020-10-20	[1]	CRAN	(R 4.0.2)
lifecycle	1.0.0	2021-02-15	[1]	CRAN	(R 4.0.3)
lubridate	1.7.10	2021-02-26	[1]	CRAN	(R 4.0.3)
magrittr	2.0.1	2020-11-17	[1]	CRAN	(R 4.0.2)
modelr	0.1.8	2020-05-19	[1]	CRAN	(R 4.0.2)
munsell	0.5.0	2018-06-12	[1]	CRAN	(R 4.0.2)
openxlsx	4.2.4	2021-06-16	[1]	CRAN	(R 4.0.3)
pbdZMQ	0.3-5	2021-02-10	[1]	CRAN	(R 4.0.3)
pillar	1.6.1	2021-05-16	[1]	CRAN	(R 4.0.3)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN	(R 4.0.2)
purrr	* 0.3.4	2020-04-17	[1]	CRAN	(R 4.0.2)
R6	2.5.0	2020-10-28	[1]	CRAN	(R 4.0.2)
Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.0.3)
readr	* 1.4.0	2020-10-05	[1]	CRAN	(R 4.0.2)
readxl	1.3.1	2019-03-13	[1]	CRAN	(R 4.0.2)
repr	1.1.3	2021-01-21	[1]	CRAN	(R 4.0.2)
reprex	2.0.0	2021-04-02	[1]	CRAN	(R 4.0.3)
rio	0.5.27	2021-06-21	[1]	CRAN	(R 4.0.3)
rlang	0.4.11	2021-04-30	[1]	CRAN	(R 4.0.3)
rstatix	0.7.0	2021-02-13	[1]	CRAN	(R 4.0.3)
rstudioapi	0.13	2020-11-12	[1]	CRAN	(R 4.0.2)
rvest	1.0.0	2021-03-09	[1]	CRAN	(R 4.0.3)
scales	1.1.1	2020-05-11	[1]	CRAN	(R 4.0.2)
sessioninfo	1.1.1	2018-11-05	[1]	CRAN	(R 4.0.2)
stringi	1.7.3	2021-07-16	[1]	CRAN	(R 4.0.3)
stringr	* 1.4.0	2019-02-10	[1]	CRAN	(R 4.0.2)
svglite	2.0.0	2021-02-20	[1]	CRAN	(R 4.0.3)
systemfonts	1.0.2	2021-05-11	[1]	CRAN	(R 4.0.3)
tibble	* 3.1.2	2021-05-16	[1]	CRAN	(R 4.0.3)
tidyr	* 1.1.3	2021-03-03	[1]	CRAN	(R 4.0.3)
tidyselect	1.1.1	2021-04-30	[1]	CRAN	(R 4.0.3)
tidyverse	* 1.3.1	2021-04-15	[1]	CRAN	(R 4.0.3)
utf8	1.2.1	2021-03-12	[1]	CRAN	(R 4.0.3)
uuid	0.1-4	2020-02-26	[1]	CRAN	(R 4.0.2)
vctrs	0.3.8	2021-04-29	[1]	CRAN	(R 4.0.3)
withr	2.4.2	2021-04-18	[1]	CRAN	(R 4.0.3)
xml2	1.3.2	2020-04-23	[1]	CRAN	(R 4.0.2)
zip	2.2.0	2021-05-31	[1]	CRAN	(R 4.0.3)

[1] /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0

[2] /usr/lib/R/library