

main

September 23, 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  3.1.4    dplyr   1.0.7
tidyr   1.1.3    stringr 1.4.0
readr   2.0.1    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	-0.04040379	-0.01089123	0.1993
	2	Caudate	ENSG000000003509_15	test_score_r2	-0.09541787	-0.03918813	0.1591
	3	Caudate	ENSG000000004468_12	test_score_r2	-0.24862373	-0.26537954	0.2069
	4	Caudate	ENSG000000004777_18	test_score_r2	-0.19261307	-0.03691556	0.3620
	5	Caudate	ENSG000000005243_9	test_score_r2	-0.09919432	-0.06594147	0.0822
	6	Caudate	ENSG000000005436_13	test_score_r2	0.17590042	0.17066234	0.1160

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. 763 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 <fct>	feature <fct>	metric <fct>	Mean <dbl>	Median <dbl>	Std <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. 37448 2. 13

A data.frame: 2 × 13		tissue <fct>	feature <fct>	metric <fct>	Mean <dbl>	Median <dbl>	Std <dbl>
1	Caudate	ENSG00000003249_13	test_score_r2	-0.04040379	-0.01089123	0.1993	
2	Caudate	ENSG00000003509_15	test_score_r2	-0.09541787	-0.03918813	0.1591	

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")
```

	tissue <fct>	Type <fct>	model <fct>	Mean <dbl>	Median <dbl>
A grouped_df: 16 × 5	Caudate	DE	Elastic Net	0.047364920	-0.008385236
	Caudate	DE	Random Forest	-0.004149059	-0.048791348
	Caudate	Random	Elastic Net	-0.031713741	-0.038995604
	Caudate	Random	Random Forest	-0.096066770	-0.096829138
	Dentate Gyrus	DE	Elastic Net	-0.045845106	-0.074908845
	Dentate Gyrus	DE	Random Forest	-0.111294729	-0.137511288
	Dentate Gyrus	Random	Elastic Net	-0.141836455	-0.124429624
	Dentate Gyrus	Random	Random Forest	-0.321247001	-0.315668145
	DLPFC	DE	Elastic Net	0.041050891	-0.016760782
	DLPFC	DE	Random Forest	-0.018942404	-0.064636371
	DLPFC	Random	Elastic Net	-0.049235705	-0.051566937
	DLPFC	Random	Random Forest	-0.130636227	-0.129728827
	Hippocampus	DE	Elastic Net	0.030346293	-0.018556250
	Hippocampus	DE	Random Forest	-0.023817650	-0.059371891
	Hippocampus	Random	Elastic Net	-0.038067283	-0.039989113
	Hippocampus	Random	Random Forest	-0.105771529	-0.103582057

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

	tissue <fct>	Type <fct>	model <fct>	Mean <dbl>	Median <dbl>
A grouped_df: 8 × 5	Caudate	DE	Elastic Net	0.04736492	-0.008385236
	Caudate	Random	Elastic Net	-0.03171374	-0.038995604
	Dentate Gyrus	DE	Elastic Net	-0.04584511	-0.074908845
	Dentate Gyrus	Random	Elastic Net	-0.14183646	-0.124429624
	DLPFC	DE	Elastic Net	0.04105089	-0.016760782
	DLPFC	Random	Elastic Net	-0.04923571	-0.051566937
	Hippocampus	DE	Elastic Net	0.03034629	-0.018556250
	Hippocampus	Random	Elastic Net	-0.03806728	-0.039989113

```
[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	-0.004149059	-0.04879135
	Caudate	Random	Random Forest	-0.096066770	-0.09682914
	Dentate Gyrus	DE	Random Forest	-0.111294729	-0.13751129
	Dentate Gyrus	Random	Random Forest	-0.321247001	-0.31566815
	DLPFC	DE	Random Forest	-0.018942404	-0.06463637
	DLPFC	Random	Random Forest	-0.130636227	-0.12972883
	Hippocampus	DE	Random Forest	-0.023817650	-0.05937189
	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	0.083358579	0.021824460
	Caudate	DE DTU	Random Forest	0.031564028	-0.024358561
	Caudate	Random DTU	Elastic Net	-0.042723875	-0.043454092
	Caudate	Random DTU	Random Forest	-0.110650265	-0.098685822
	Dentate Gyrus	DE DTU	Elastic Net	-0.001453411	-0.057626652
	Dentate Gyrus	DE DTU	Random Forest	-0.054620608	-0.073247823
	Dentate Gyrus	Random DTU	Elastic Net	-0.124886491	-0.098911284
	Dentate Gyrus	Random DTU	Random Forest	-0.263556534	-0.221790579
	DLPFC	DE DTU	Elastic Net	0.110945347	0.030203374
	DLPFC	DE DTU	Random Forest	0.060444210	-0.009606728
	DLPFC	Random DTU	Elastic Net	-0.032211432	-0.047589984
	DLPFC	Random DTU	Random Forest	-0.104759586	-0.124457006
	Hippocampus	DE DTU	Elastic Net	0.082646771	0.008704045
	Hippocampus	DE DTU	Random Forest	0.041500481	-0.024132825
	Hippocampus	Random DTU	Elastic Net	-0.027494227	-0.033656459
	Hippocampus	Random DTU	Random Forest	-0.086672716	-0.096477614

```
[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	0.083358579	0.021824460
	Caudate	Random DTU	Elastic Net	-0.042723875	-0.043454092
	Dentate Gyrus	DE DTU	Elastic Net	-0.001453411	-0.057626652
	Dentate Gyrus	Random DTU	Elastic Net	-0.124886491	-0.098911284
	DLPFC	DE DTU	Elastic Net	0.110945347	0.030203374
	DLPFC	Random DTU	Elastic Net	-0.032211432	-0.047589984
	Hippocampus	DE DTU	Elastic Net	0.082646771	0.008704045
	Hippocampus	Random DTU	Elastic Net	-0.027494227	-0.033656459

```
[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	0.22643875	0.1834459
	Dentate Gyrus	0.09967526	0.1104247
	DLPFC	0.26033747	0.2069071
	Hippocampus	0.24436662	0.1934286

```
[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	0.22586916	0.14214099
	Dentate Gyrus	0.07252514	0.08457646
	DLPFC	0.22460423	0.16552777
	Hippocampus	0.19047202	0.14687688

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	-0.08549819	0.002645479	-32.31861	1.326054e-219	5.3042
	Dentate Gyrus	TypeRandom	-0.15297181	0.007802110	-19.60647	8.662432e-81	3.4649
	DLPFC	TypeRandom	-0.10110127	0.002959698	-34.15932	4.862027e-243	1.9448
	Hippocampus	TypeRandom	-0.07518373	0.002420087	-31.06654	1.121866e-203	4.4874

```
[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	0.04113378	0.006818216	6.032923	1.708711e-09	6.8
	Dentate Gyrus	New_TypeDE DTU	0.05365651	0.026285552	2.041293	4.139144e-02	1.6
	DLPFC	New_TypeDE DTU	0.08198720	0.008949757	9.160829	7.166234e-20	2.8
	Hippocampus	New_TypeDE DTU	0.06359357	0.007985402	7.963728	1.992383e-15	7.9

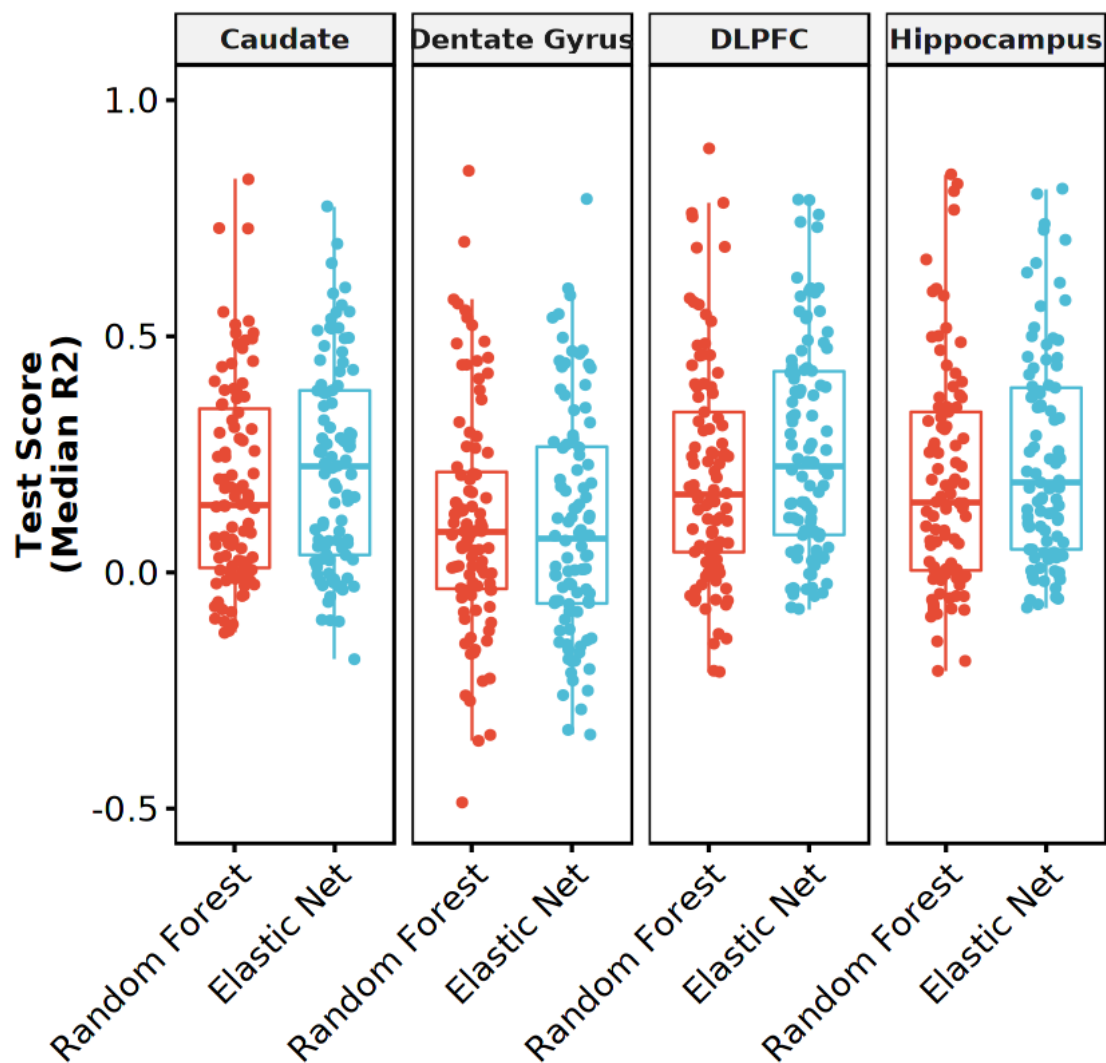
```
[19]: df %>% filter(Type == "Random") %>% group_by(tissue) %>%
  do(fit = broom::tidy(lm(Median ~ New_Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	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

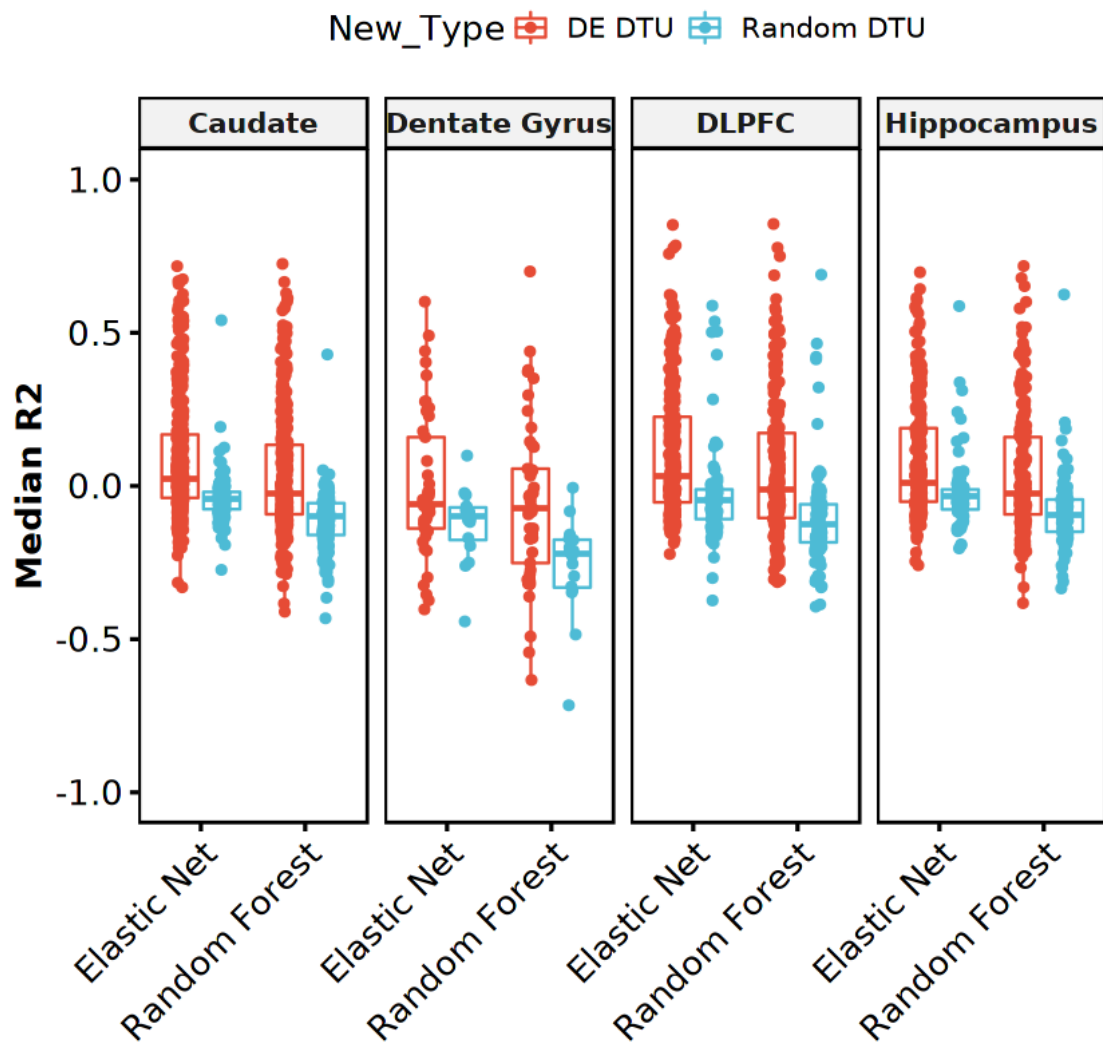
1.2.3 Plot

Boxplots

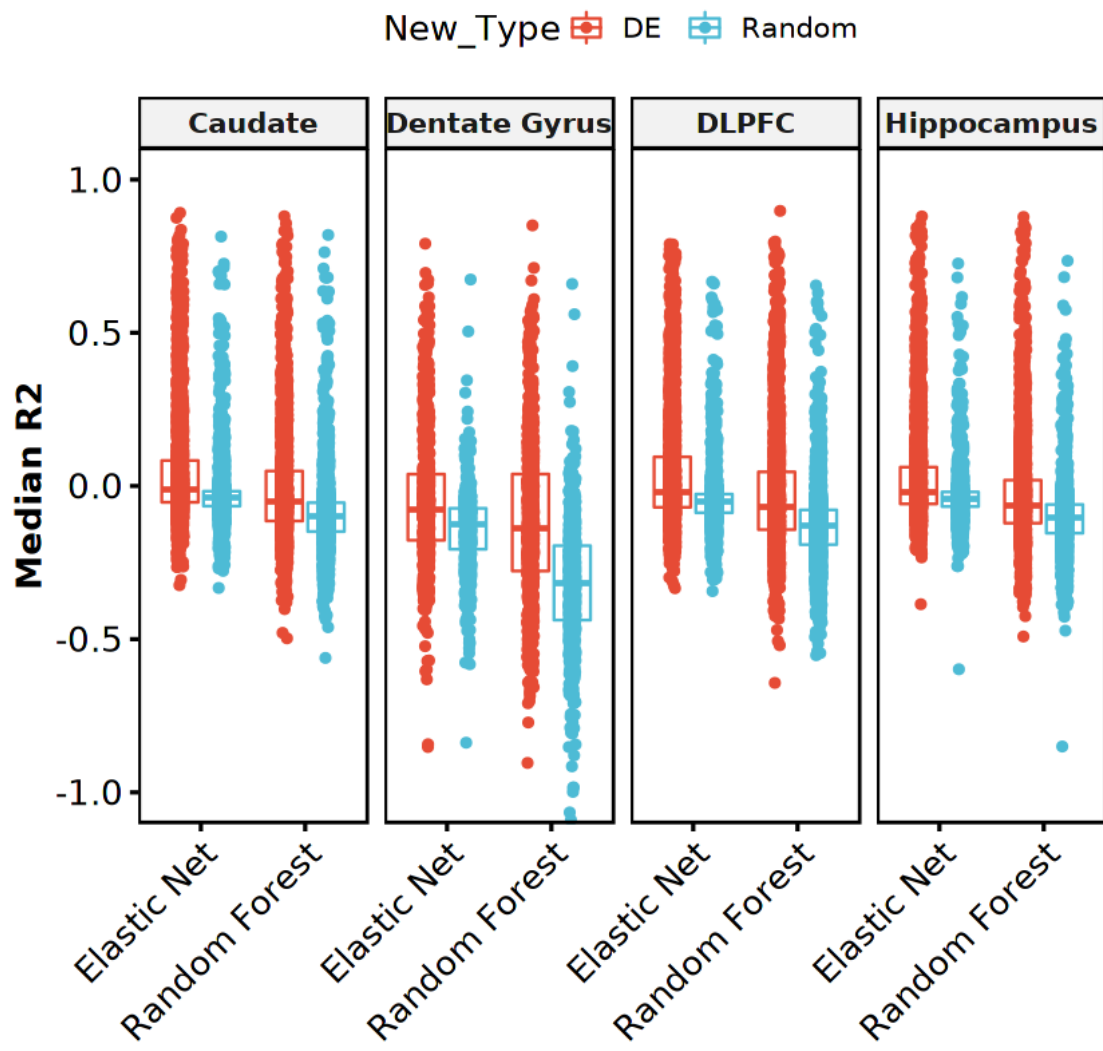
```
[20]: bxp = de100 %>%
  ggboxplot(x="model", y="Median", color="model", add="jitter",
    facet.by="tissue", legend="none", palette="npg",
    ylim=c(-0.5, 1), ylab="Test Score\n(Median R2)",
    xlab="", ncol=4, panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE))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



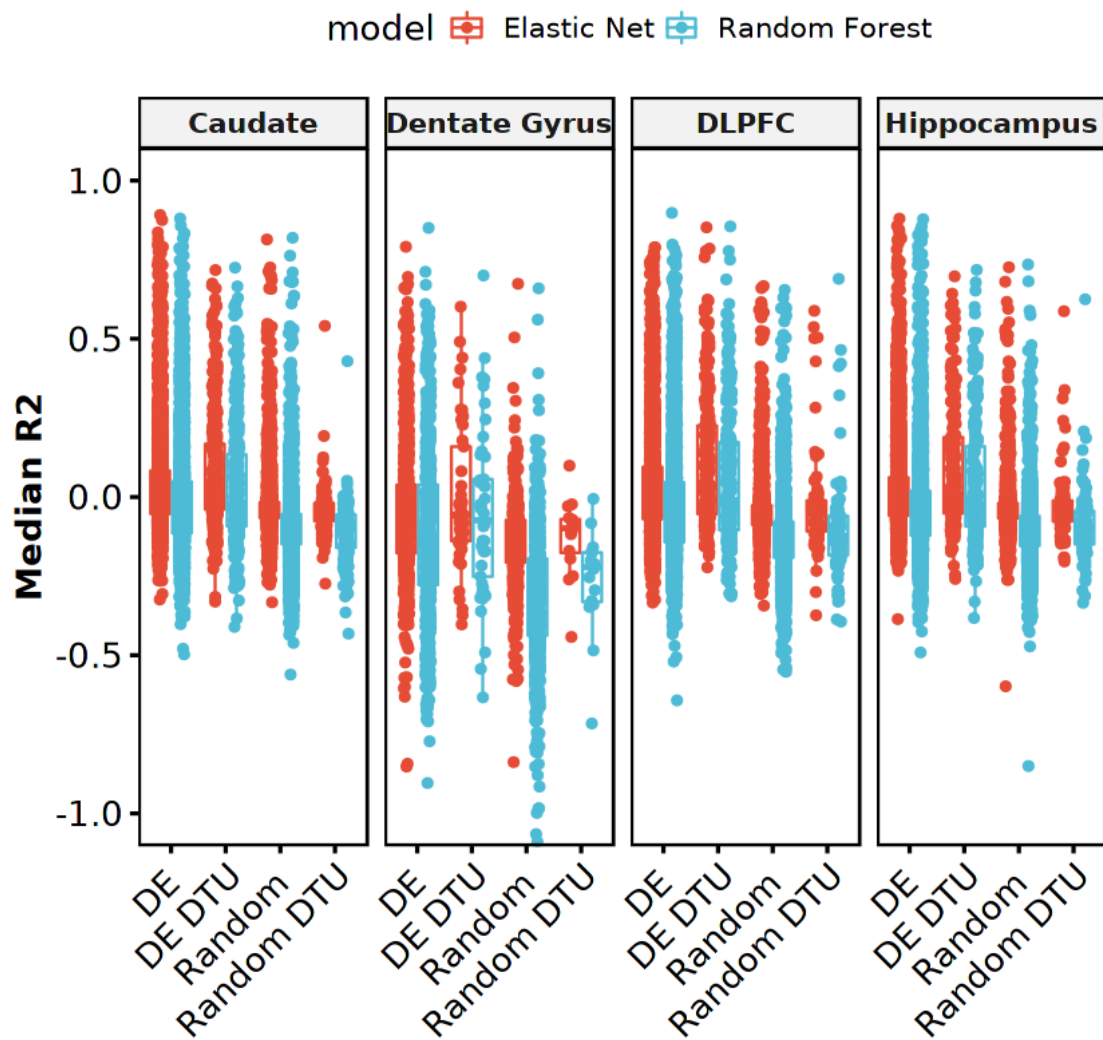
```
[21]: bxp = df %>% filter(DTU == "DTU") %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
 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

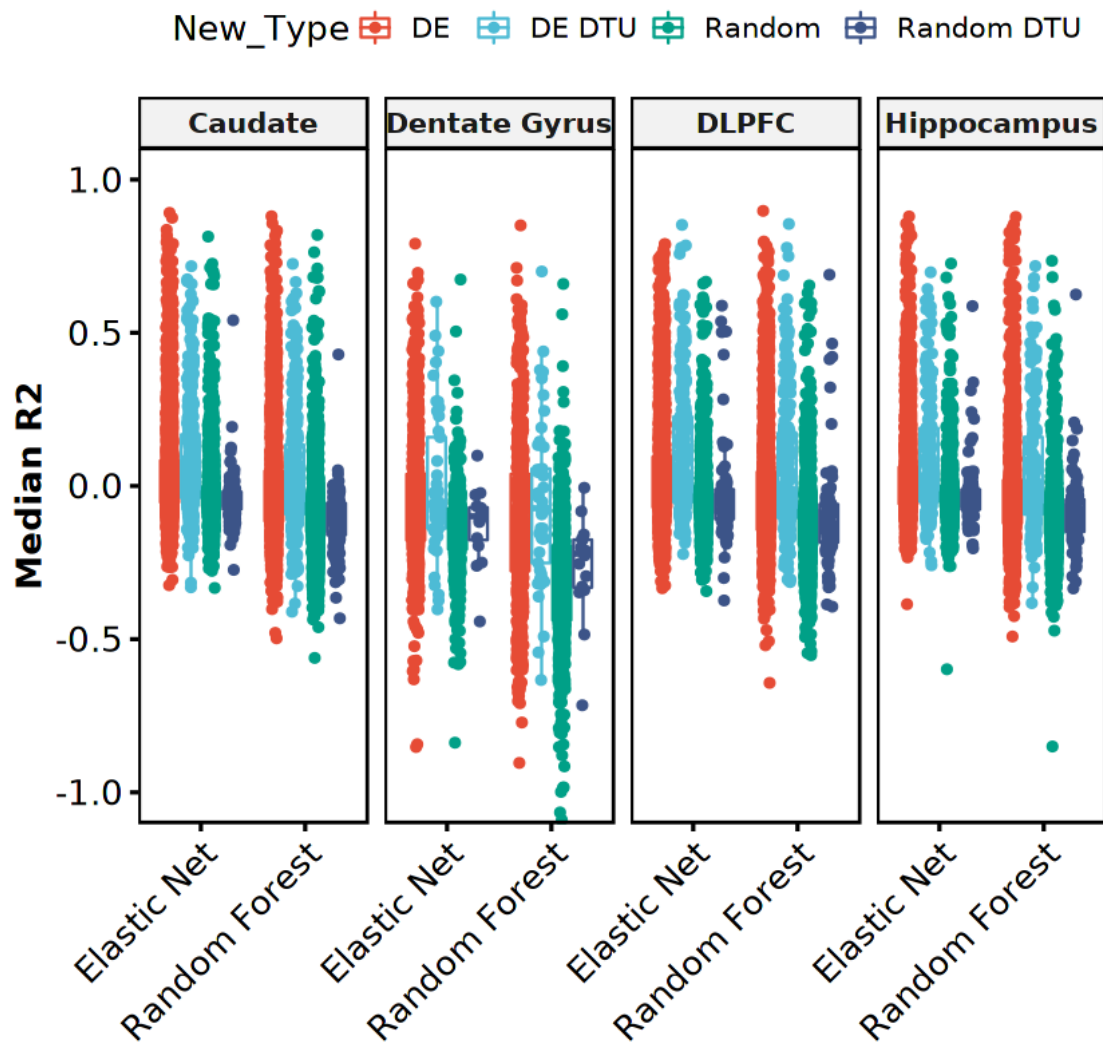
```
[22]: bxp = df %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15,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



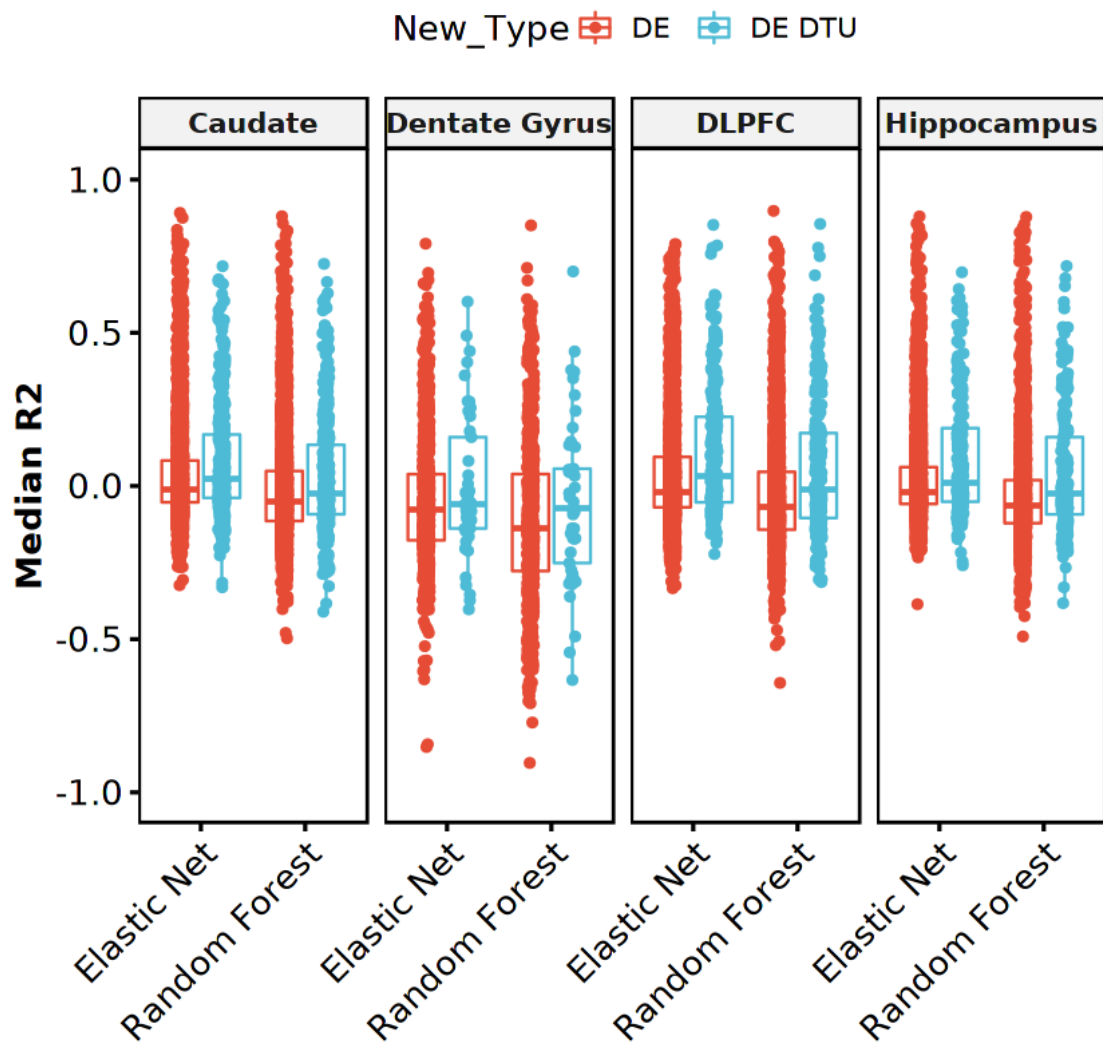
```
[23]: bxp = df %>% ggboxplot(x="New_Type", y="Median", color="model", add="jitter",
  facet.by="tissue", palette="npg", ylim=c(-1, 1),
  ylab="Median R2", xlab="", ncol=4,
  panel.labs.font=list(face='bold'),
  ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title",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



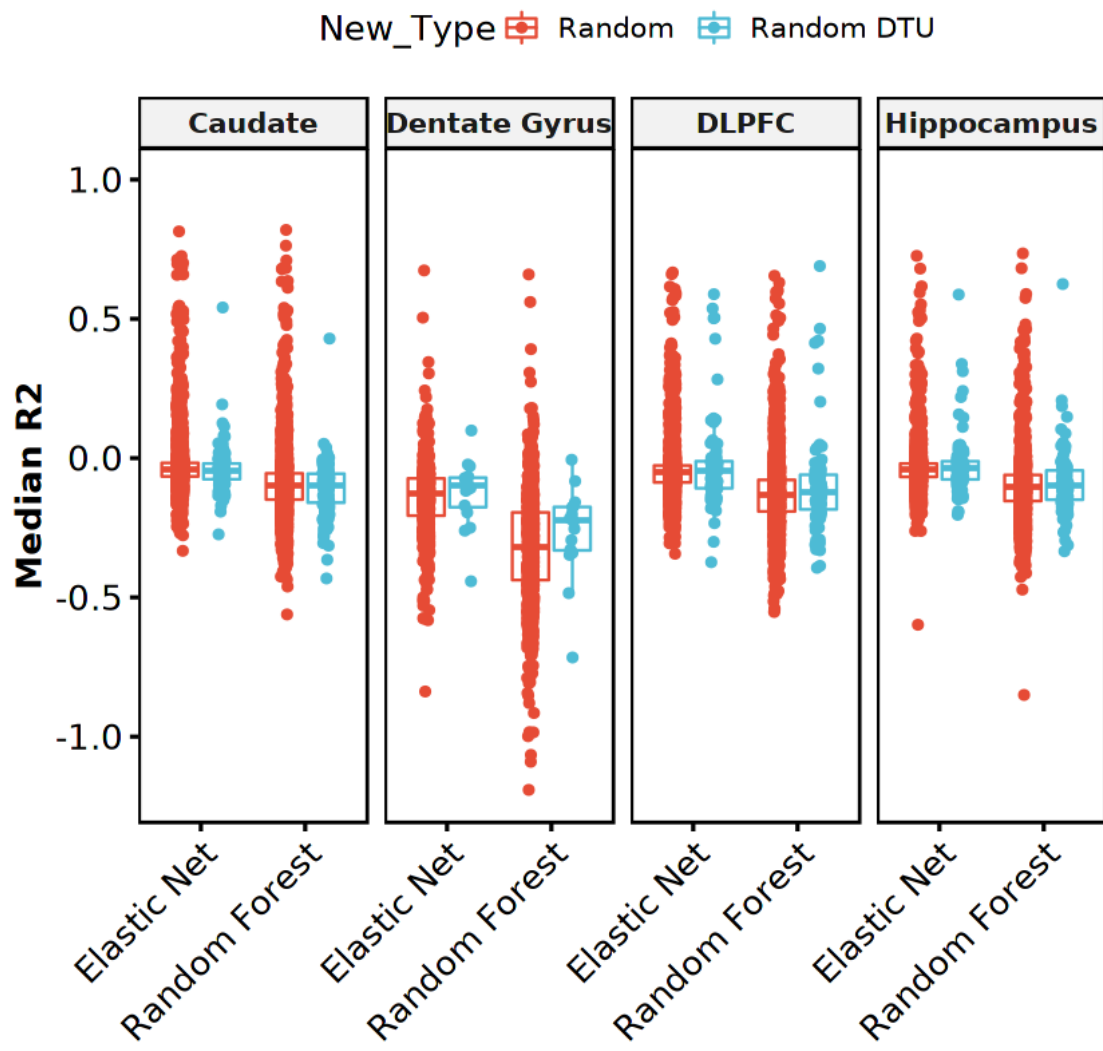
```
[24]: bxp = df %>% ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
  facet.by="tissue", palette="npg", ylim=c(-1, 1),
  ylab="Median R2", xlab="", ncol=4,
  panel.labs.font=list(face='bold'),
  ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title",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

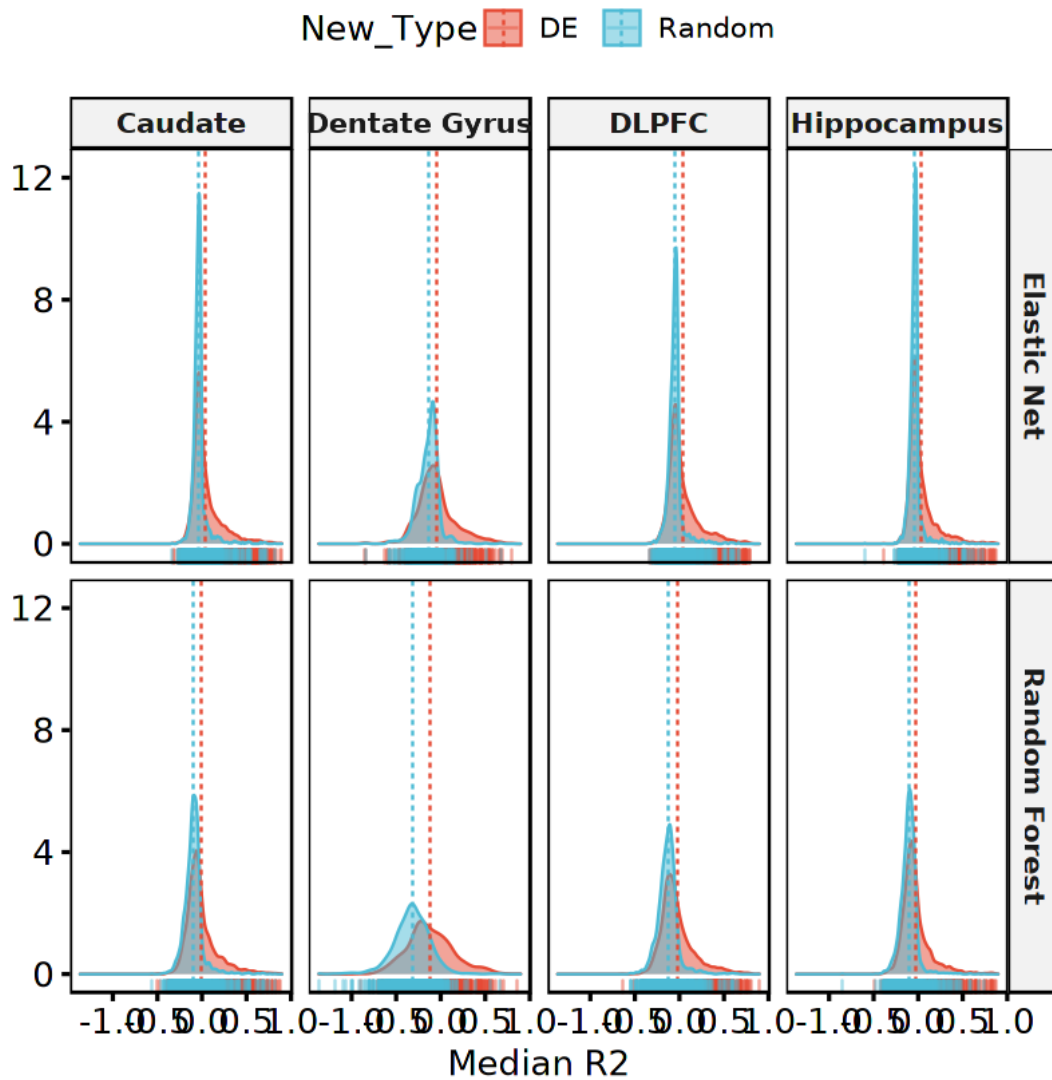


```
[26]: bxp = df %>% filter(Type == "Random") %>%
  ggboxplot(x="model", y="Median", color="New_Type", add="jitter",
    facet.by="tissue", palette="npg", ylim=c(-1.2, 1),
    ylab="Median R2", xlab="", ncol=4,
    panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE))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

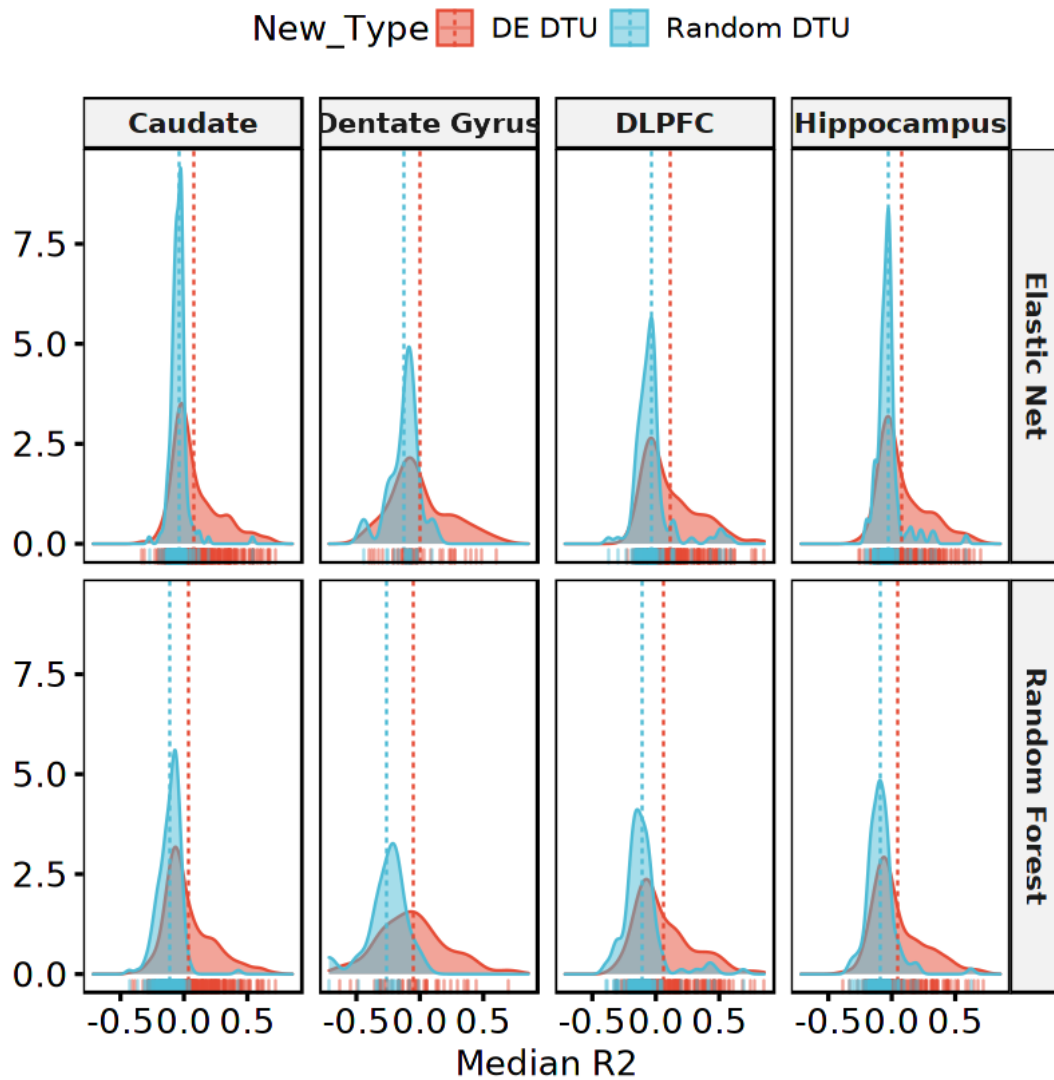


Distribution

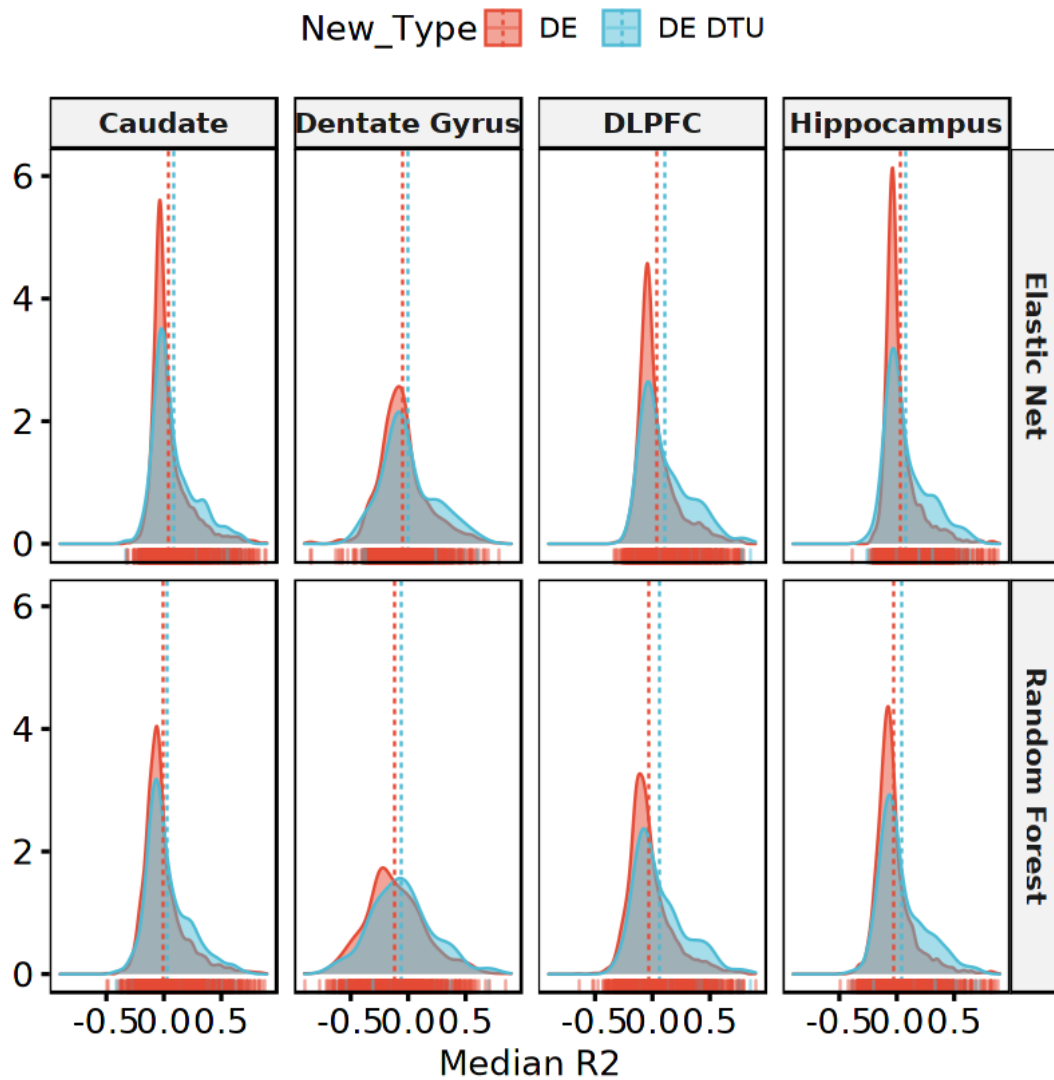
```
[27]: freq = df %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggdensity(x = "Median", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("model", "tissue"), xlab="Median R2",
    panel.labs.font=list(face='bold'), ylab="",
    ggtheme=theme_pubr(base_size=15,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



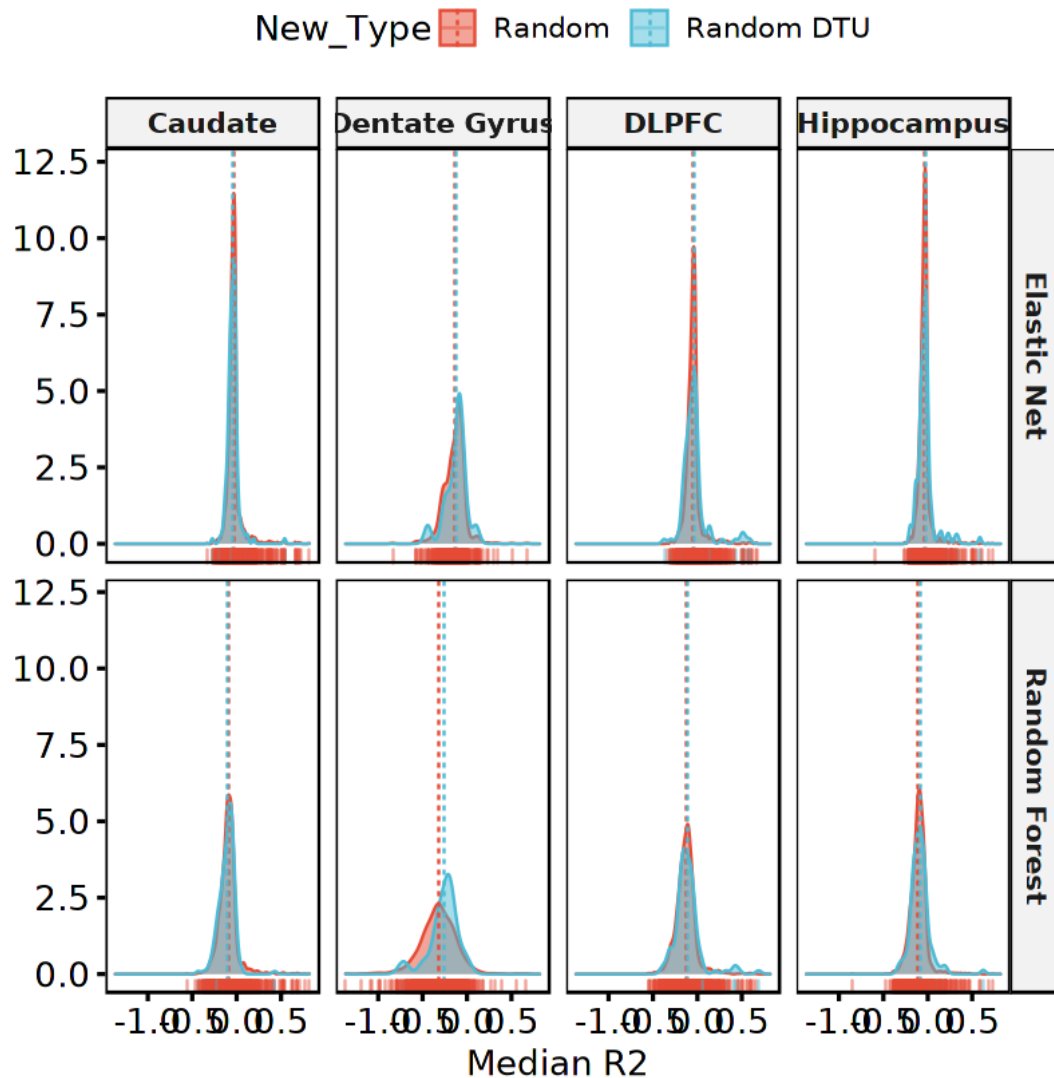
```
[28]: freq = df %>% filter(DTU == "DTU") %>%
  ggdensity(x = "Median", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("model", "tissue"), xlab="Median R2",
    panel.labs.font=list(face='bold'), ylab="",
    ggtheme=theme_pubr(base_size=15, border=TRUE))
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/enet_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	19	0.17320347	187.4661	226.7379	Caudate
	2	ENSG00000003509.15	11	0.05478319	215.9147	228.4288	Caudate

```
[32]: de100_v2 = data.table::fread("../de_genes/partial_r2/enet_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	11	0.09353468	168.3528	185.7245	Caudate	DI
		ENSG000000034053.14	10	0.02726263	191.4707	196.8370	Caudate	DI

```
[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	19	0.17320347	187.4661	226.7379	Caudate
	2	ENSG000000003509.15	11	0.05478319	215.9147	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	0.12661422	0.06633090	0.15369961
	Caudate	Random	0.05462609	0.01699683	0.09628326
	Dentate Gyrus	DE	0.16093977	0.10535294	0.15987563
	Dentate Gyrus	Random	0.12287941	0.07247299	0.14869899
	DLPFC	DE	0.12729445	0.06716576	0.15062524
	DLPFC	Random	0.05820650	0.02022916	0.09648807
	Hippocampus	DE	0.10697598	0.05497714	0.13713545
	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	0.2125584	0.1529605	0.1907122
	Dentate Gyrus	0.2341949	0.1873381	0.1761255
	DLPFC	0.2232828	0.1588400	0.2047702
	Hippocampus	0.2253557	0.1546401	0.1959000

```
[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	0.12313659	0.06344694	0.15189444
	Caudate	DE DTU	0.15022701	0.09273527	0.16370508
	Caudate	Random	0.05484748	0.01691896	0.09715149
	Caudate	Random DTU	0.05096409	0.02011857	0.08074849
	Dentate Gyrus	DE	0.15886881	0.10251711	0.15851589
	Dentate Gyrus	DE DTU	0.19444336	0.14792580	0.17909600
	Dentate Gyrus	Random	0.12350021	0.07247299	0.14955208
	Dentate Gyrus	Random DTU	0.09354702	0.05169564	0.09920611
	DLPFC	DE	0.12302194	0.06478967	0.14658899
	DLPFC	DE DTU	0.17088470	0.11033965	0.18158222
	DLPFC	Random	0.05714081	0.02007225	0.09325817
	DLPFC	Random DTU	0.08482760	0.02604147	0.15528999
	Hippocampus	DE	0.10344327	0.05367530	0.13295824
	Hippocampus	DE DTU	0.15040085	0.07286198	0.17541652
	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] 5.782183e-97
[1] 6.967195e-07
[1] 7.871218e-86
[1] 5.509819e-90
```

```
[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	-0.07198814	0.003411066	-21.104294	2.887141e-95	1.1548
	Dentate Gyrus	TypeRandom	-0.03806036	0.007855809	-4.844868	1.394341e-06	5.5773
	DLPFC	TypeRandom	-0.06908795	0.003475701	-19.877417	7.081244e-85	2.8324
	Hippocampus	TypeRandom	-0.06109702	0.003020795	-20.225474	7.048574e-88	2.8194

```
[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.l <dbl>
A tibble: 4 × 7	Caudate	New_TypeDE DTU	0.02709042	0.008476780	3.195838	1.409088e-03	5.6
	Dentate Gyrus	New_TypeDE DTU	0.03557455	0.024540901	1.449602	1.475760e-01	5.9
	DLPFC	New_TypeDE DTU	0.04786276	0.010106107	4.736024	2.291791e-06	9.1
	Hippocampus	New_TypeDE DTU	0.04695758	0.009598573	4.892142	1.051727e-06	4.2

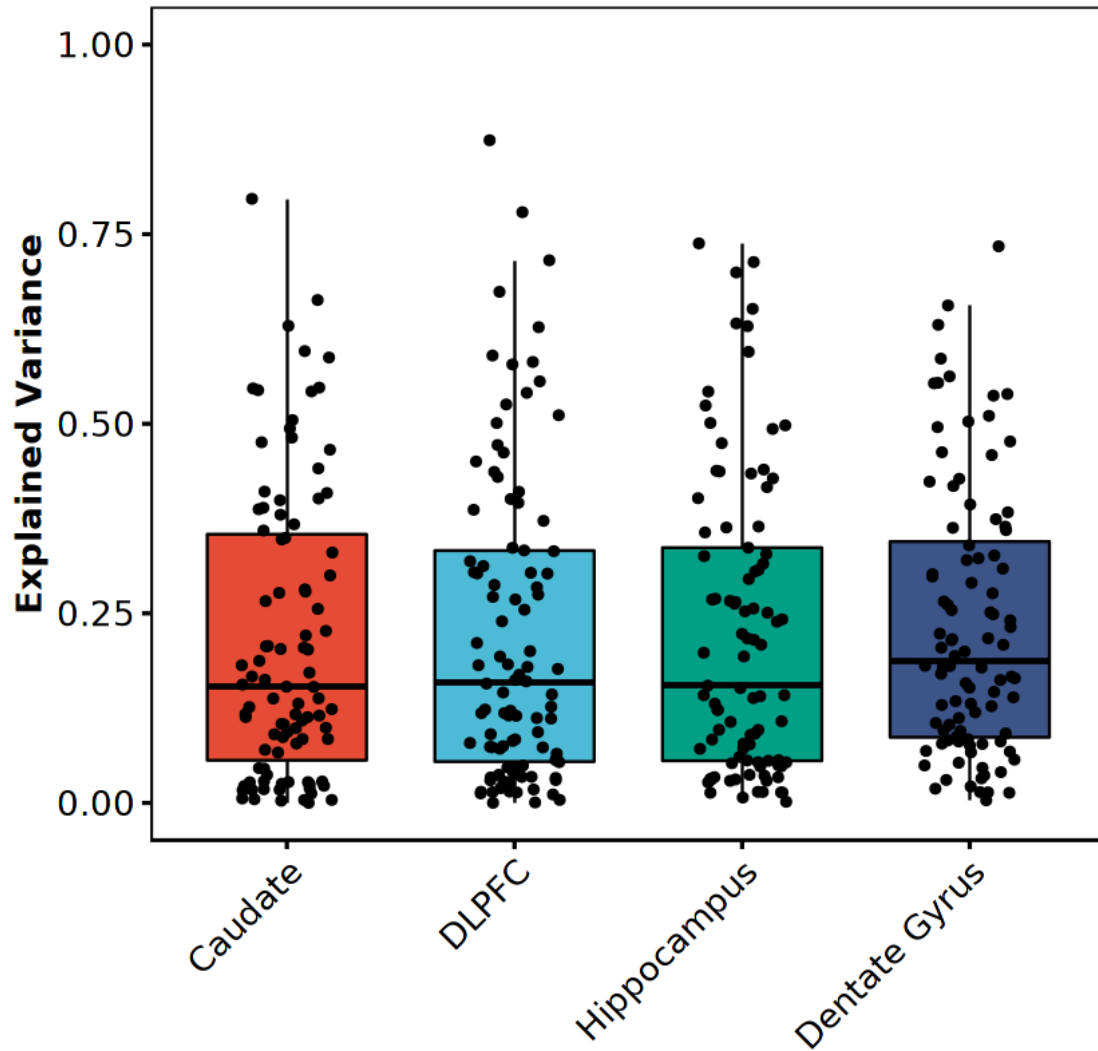
```
[41]: df2 %>% filter(Type == "Random") %>% group_by(Tissue) %>%
  do(fit = broom::tidy(lm(Partial_R2 ~ New_Type, data=))) %>%
  unnest(fit) %>% filter(term != '(Intercept)') %>%
  mutate(p.bonf = p.adjust(p.value, "bonf"))
```

	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

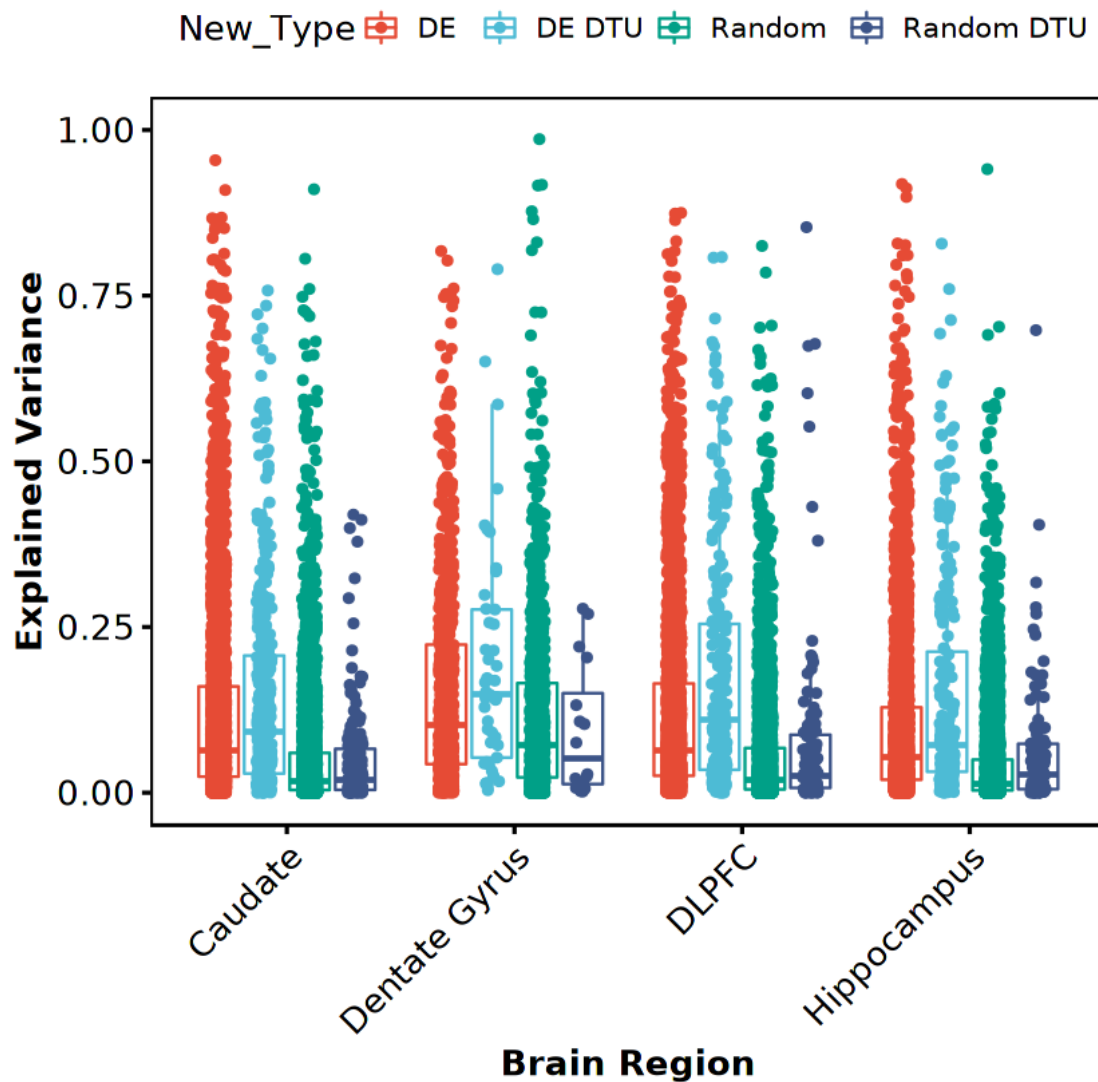
1.3.4 Plot

Boxplots

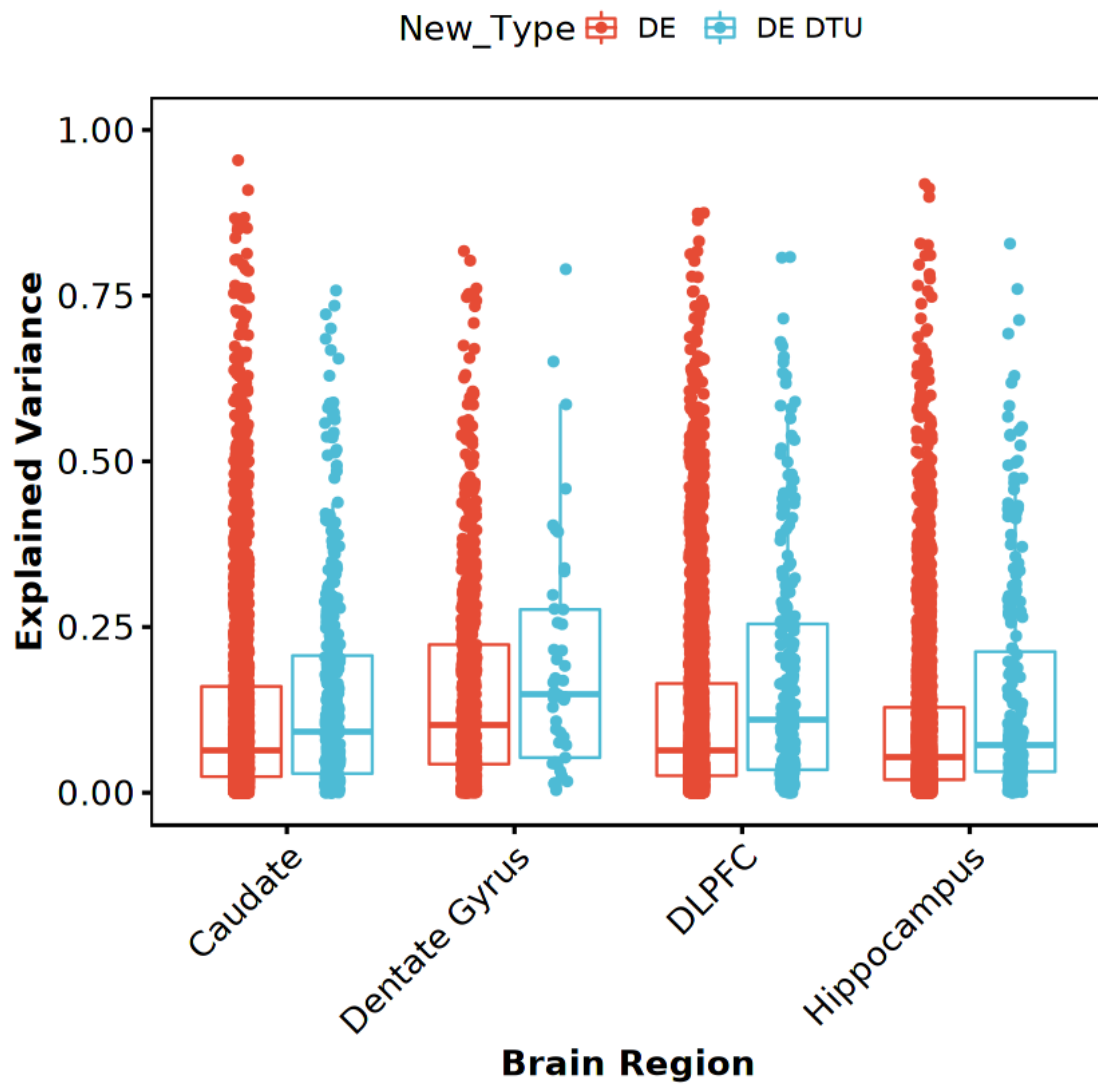
```
[42]: bxp = de100_v2 %>%
  ggboxplot(x="Tissue", y="Partial_R2", fill="Tissue", add="jitter",
    palette="npg", legend="none", ylim=c(0, 1), xlab="", ncol=4,
    ylab="Explained Variance", panel.labs.font=list(face='bold'),
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45)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



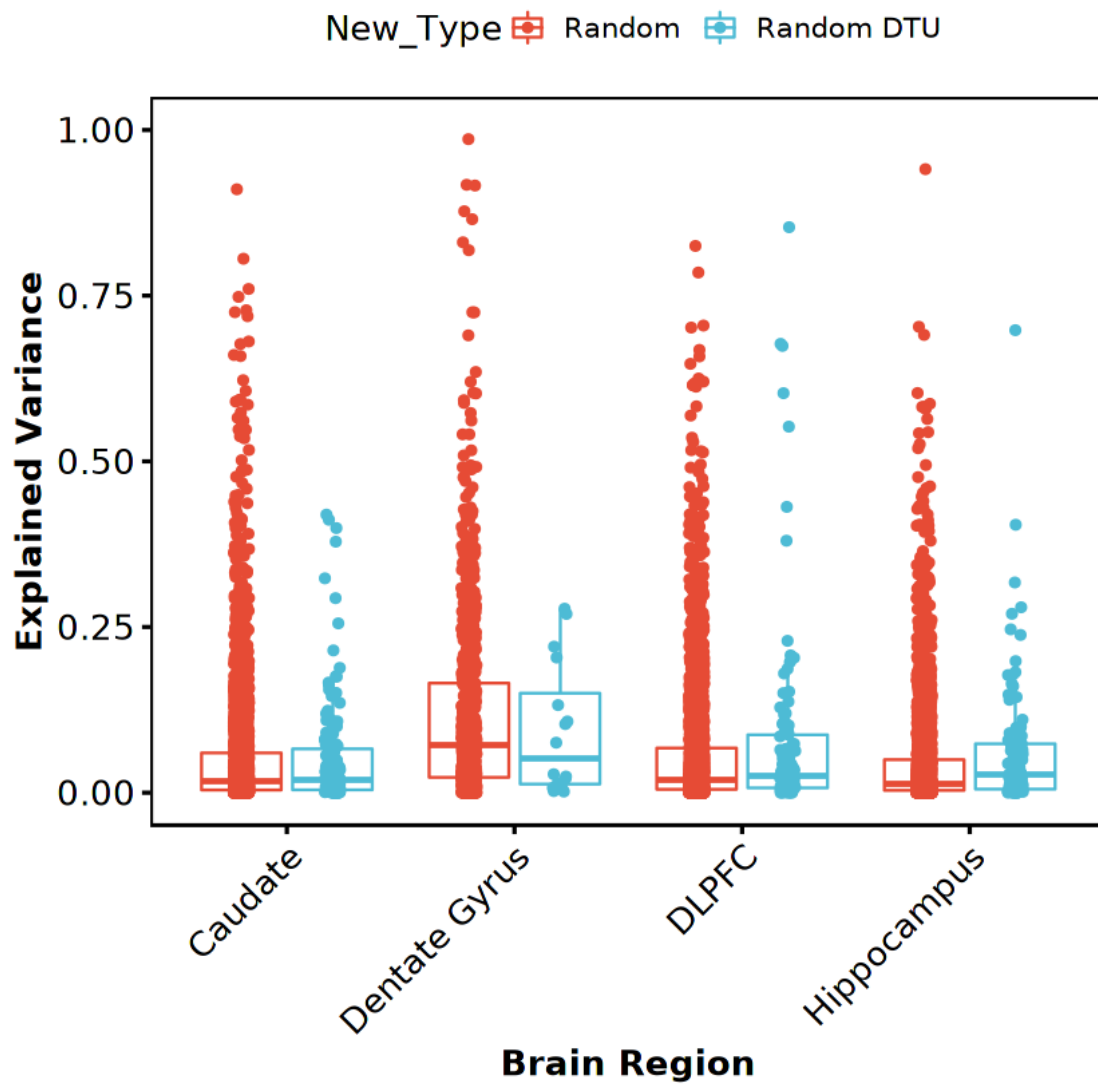
```
[43]: bxp = df2 %>% #filter(Type == "DE") %>%
      ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
                palette="npg", ylim=c(0, 1), ylab="Explained Variance",
                xlab="Brain Region", ncol=4,
                ggtheme=theme_pubr(base_size=15, border=TRUE))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



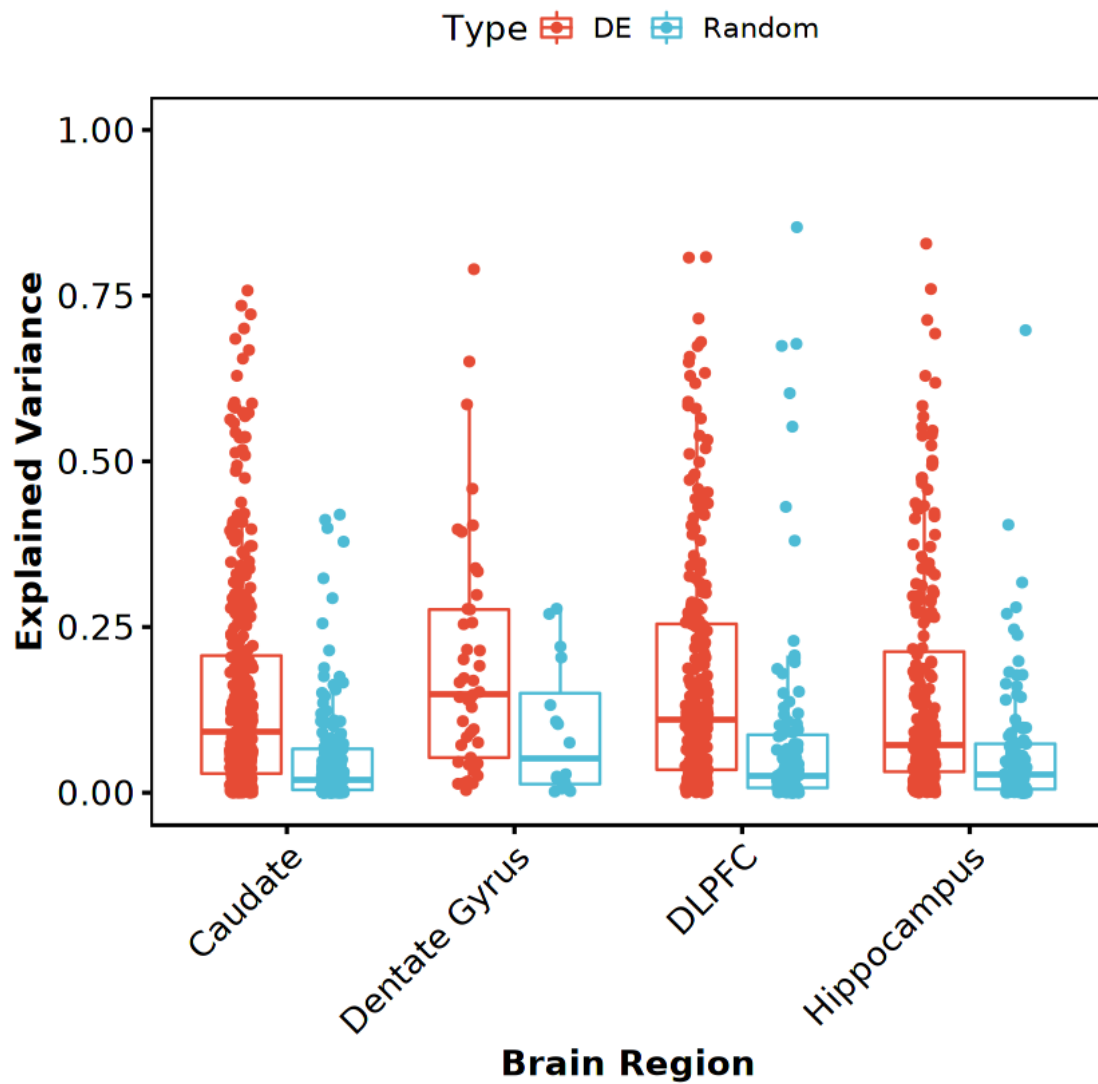
```
[44]: bxp = df2 %>% filter(Type == "DE") %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title",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

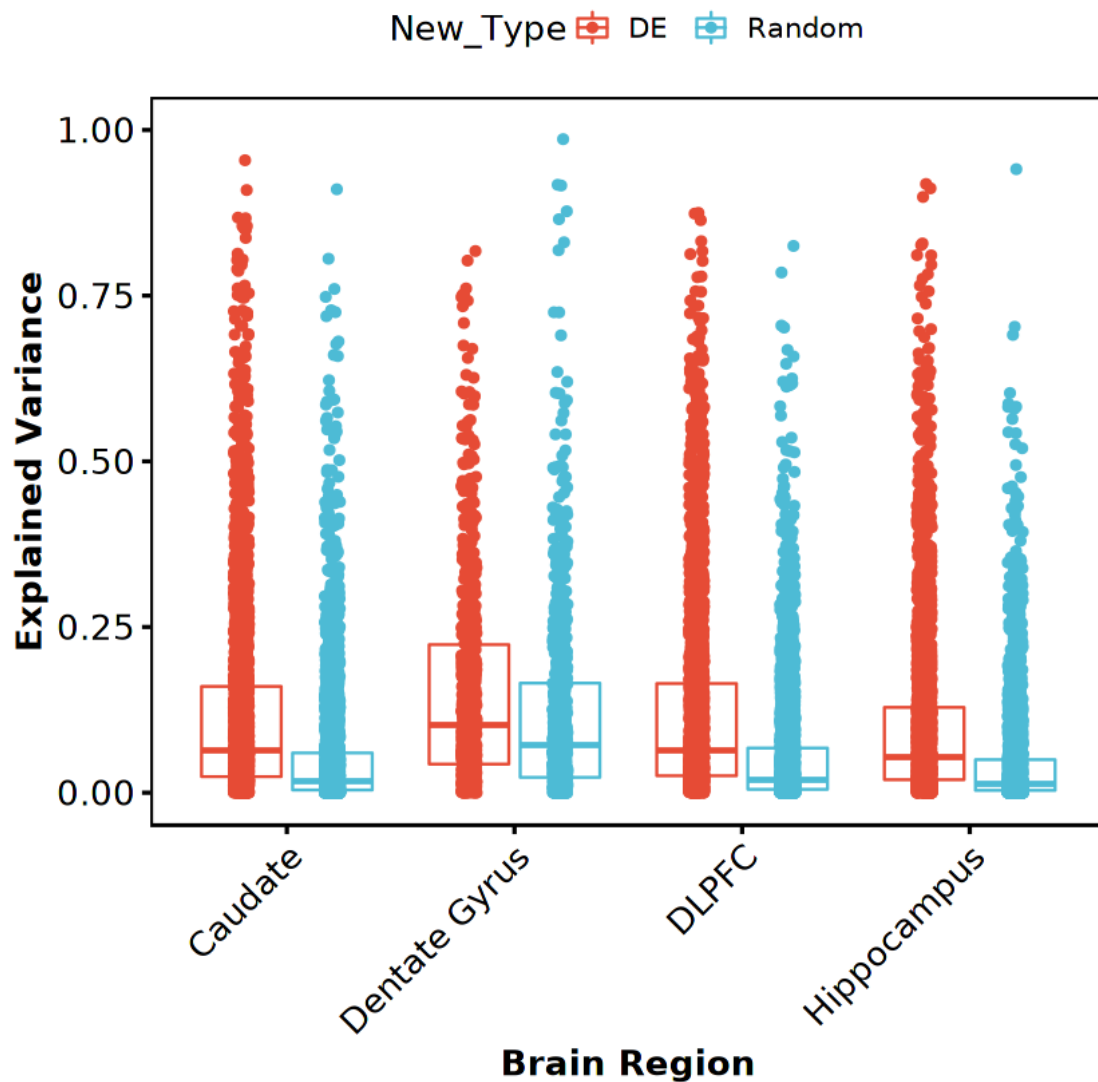
```
[45]: bxp = df2 %>% filter(Type == "Random") %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45) + font("xy.title",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

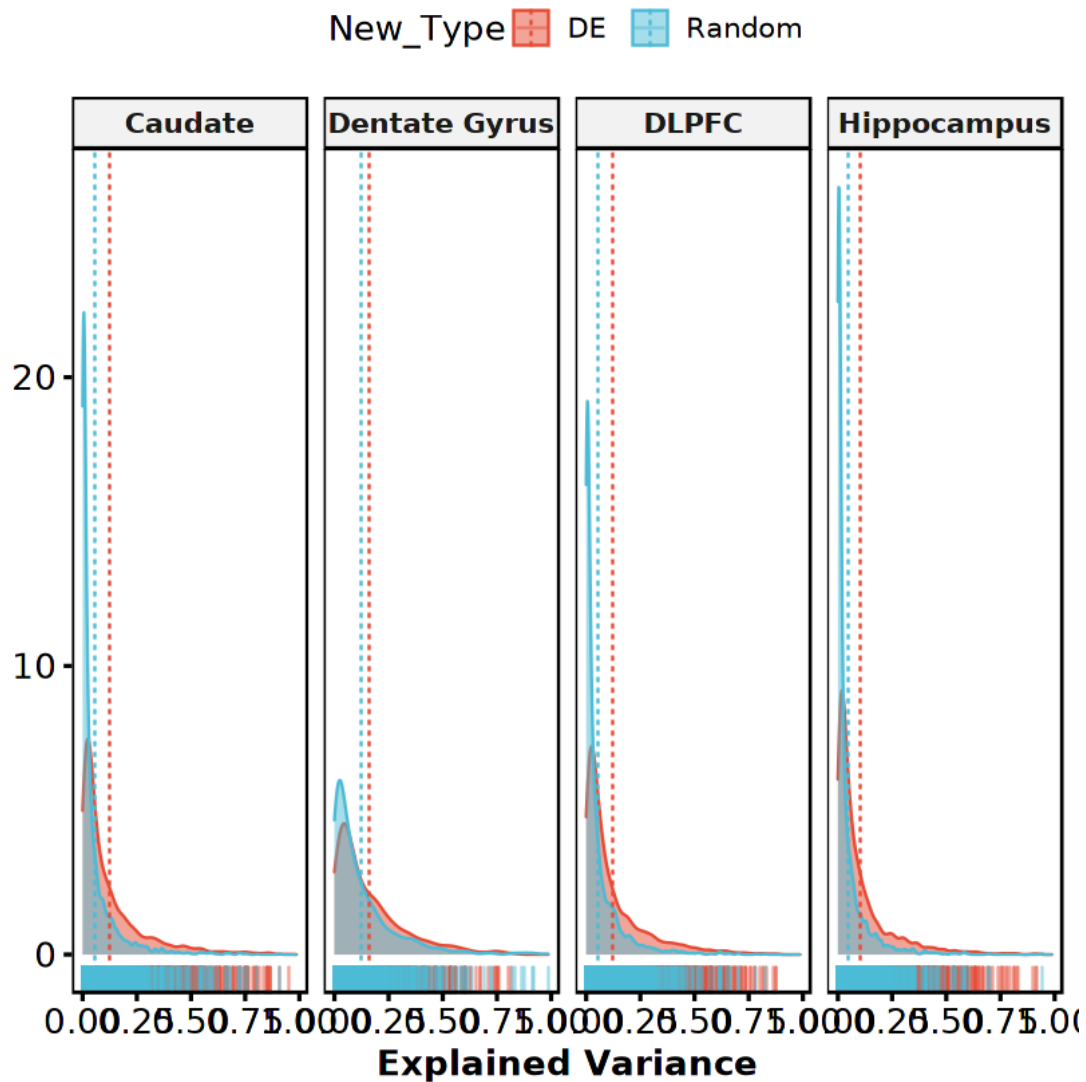


```
[47]: bxp = df2 %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggboxplot(x="Tissue", y="Partial_R2", color="New_Type", add="jitter",
    palette="npg", ylim=c(0, 1), ylab="Explained Variance",
    xlab="Brain Region", ncol=4,
    ggtheme=theme_pubr(base_size=15, border=TRUE)) +
  rotate_x_text(45)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

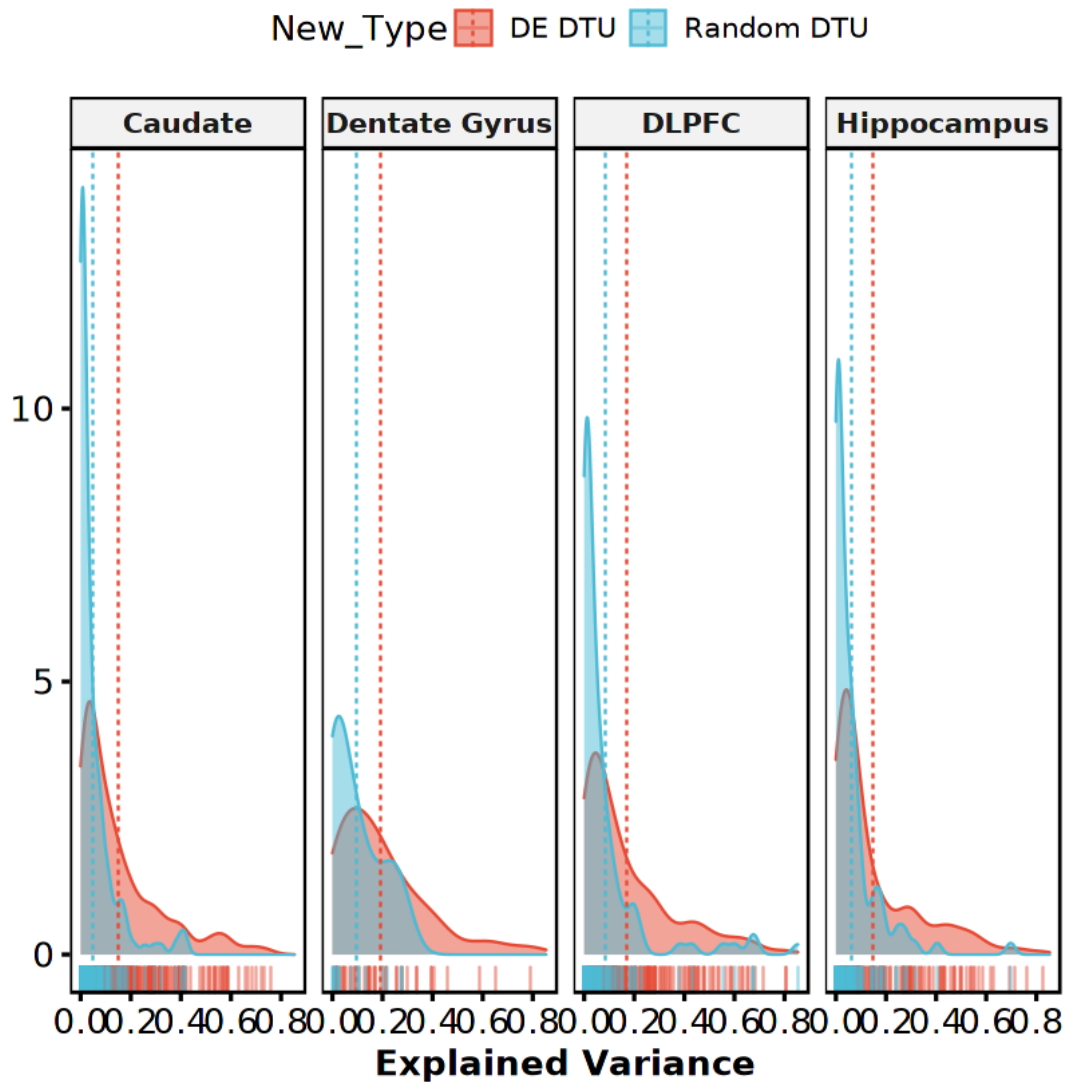


Distribution

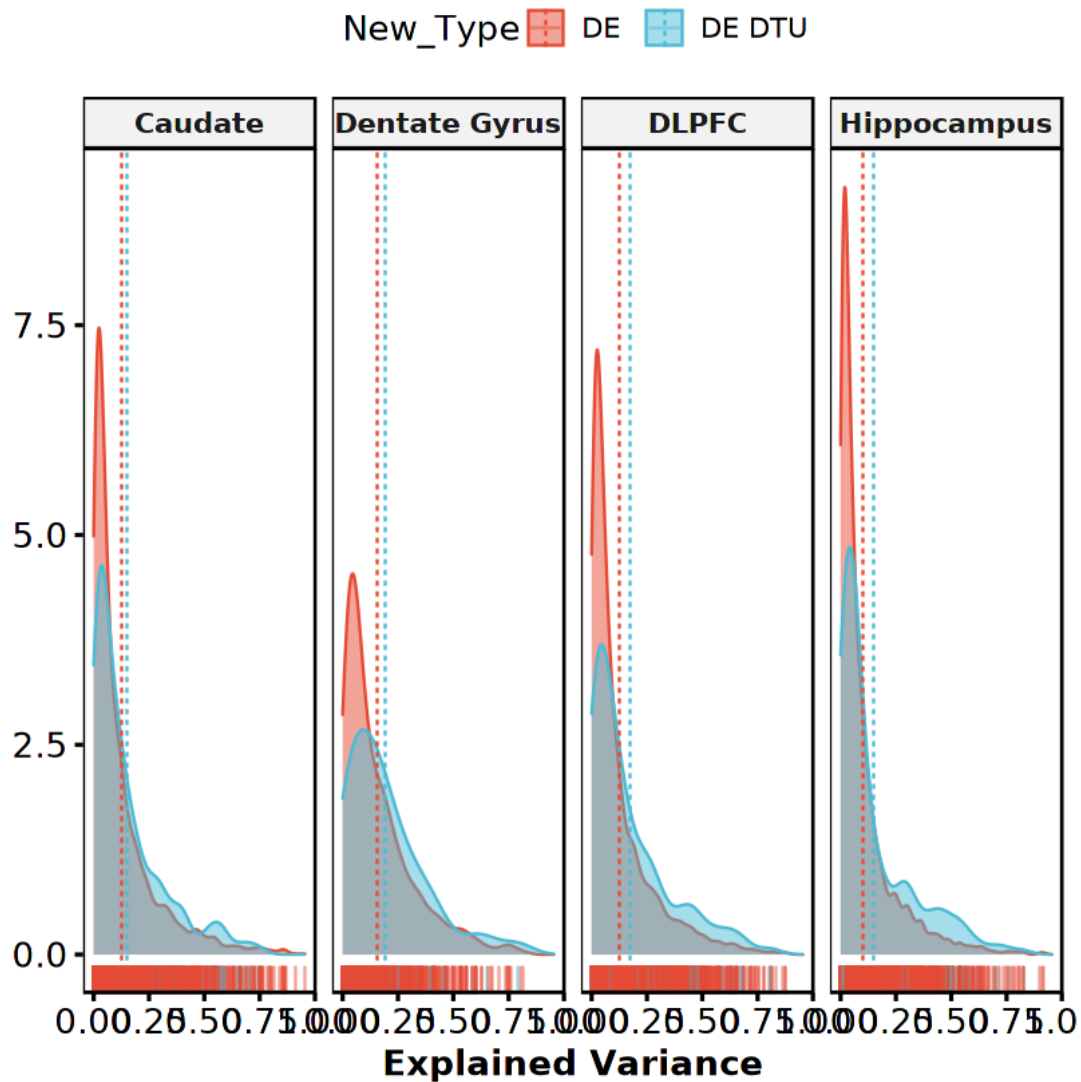
```
[48]: freq = df2 %>% filter(New_Type %in% c("DE ", "Random ")) %>%
  ggdensity(x = "Partial_R2", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("Tissue"), xlab="Explained Variance",
    panel.labs.font=list(face='bold'), ylab="", ncol=4,
    ggtheme=theme_pubr(base_size=15,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



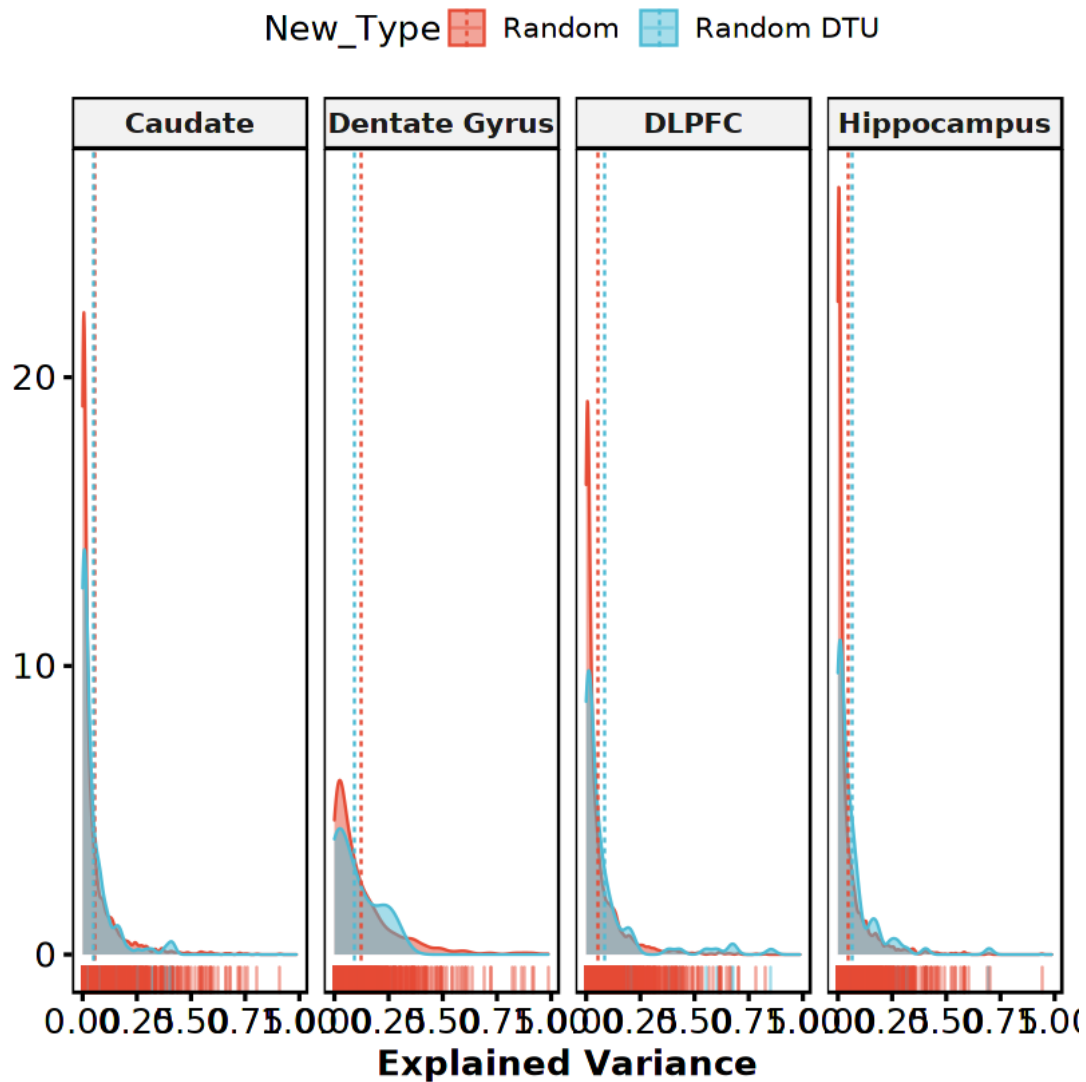
```
[49]: freq = df2 %>% filter(DTU == "DTU") %>%
  ggdensity(x = "Partial_R2", add = "mean", rug = TRUE,
    color = "New_Type", fill = "New_Type", palette="npg",
    facet.by=c("Tissue"), xlab="Explained Variance",
    panel.labs.font=list(face='bold'), ylab="", ncol=4,
    ggtheme=theme_pubr(base_size=15, 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] "2021-09-23 08:20:22 EDT"
```

```
      user  system elapsed
126.315    3.370  107.736
```

```
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    2021-09-23

```

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	0.7.9	2021-07-27	[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	3.0.1	2021-07-17	[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)
fastmap	1.1.0	2021-01-25	[1]	CRAN (R 4.0.2)
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	2.4.3	2021-08-04	[1]	CRAN (R 4.0.3)
hms	1.1.0	2021-05-17	[1]	CRAN (R 4.0.3)
htmltools	0.5.2	2021-08-25	[1]	CRAN (R 4.0.3)

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	1.6.2	2021-07-29	[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	2.5.1	2021-08-19	[1]	CRAN	(R 4.0.3)
Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.0.3)
readr	* 2.0.1	2021-08-10	[1]	CRAN	(R 4.0.3)
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	2.0.1	2021-08-05	[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	1.0.1	2021-07-26	[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	1.7.4	2021-08-25	[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	* 3.1.4	2021-08-25	[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)
tzdb	0.1.2	2021-07-20	[1]	CRAN	(R 4.0.3)
utf8	1.2.2	2021-07-24	[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