Pre-registered report: Space Sequence Synesthesia Diagnostic using form mapping

2025-10-29

Table of contents

1 Abstract:

Existent diagnostic tools for space sequence synethesia are based on questionnaire and response consistency. Consistency is calculated as the area between repetitions for the same inducer. In the first present phase, available data from 467 participants is used to explore new geometrical features to discriminate syntheses from controls. Conceptually, our goal is to take advantage of the inducer's ordinality that create synesthetic forms. For this aim, we harness a geography package to extract geometrical features to use as a test for synethesia. Reciever Operator Characteristics are used to select the features that best diagnose synesthesia. In a second phase to come, we test the predictive power of the new diagnostic features onto newly collected dataset.

2 Introduction

Humans with Sequence Space Synesthesia (SSS) represent ordered sequences in particular spatial positions. For example, August (i.e. the *inducer)* might be represented in the bottom left position (i.e. the position is here the *concurrent),* this position is relative to the concurrent position of the other months which could form a circle all together. In addition time units, numbers also take particular forms (Galton 1880), not to be confused with the Mental Number Line [Dehaene to be found]. These forms are idiosyncratic, meaning they might vary across individuals. This makes it difficult to detect authentic SSS and therefore give precise estimates of prevalence in the population. Estimated prevalence for SSS in the general population spans between 4.4 % (Brang et al. 2013) and 14.2 % (Seron et al. 1992), see also (Ward et al. 2018; Sagiv et al. 2006). Hence a reliable diagnostic tool to detect SSS would also be useful to investigate SSS.

Diagnostic depends on the definition of the conditions under investigation. A strict definition of Synestheesia requires five different criteria (Deroy and Spence 2013). *Automaticity*: the *inducer* automatically triggers the *concurrent*. For example august might automatically trigger it's specific spatial location. *Unidirectionality*: while the *inducer* triggers the concurrent, the concurrent does not trigger the inducer. Hence the bottom left position doe not trigger August. *Consciousness*: The concurrent is consciously percieved. *Developmentally early*: the experience was already present during childhood. *Consistency*: the inducer-concurrent pair remains stable in time. For example, August triggers the same bottom left position. Consistency is arguably the most suited criteria to develop a diagnostic tool since it is relatively simple to implement in a behavioral task and quantify.

Hence given consistency, similar concurrent responses triggered by the same inducers can be used as a marker for authentic SSS. Consistency test have become golden standart for colour-grapheme synesthesia, where an inducer is presented (i.e. letters of the alphabet) and the participant is requested to selected the concurrent colour, using a colour picker. Individual consistency is then calculated as the distance between repeated colour responses to the same inducers. Interestingly, the best colour space to detect colour-grapheme synesthesia is CIE*LUV, a colour space developed to be isoform to human perception (Rothen et al. 2013). Analogously to grapheme-colour synesthetes, consistency test can be used to diagnose SSS. In that tasks, it is repeatedly asked to report the position of the inducers on a screen. The total area between the responses of same inducer (i.e. a triangle if repeated three times) is then used as characteristic to diagnose SSS. The rationale being that consistent responses would lead to smaller area than inconsistent ones (Rothen et al. 2016). This method resembles how number forms are describe in the single case study (Piazza, Pinel, and Dehaene 2006), see Experiment 1.

However characterizing synesthetes from non synesthetes using total area has several limitations. For example high consistency by non-synesthetes can be achieved by giving all responses on the same screen position (i.e. false positive). Moreover, this kond of criteria might bias the diagnosis to include synesthetes with straigth lines which leads to less variability than more complex forms(**?**).

The goal of the present registered report is to first identify new features characterizing synesthetes responses based on already available datasets and test the best working features on a future dataset. The new features are desgined to take advantage of two properties of synesthetic responses that have not been included in precedent consistency tests. First, sequentiality on top of single inducer responses the ordered position between subsequent induces is important. For example the relative position of August and the other months. From numerical cognition, ordinality has been aknowledged to be an important semantic property of numbers, also given their sequential acquisition (i.e. 1 is learned before 2). Second, thee particular synthetic forms of the sequential spatial location. These forms might have geometrical properties. For example months of the year might be represented circularly (as already described by (Galton 1880) for numbers).

To take advantage of sequential and geometrical synesthetic forms, we harnessed a geo-spatial package(Pebesma 2018) to extract geometrical features from participant x and y coordinate responses. This packages allows for example to build string or polygons for each repetition and compare different geometrical features. Those individual geometrical features are then compared using Receiver Operator Charactheristics (ROC) between individuals grouped as synesthes and control. In the present *phase I*,we compare ROC on three merged derivation-datasets using the same task on SSS Ward (n.d.). In future *phase II*, we compare whether the features selected to diagnose SSS in *phase I,* on a validation dataset that is not yet acquired (registered report on the open science foundation: https://osf.io/9efjb/**).**

3 Methods

Phase I: present analyses. First, we reproduce the diagnostic criteria of each respective dataset. Second, we merge the dataset and compare the diagnostic criteria across datasets using Reciever Operator Charachteristics (ROC). Third, we compare wheter the featuers lead to somilar ROC charachteristics across the different sets (i.e. for months, weeks and numbers). Fourth, we compute new candidate geometrical features that could be used to diagnose SS. Finally we summarize and compare all ROC and select the best features that class synesthetes from control with the merged dataset.

Phase II: future analyses. On a future dataset using the same task, we will compare the predictive power of the selected features using ROC.

3.1 Materials

A the exception of (Rothen et al. 2016) (see https://osf.io/6hq94/files/osfstorage), the data from (Ward, n.d.; Van Petersen et al. 2020a) were collected online. The 29 inducers were: the 12 months of a year, 7 days of the week and 10 numbers (i.e. hindo-arabic numerals from 0 to 9). (Van Petersen et al. 2020a) Also presented 50 and 100 numerals, which we excluded here. (Ward, n.d.) data is collected using the Syntoolkit.

3.2 Procedure

The details for each procedure is described in each respective article (Rothen et al. 2016; Ward, n.d.; Van Petersen et al. 2020a), here we describe the common task.

Each participant is presented with one one inducer at a time at the center of a otherwise white screen. The participant is instructed to click at the screen position that they visualize them. Inducers order is randomized and each inducer is repeated three times.

The order of the stimuli was randomised, but such that no stimulus was repeated until the previous batch of unique stimuli (N = 29) had been presented.

4 *Phase I* Methods

The data for *phase I,* comes from: (Rothen et al. 2016),(Ward, n.d.) (from: https://osf.io/p5xsd/files/osfstorage) and (Van Petersen et al. 2020b)

- Root (Root et al. 2021)

```
[1] 0
```

4.1 *Phase I.* Population

dataSource	Ctl	Syn
PeterCor	21	12
Rothen	37	33
Ward	215	252

4.2 *Phase I.* Analysis

First, we replicate consistency methods found in the literature using the same task ((Rothen et al. 2016; Ward, n.d.; Van Petersen et al. 2020a; Root et al. 2021)) and compare the results.

Second, we extract features based on the form. (C) We harness a geography package to compute segment based features (D) We compute polygon based features. (E) Convex Hull (F) Angles.

- Each feature is presented with the following structure:
  - Compute Feature
  - Example
  - Receiver Operator Characteristics (ROC)

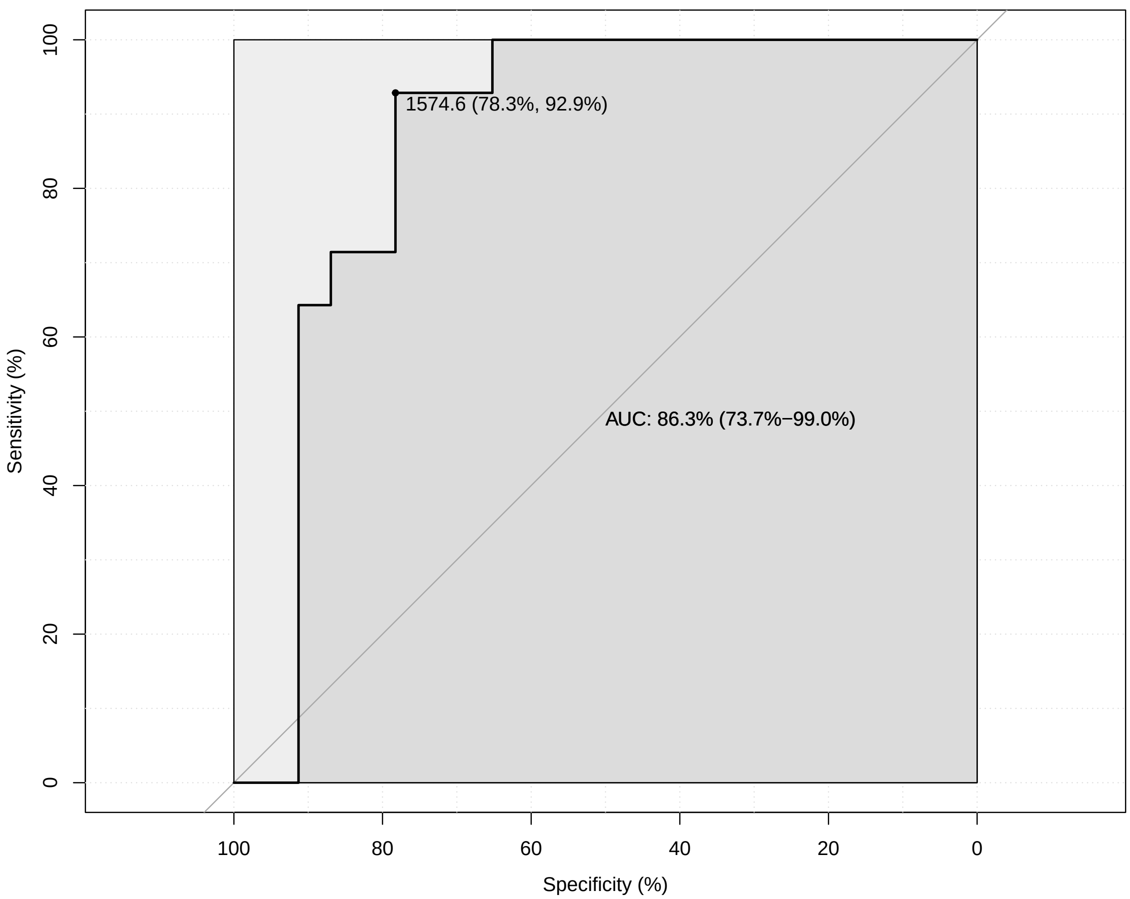
5 *Phase I.* Results Reproduce

5.1 Triangle area

Definition: Calculating consistency Each stimulus is represented by three xy coordinates - (x1, y1), (x2, y2), (x3, y3) - from the three repetitions. For each stimulus, the area of the triangle bounded by the coordinates is calculated as follows:
$Area = (x1y2 + x2y3 + x3y1 – x1y3 – x2y1 – x3y2)/2$

5.2 Reproduce Rothen et al., 2016.

Here we reproduce (Rothen et al. 2016) ROC results:



Feature	AUC	threshold	sensitivity	specificity	ppv	npv	ci_low	ci_high
triangle_area_GA_Rothen	86.3354	1574.552	92.85714	78.26087	72.22222	94.73684	73.65815	99.01265

group	n	Mean	SD
Ctl	37	8084.1979	12753.8026
Syn	33	834.9606	516.0656

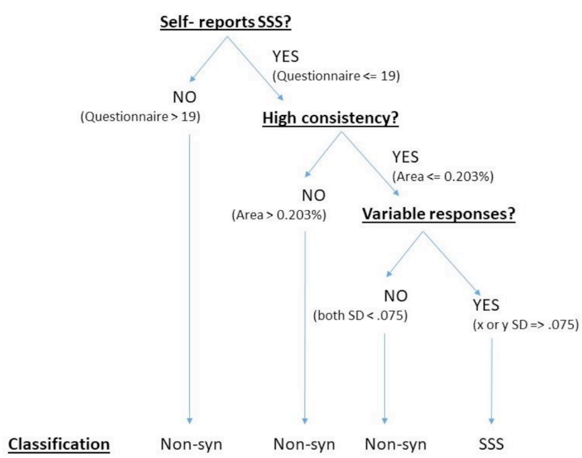
```
                  Feature     AUC threshold sensitivity specificity      ppv
1 triangle_area_GA_Rothen 86.3354  1574.552    92.85714    78.26087 72.22222
       npv   ci_low  ci_high
1 94.73684 73.65815 99.01265
```

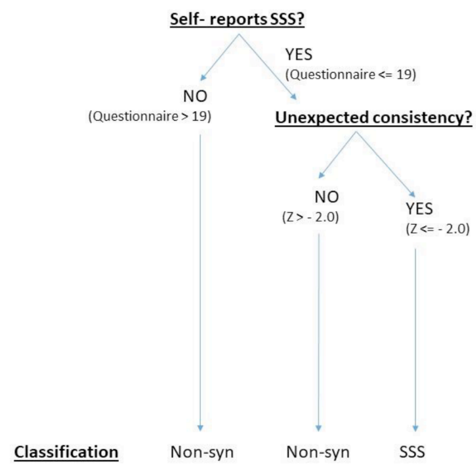
5.2.1 Summary Rothen vs Reproduction

	Description	DP	AUC	Mean (syn)	Mean (con)	SD (syn)	SD (con)	Sensitivity	Specificity	Cut-off
Rothen	Area	1.57	0.76	1’079	7’031	1365	11’149	88	70	1’596
Repro			0.75	**1’312**	7’031	**1829**	11’303	**85**	70	1’575
Repro_Na			0.76	930	7’031	745	11’303	90	70	1’574

5.3 Reproduce Ward, 2020:

(Ward, n.d.) combines different individual measures and features to diagnose synesthesia in comparison to randomly permuted z-score chance level thresholds:





Since we do not have questionnaire for all the data, we will only try to reproduce the consistency and sd combination.

The mean area is calculated by adding together the area for each stimulus and dividing by 29. This unit is transformed into a percentage area taking into account the different pixel resolution of each participant.Mean area $= SummedArea/ScreenArea$, where: $ScreenArea = Xpixels * Ypixels$

```
Adding missing grouping variables: `ID`
```

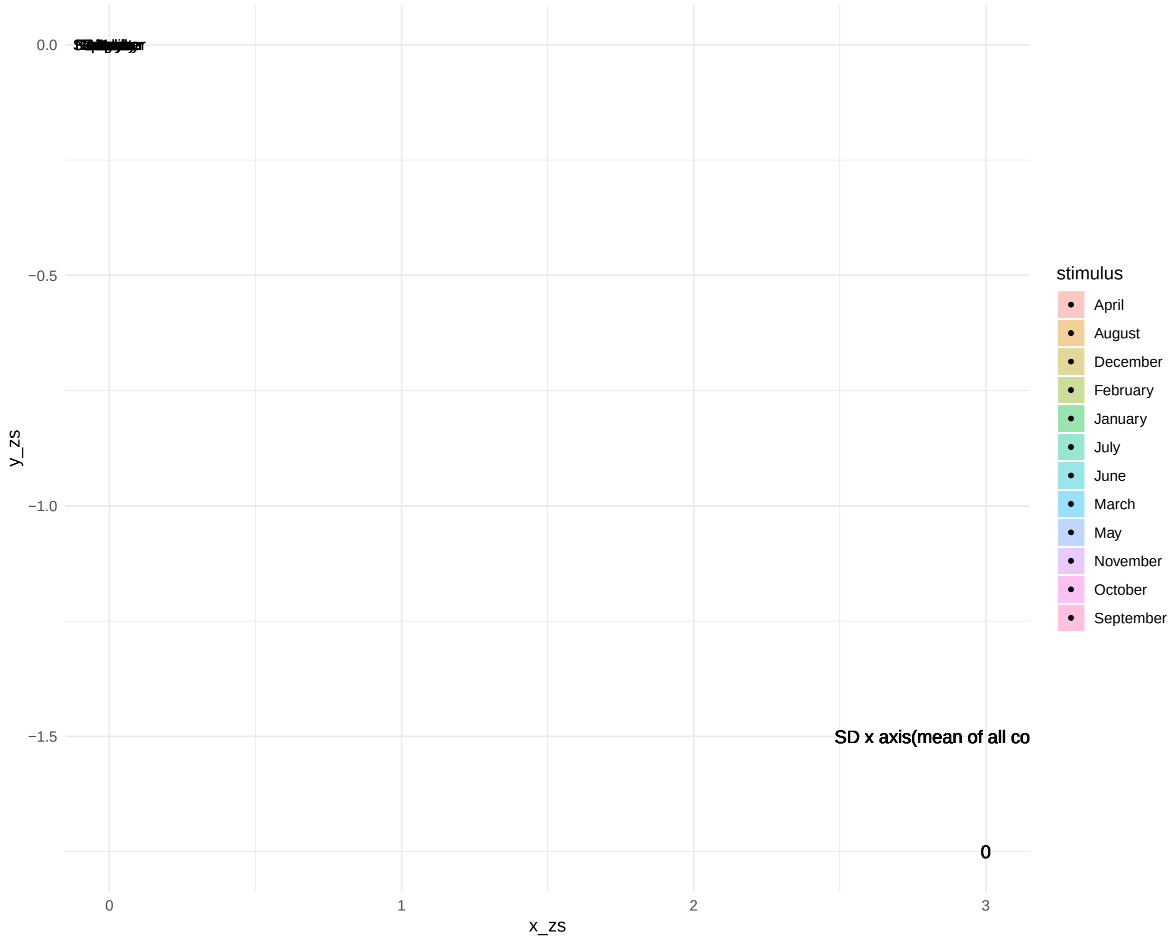
5.3.0.1 WIP here:

5.4 Add SD

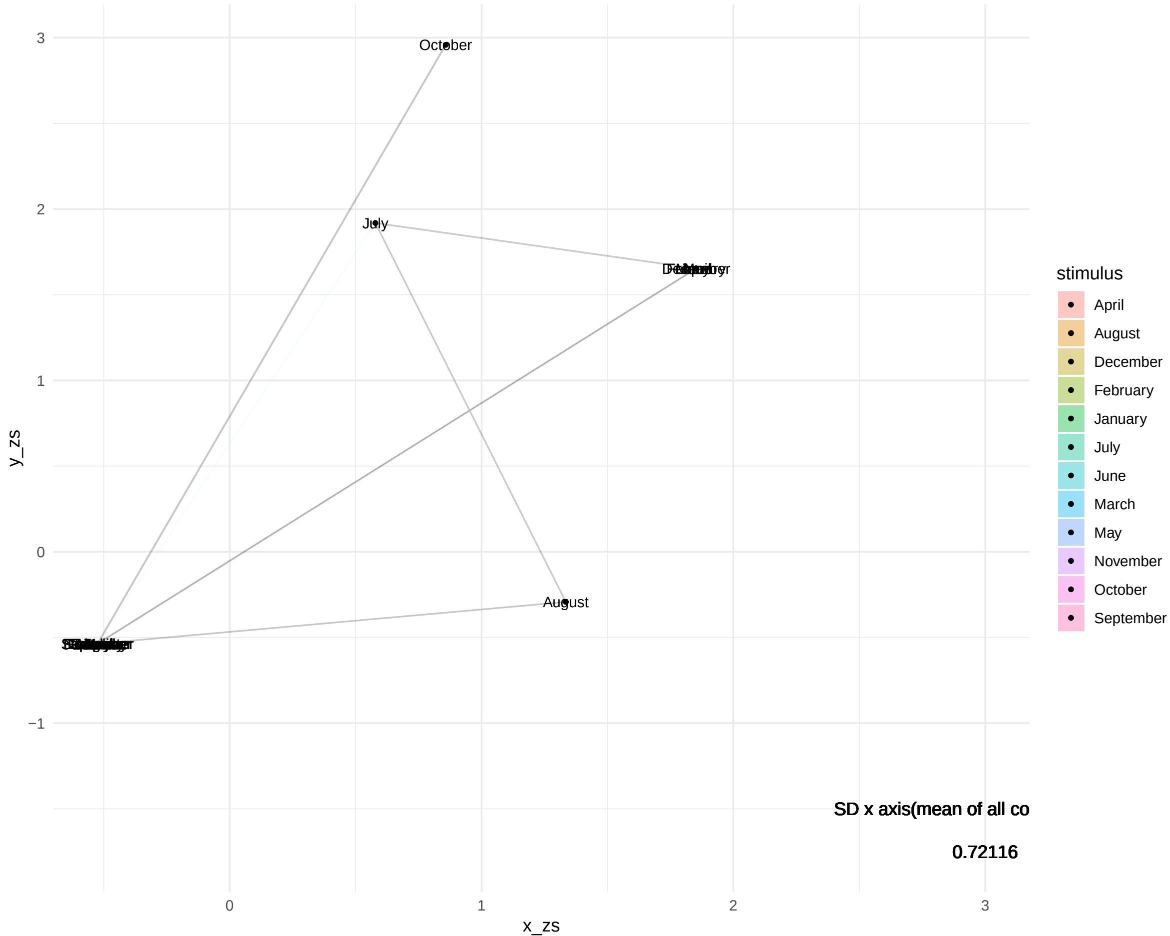
5.4.1 Example

Would need an example with all in the center

```
Warning: There were 570 warnings in `filter()`.
The first warning was:
i In argument: `ID == ds_Q$ID[ds_Q$SD_ID_xsc == min(ds_Q$SD_ID_xsc)]`.
i In group 1: `ID = "1005_SeMi"`.
Caused by warning in `ID == ds_Q$ID[ds_Q$SD_ID_xsc == min(ds_Q$SD_ID_xsc)]`:
! longer object length is not a multiple of shorter object length
i Run `dplyr::last_dplyr_warnings()` to see the 569 remaining warnings.
```

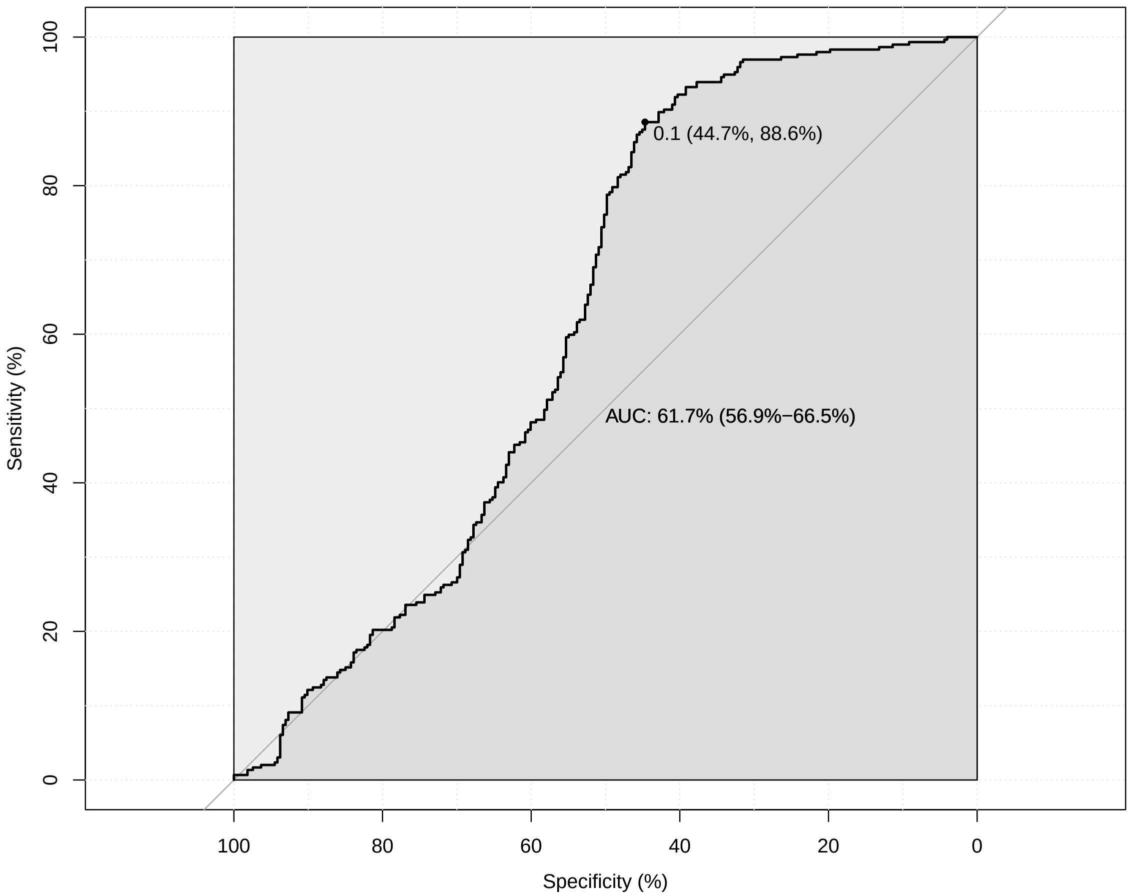


0.0
−0.5
−1.0
−1.5
y_zs
0
1
2
3
x_zs
SD x axis(mean of all co
0
stimulus
April
August
December
February
January
July
June
March
May
November
October
September



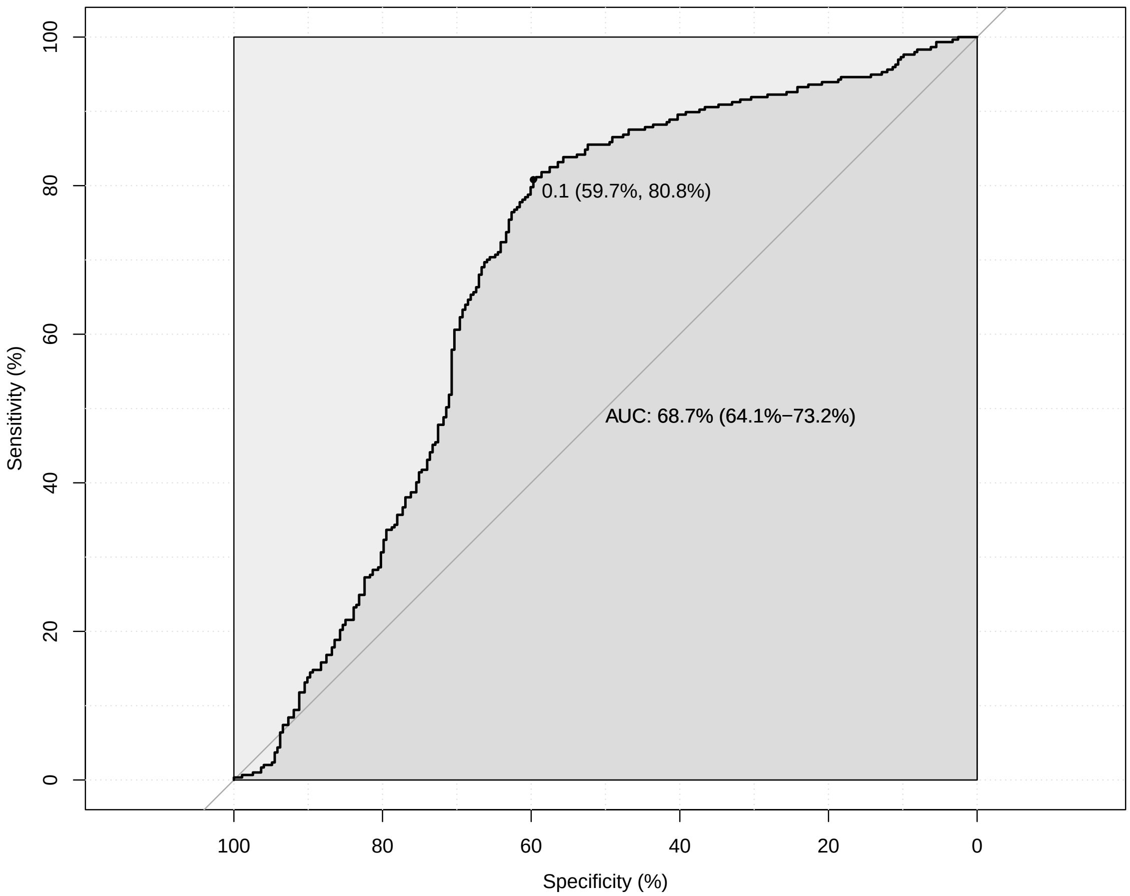
5.4.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
    Feature    AUC threshold sensitivity specificity      ppv      npv   ci_low
1 SD_ID_xsc 61.684 0.1216165    88.55219    44.68864 63.52657 78.20513 56.86332
   ci_high
1 66.50467

$Coningency_table

      Ctl            Syn
  Ctl "122 (44.7%)" "151 (55.3%)"
  Syn "34 (11.4%)"  "263 (88.6%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean     SD
```

```
  <fct> <int> <dbl>  <dbl>
1 Ctl     273 0.165 0.124
2 Syn     297 0.224 0.0827

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 61.68%
95% CI: 56.86%-66.5% (DeLong)

Setting levels: control = Ctl, case = Syn

Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
    Feature     AUC threshold sensitivity specificity      ppv      npv
1 SD_ID_ysc 68.6696 0.1157444    80.80808    59.70696 68.57143 74.09091
    ci_low  ci_high
1 64.12539 73.21381

$Coningency_table

      Ctl            Syn
  Ctl "163 (59.7%)" "110 (40.3%)"
  Syn "57 (19.2%)"  "240 (80.8%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean     SD
```

```
  <fct> <int> <dbl>  <dbl>
1 Ctl     273 0.124 0.117
2 Syn     297 0.196 0.0940

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 68.67%
95% CI: 64.13%-73.21% (DeLong)
```

```
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

5.5 Reproduce Root 2021

(Root et al. 2021) suggested to use random permutations to calculate individual chance levels of consistency. Individual x and y coordinates (29 (inducers) *3 (repetitions) = 87) are randomly shuffled across conditions and inducers and areas are calculated for each 1000 permutations. Hence giving rise to individual distribution for chance level of consistency. Z-score is then computed to compare the observed with the permuted consistencies:

$Zscore = \frac{ObsConsistency - mean(PermConsistency)}{SD(PermConsistency)}$

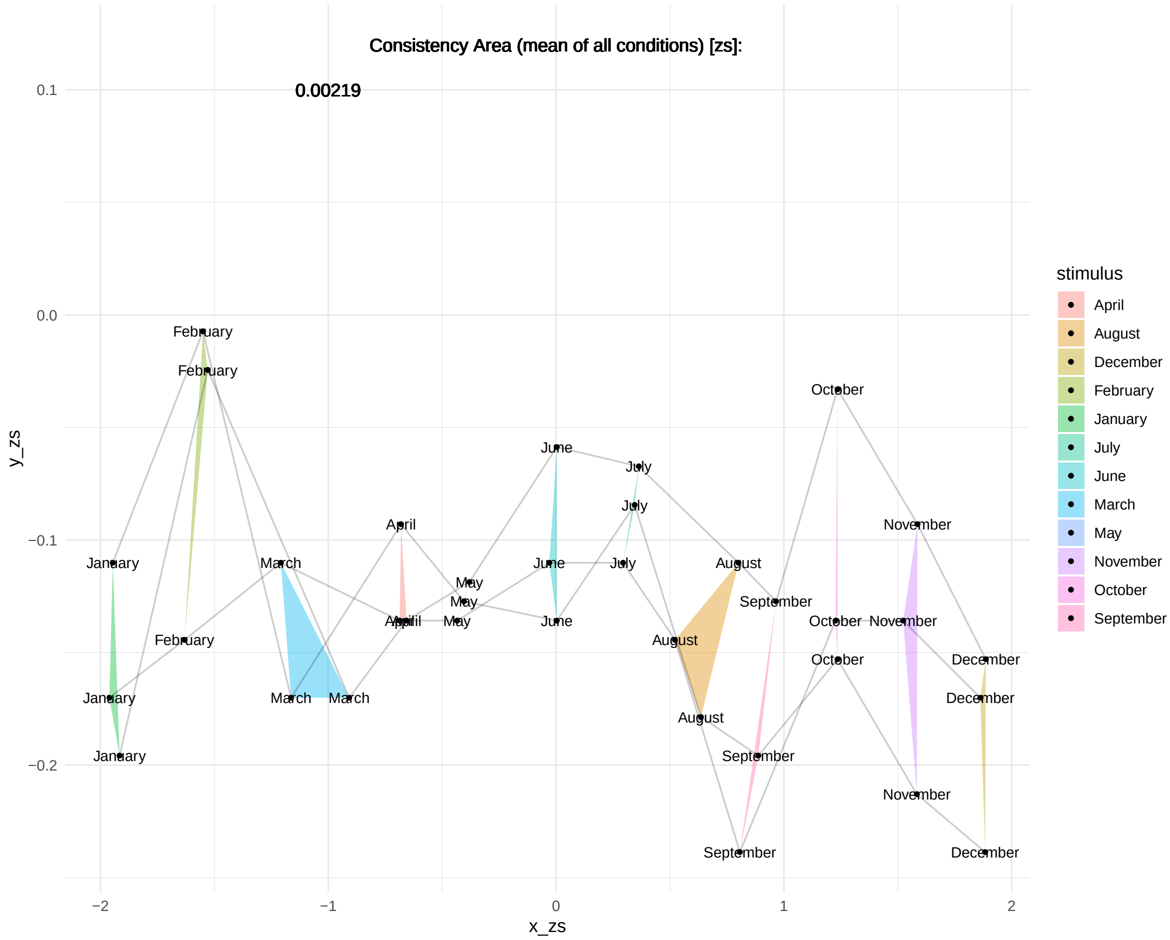
5.5.1 Example

5.5.2 ROC

5.6 Reproduced features on merged data:

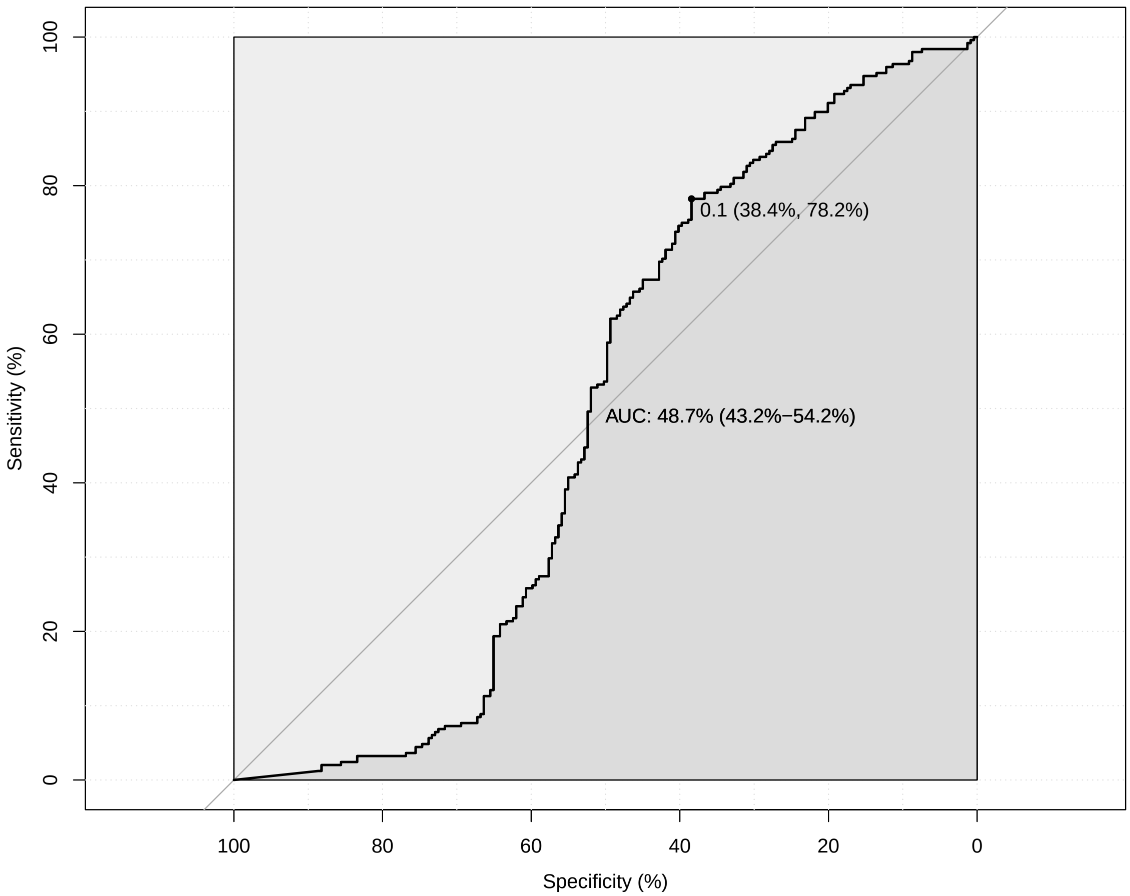
5.6.1 Area Consistency

5.6.1.1 Example



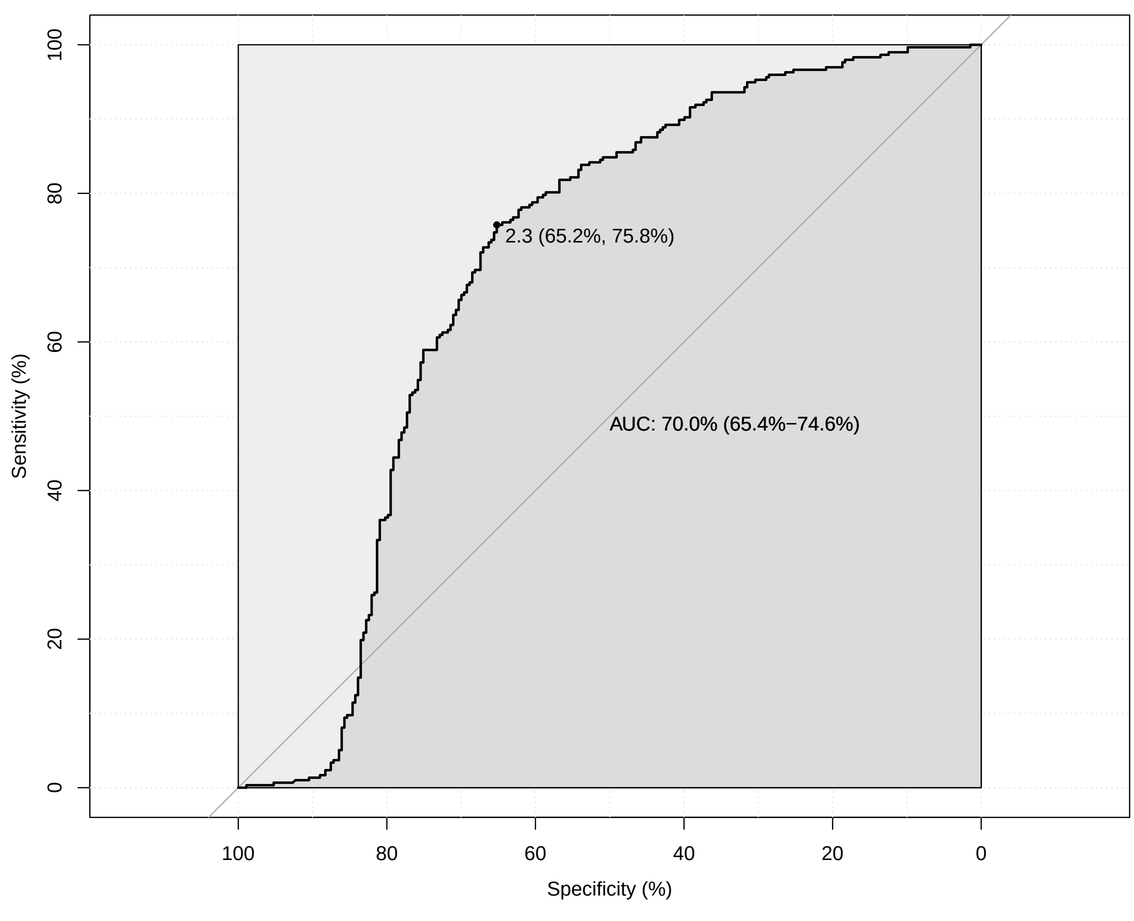
5.6.1.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
      Feature     AUC  threshold sensitivity specificity      ppv      npv
1 Consistency 48.7146 0.05868314    78.22581    38.42795 57.91045 61.97183
    ci_low  ci_high
1 43.20307 54.22614

$Coningency_table

      Ctl          Syn
  Ctl "88 (38.4%)" "141 (61.6%)"
  Syn "54 (21.8%)" "194 (78.2%)"

$Descr_table
# A tibble: 2 x 4
  group     n   Mean    SD
```

```
  <fct> <int>  <dbl> <dbl>
1 Ctl     273 0.116  0.227
2 Syn     297 0.0682 0.142

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 229 controls (data[[group_col]] Ctl) > 248 cases (data[[group_col]]
Area under the curve: 48.71%
95% CI: 43.2%-54.23% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```



```
$ROC_properties
         Feature     AUC threshold sensitivity specificity     ppv  npv
1 Consistency_zs 69.9813  2.272523    75.75758    65.20147 70.3125 71.2
    ci_low  ci_high
1 65.40126 74.56125

$Coningency_table

      Ctl           Syn
  Ctl "178 (65.2%)" "95 (34.8%)"
  Syn "72 (24.2%)"  "225 (75.8%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
  <fct> <int> <dbl> <dbl>
1 Ctl     273  7.69  9.35
2 Syn     297  2.22  4.01

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) > 297 cases (data[[group_col]]
Area under the curve: 69.98%
95% CI: 65.4%-74.56% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

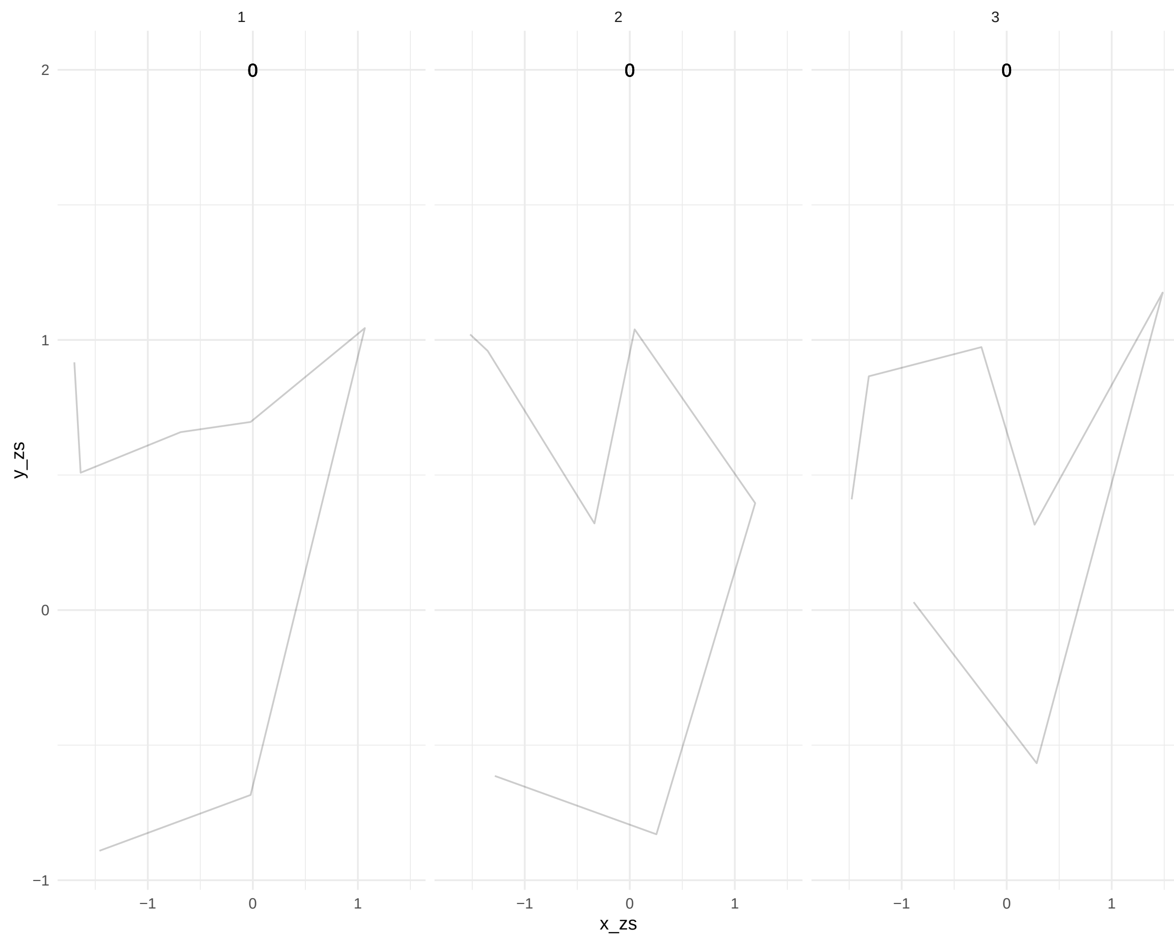
6 *Phase I.* Results: Novel features

6.1 Segment self-intersection

For each category we compute the number of times the path intersects within each repetition. This can be conceptualized as drawing a segment between the ordered inducers of each category (i.e. between 0 and 9 for numbers) and count how many line intersect. Hence the number of segment is `length(stimuli)-1`, for each participant we sum the number of self-intersections.

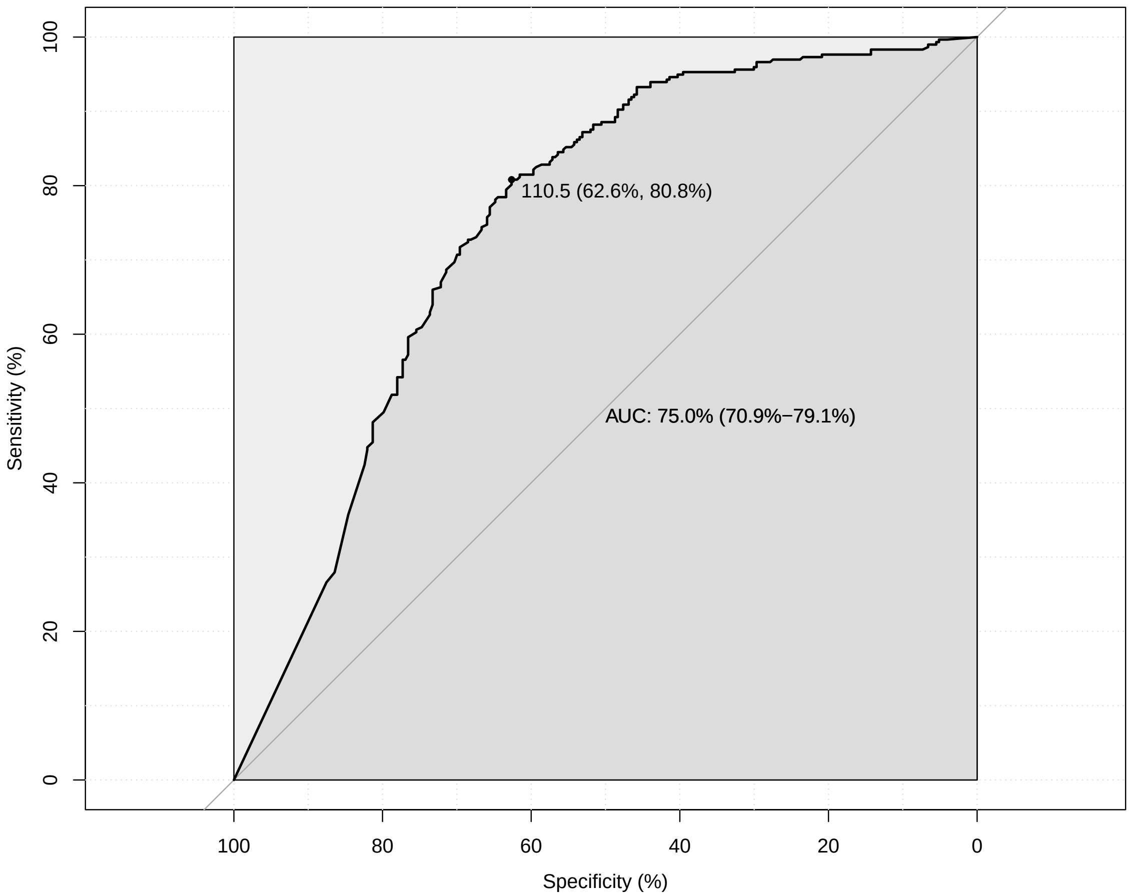
Note: ideally we should shuffle the repetitions.

6.1.1 Example



6.1.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
          Feature     AUC threshold sensitivity specificity      ppv npv
1 SumID_lineInter 74.9769     110.5    80.80808    62.63736 70.17544  75
   ci_low  ci_high
1 70.8766 79.07715

$Coningency_table

      Ctl           Syn
  Ctl "171 (62.6%)" "102 (37.4%)"
  Syn "57 (19.2%)"  "240 (80.8%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

```
  <fct> <int> <dbl> <dbl>
1 Ctl     273  960. 1062.
2 Syn     297  173.  498.

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) > 297 cases (data[[group_col]]
Area under the curve: 74.98%
95% CI: 70.88%-79.08% (DeLong)
```

```
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

7 Segments (with sf)

We will take advantage of the `sf` package and connect the x and y coordinates of ordered inducer with a segment. Sf hates NaN's. Either convert them to 0 (as originally) or remove them. I'll start converting to 0.

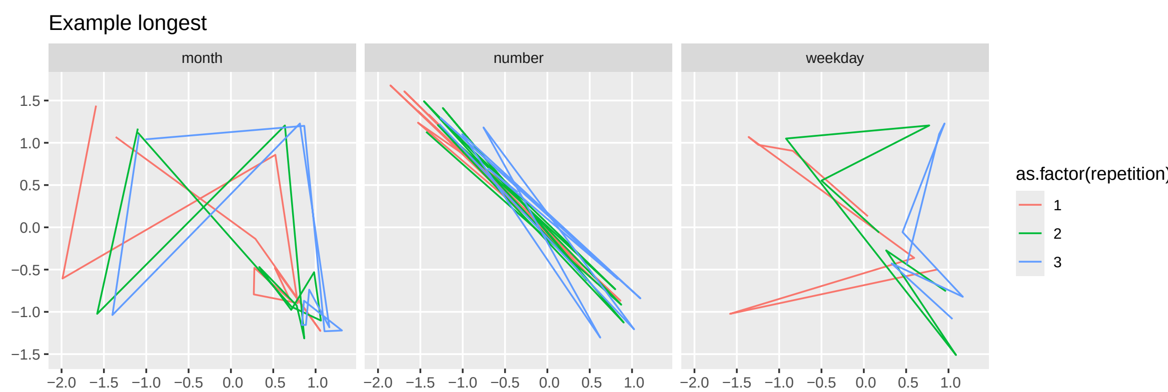
```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
Spherical geometry (s2) switched off
```

```
Warning: Using one column matrices in `filter()` was deprecated in dplyr 1.1.0.
i Please use one dimensional logical vectors instead.
```

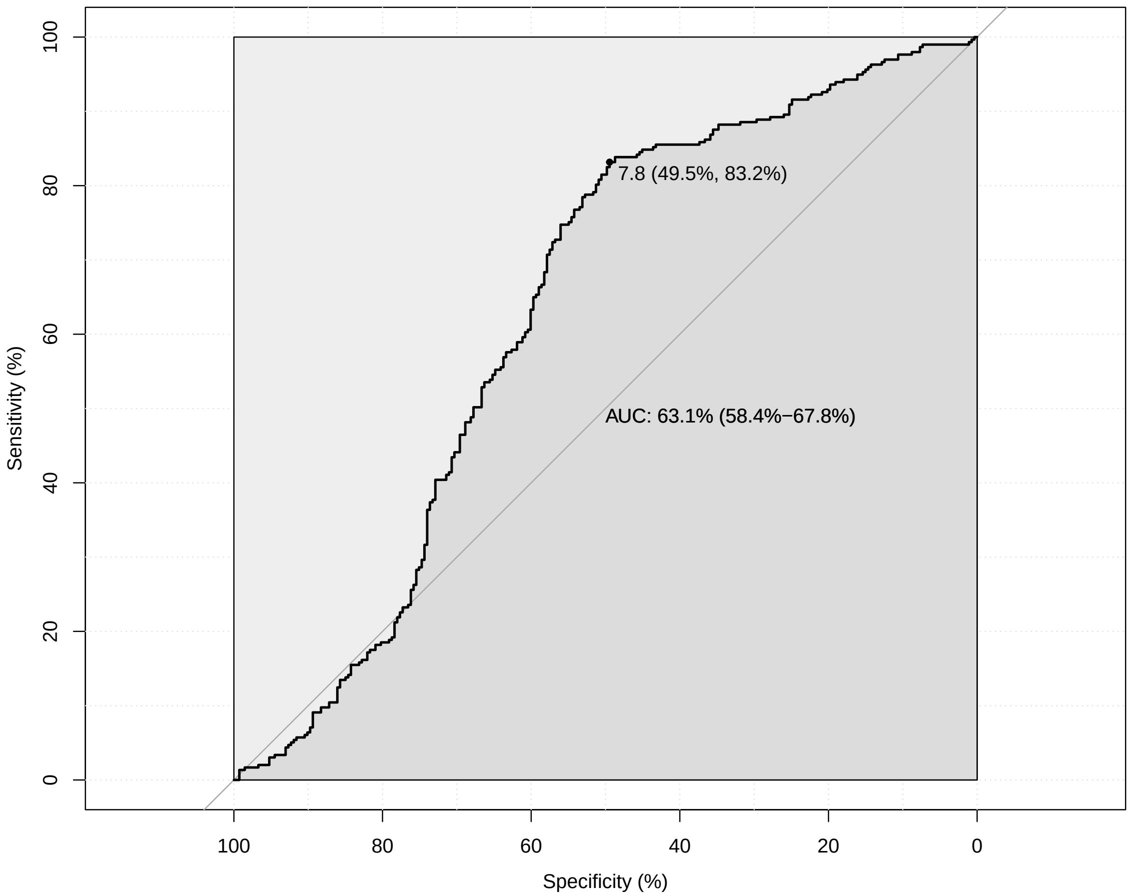
8 Segment length (should replicate Rothen)

8.1 Example



8.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
      Feature     AUC threshold sensitivity specificity      ppv      npv
1 GA_segm_leng 63.1036  7.755259    83.16498    49.45055 64.15584 72.97297
    ci_low  ci_high
1 58.36623 67.84089

$Coningency_table

      Ctl            Syn
  Ctl "135 (49.5%)" "138 (50.5%)"
  Syn "50 (16.8%)"  "247 (83.2%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

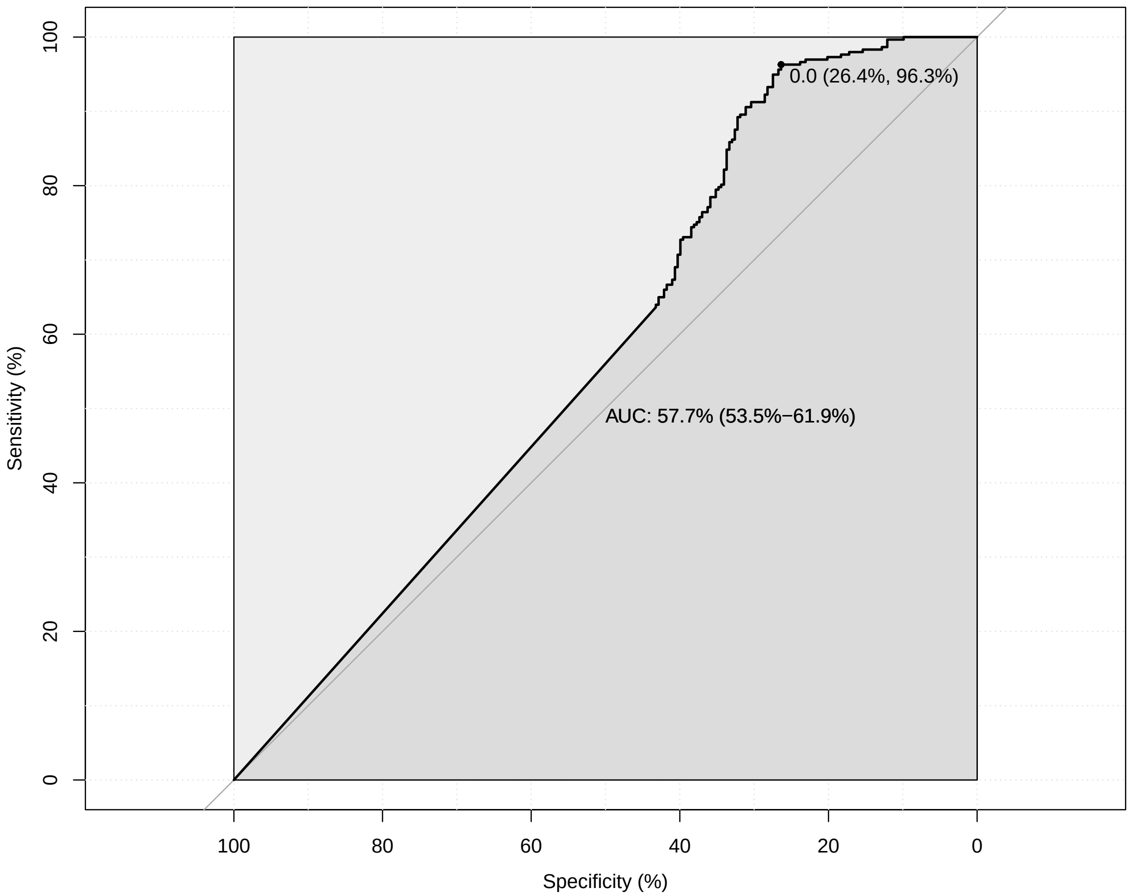
```
  <fct> <int> <dbl> <dbl>
1 Ctl     273  7.56  3.13
2 Syn     297  6.22  2.24

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) > 297 cases (data[[group_col]]
Area under the curve: 63.1%
95% CI: 58.37%-67.84% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

9 Distances between repetitions

9.1 Example

9.2 ROC

```
Setting levels: control = Ctl, case = Syn
```



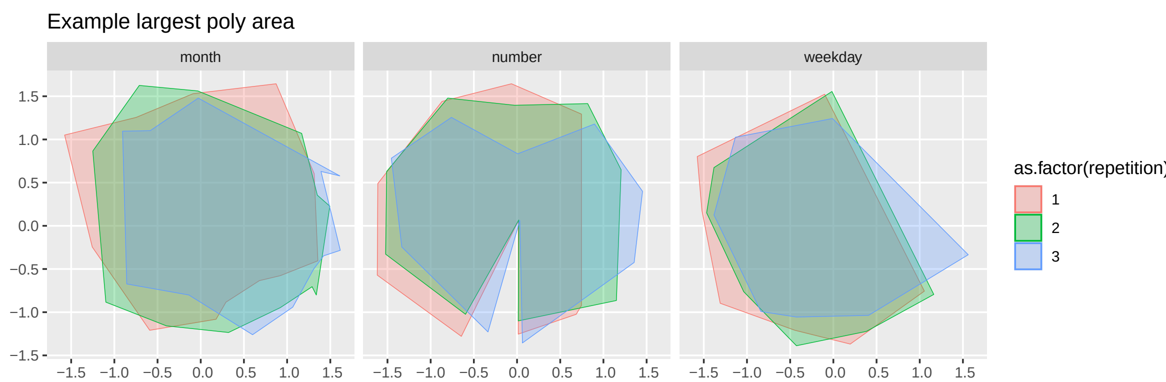

```
$ROC_properties
     Feature     AUC  threshold sensitivity specificity      ppv      npv
1 GA_BtwDist 57.6664 0.03093695     96.2963    26.37363 58.7269 86.74699
    ci_low  ci_high
1 53.45119 61.88163

$Coningency_table

      Ctl          Syn
  Ctl "72 (26.4%)" "201 (73.6%)"
  Syn "11 (3.7%)"  "286 (96.3%)"

$Descr_table
# A tibble: 2 x 4
  group     n    Mean     SD
```

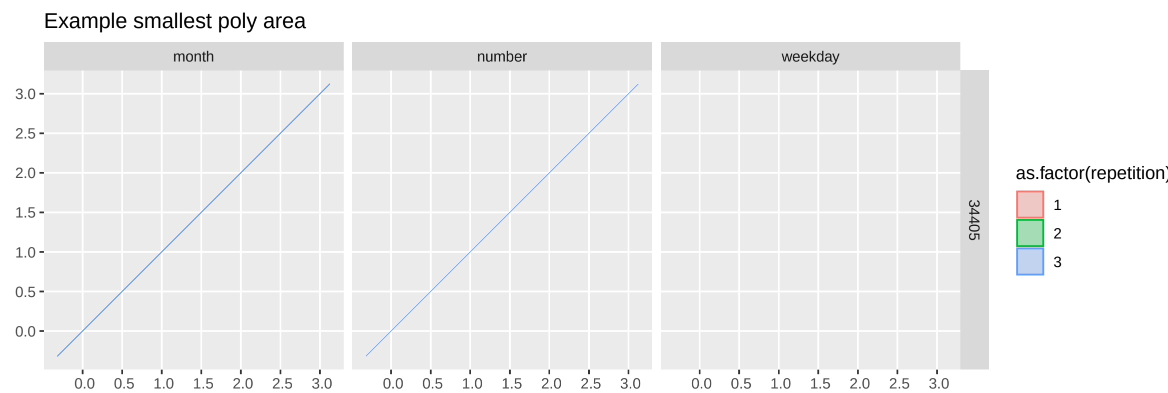
```
  <fct> <int>   <dbl>  <dbl>
1 Ctl     273 0.0785  0.172
2 Syn     297 0.00865 0.0410

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) > 297 cases (data[[group_col]]
Area under the curve: 57.67%
95% CI: 53.45%-61.88% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

10 Polygon based geometries

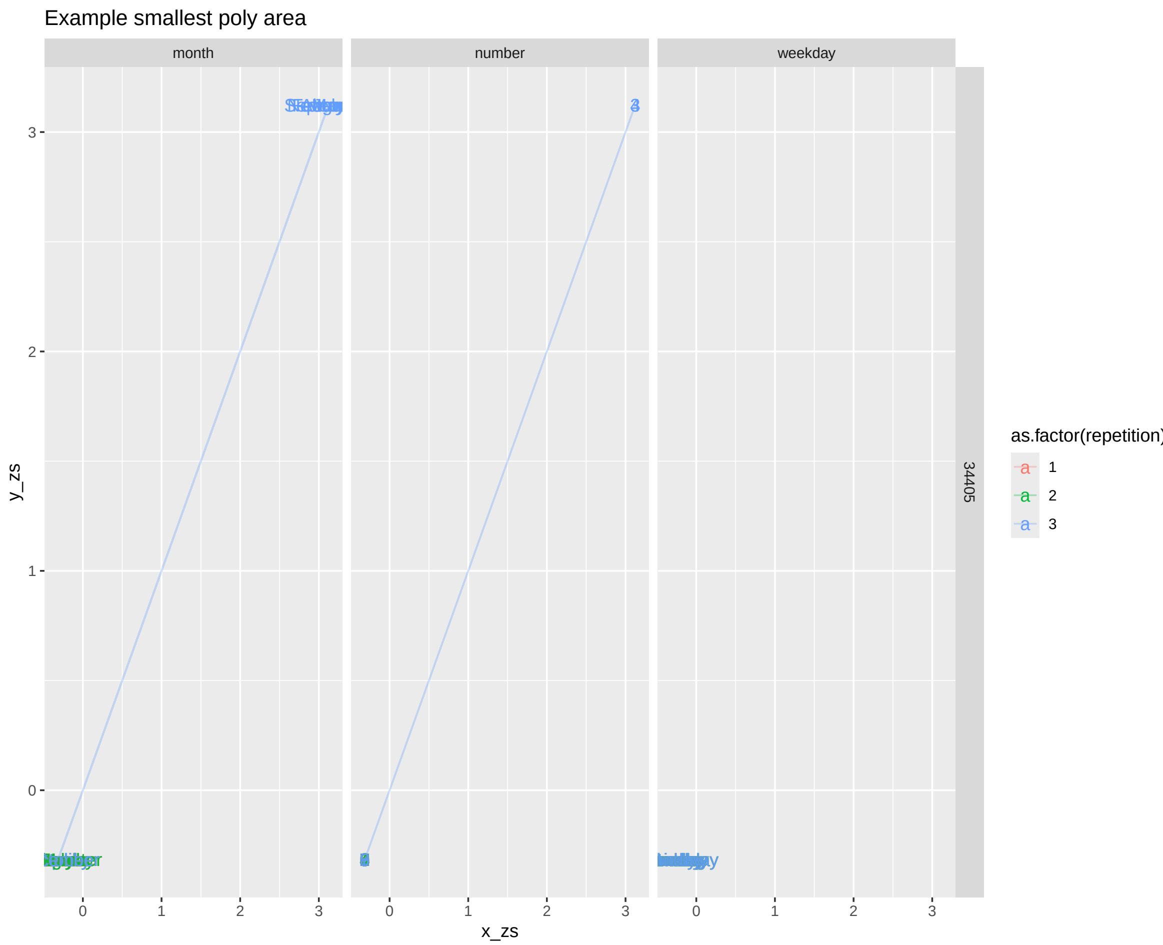
11 Polygon area

11.1 Example



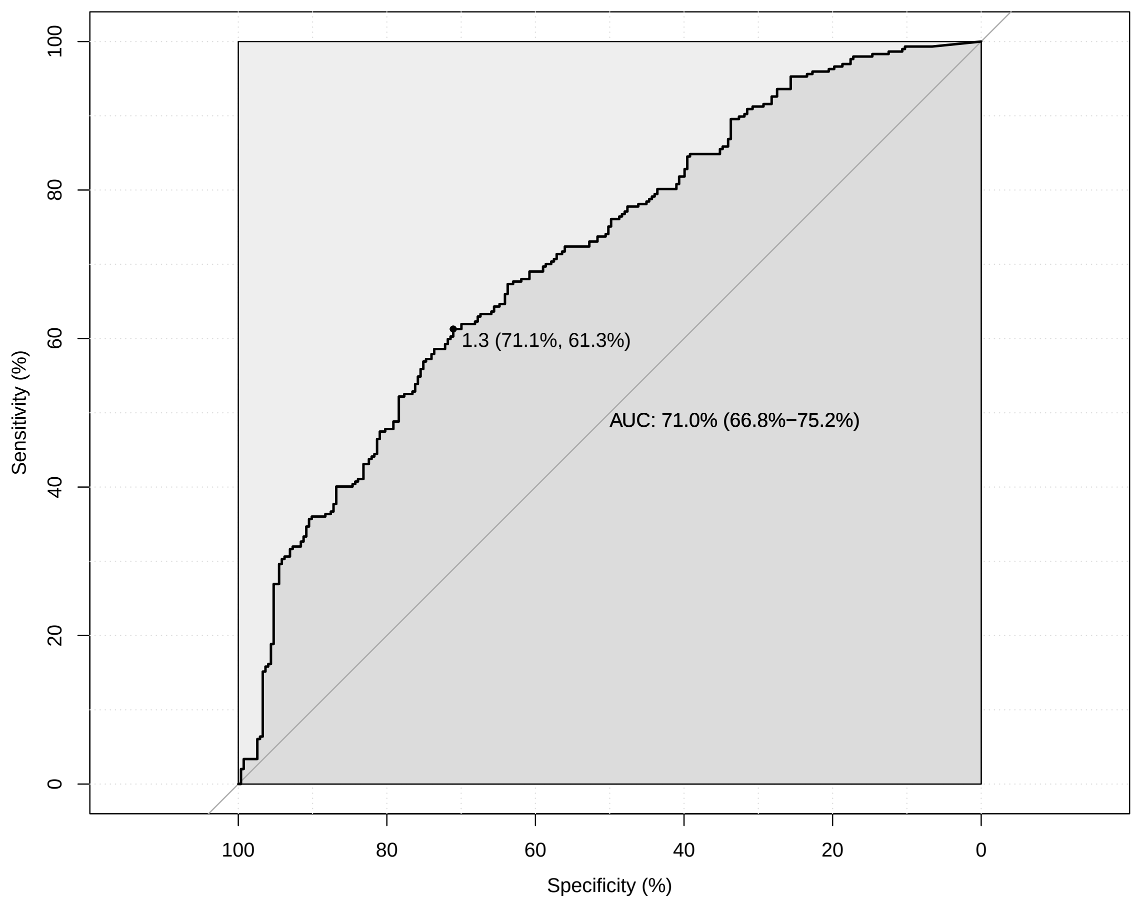


Example smallest poly area
month
number
weekday
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
34405
as.factor(repetition)
1
2
3



11.2 ROC

```
Setting levels: control = Ctl, case = Syn
```

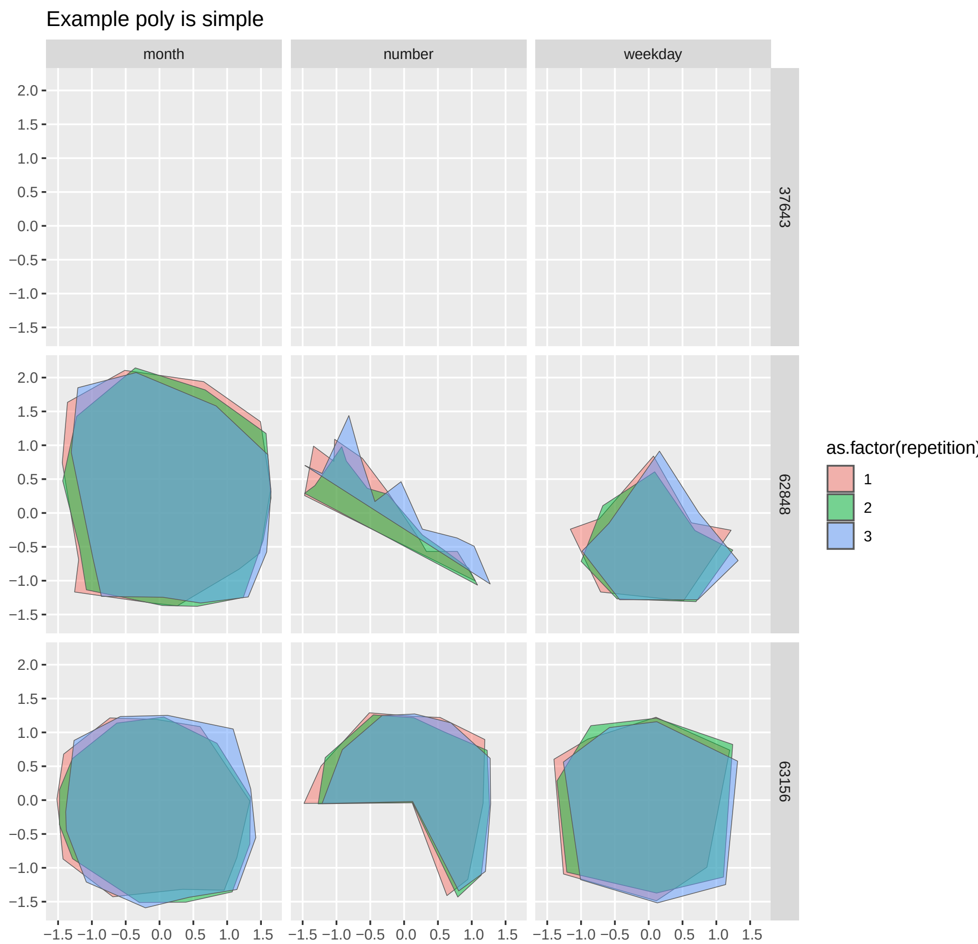


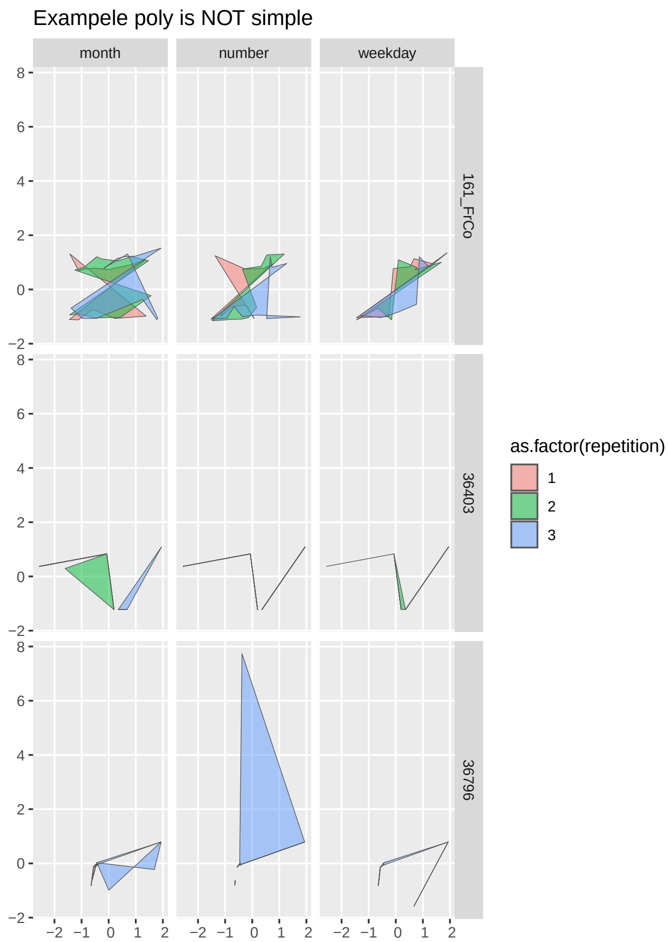

```
$ROC_properties
      Feature     AUC threshold sensitivity specificity      ppv      npv
1 GA_areaPoly 71.0339  1.288086    61.27946    71.06227 69.7318 62.78317
    ci_low  ci_high
1 66.84183 75.22598

$Coningency_table

      Ctl             Syn
  Ctl "194 (71.1%)" "79 (28.9%)"
  Syn "115 (38.7%)" "182 (61.3%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

```
  <fct> <int> <dbl> <dbl>
1 Ctl     273 0.958 0.989
2 Syn     297 1.77  1.24

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirh

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 71.03%
95% CI: 66.84%-75.23% (DeLong)


Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

12 Polygon simplicity

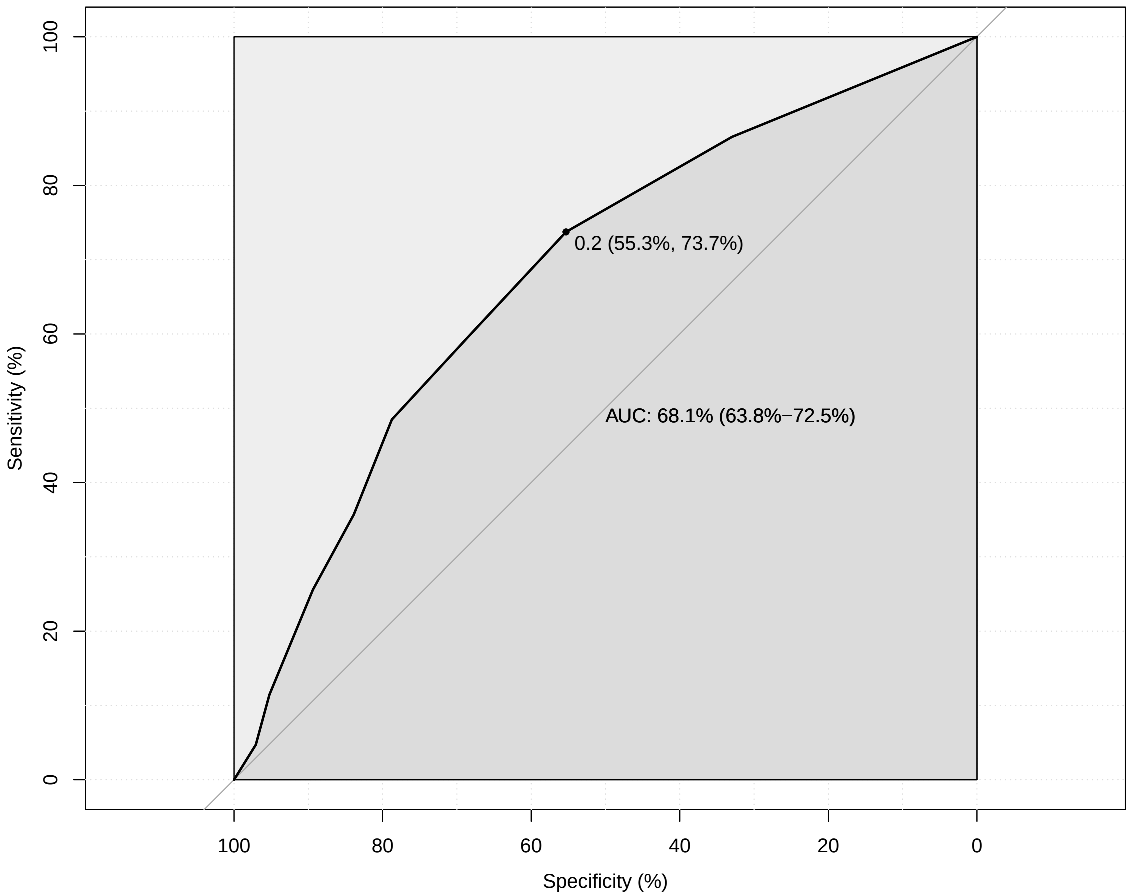
12.1 Example





12.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
      Feature     AUC threshold sensitivity specificity      ppv      npv
1 GA_isSimple 68.1306 0.1666667    73.73737    55.31136 64.22287 65.93886
    ci_low  ci_high
1 63.79175 72.46952

$Coningency_table

      Ctl           Syn
  Ctl "151 (55.3%)" "122 (44.7%)"
  Syn "78 (26.3%)"  "219 (73.7%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

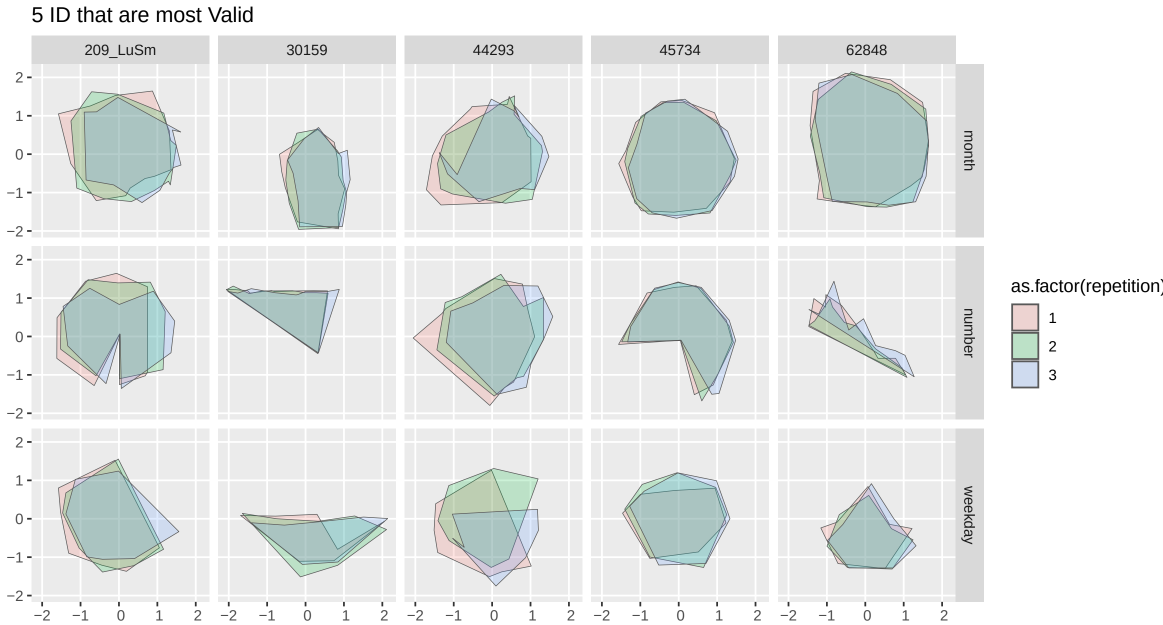
```
  <fct> <int> <dbl> <dbl>
1 Ctl     273 0.223 0.249
2 Syn     297 0.387 0.272

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 68.13%
95% CI: 63.79%-72.47% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

13 Topological validity Structure

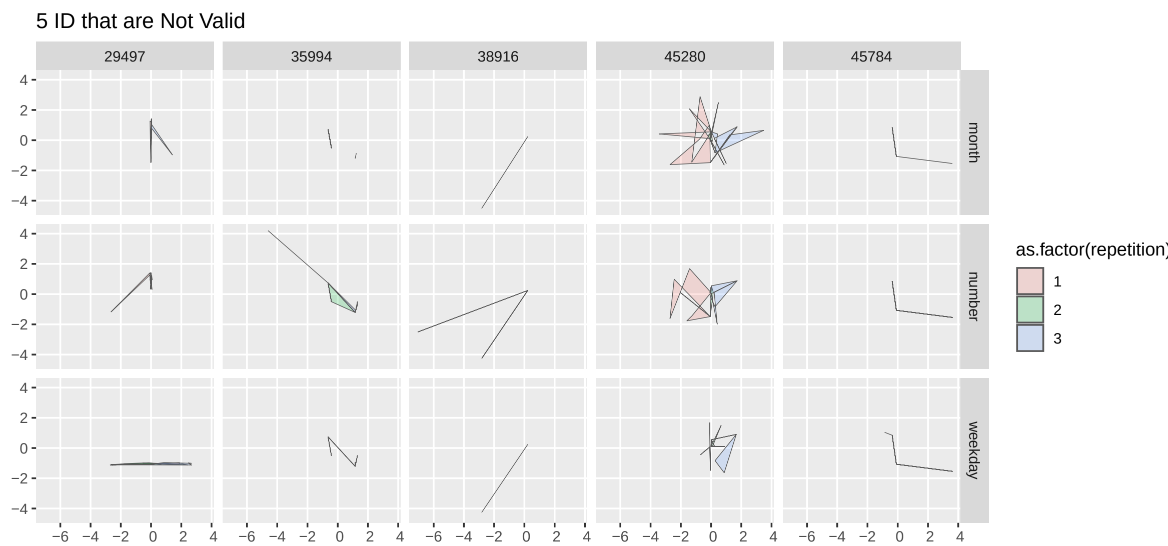
is topologically valid:

From the package description: *"For projected geometries, st_make_valid uses the lwgeom_makevalid method also used by the PostGIS command ST_makevalid if the GEOS version linked to is smaller than 3.8.0, and otherwise the version shipped in GEOS; for geometries having ellipsoidal coordinates s2::s2_rebuild is being used."* From https://postgis.net/docs/ST_IsValid.html: *value is well-formed and valid in 2D according to the OGC rules.* (Open Geopsatial Consotrtium)

13.1 Example

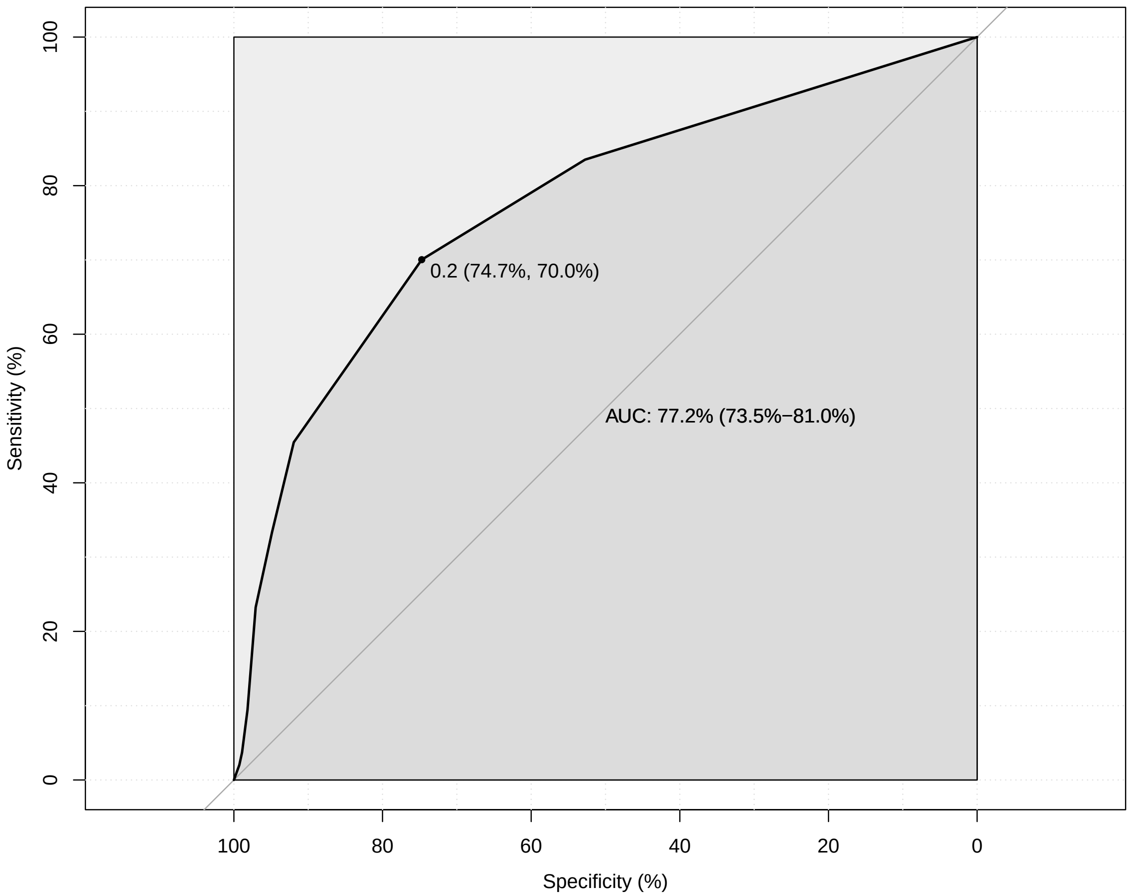


5 ID that are most Valid
209_LuSm
30159
44293
45734
62848
month
number
weekday
as.factor(repetition)
1
2
3



13.2 ROC

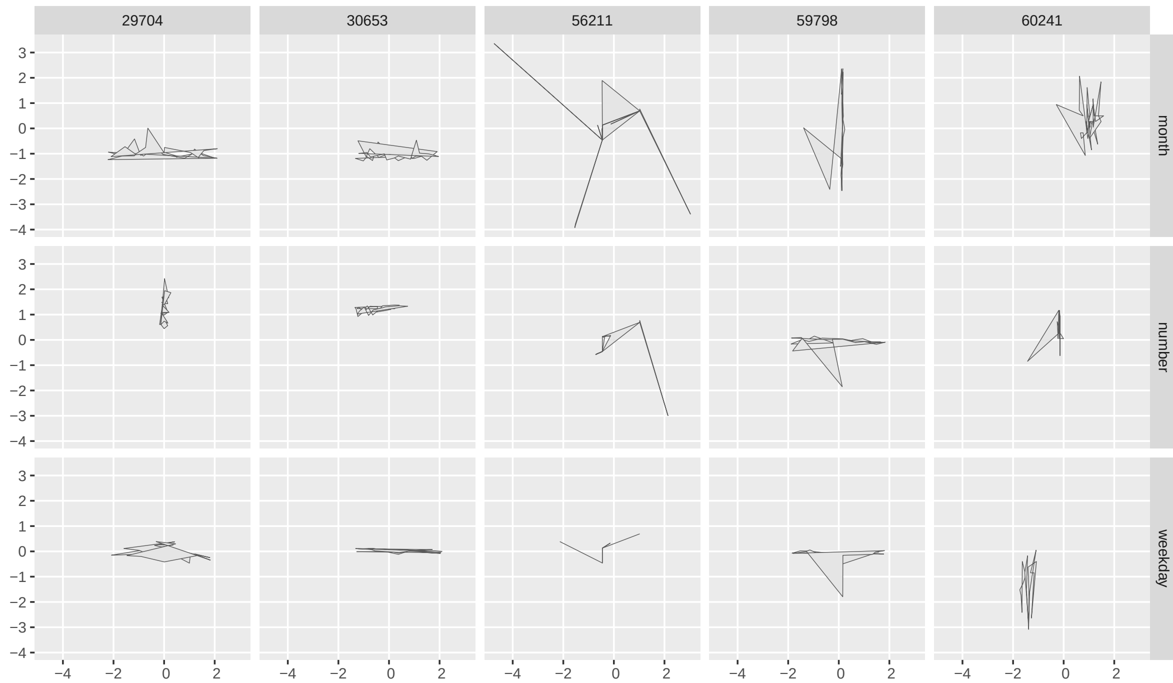
```
Setting levels: control = Ctl, case = Syn
```

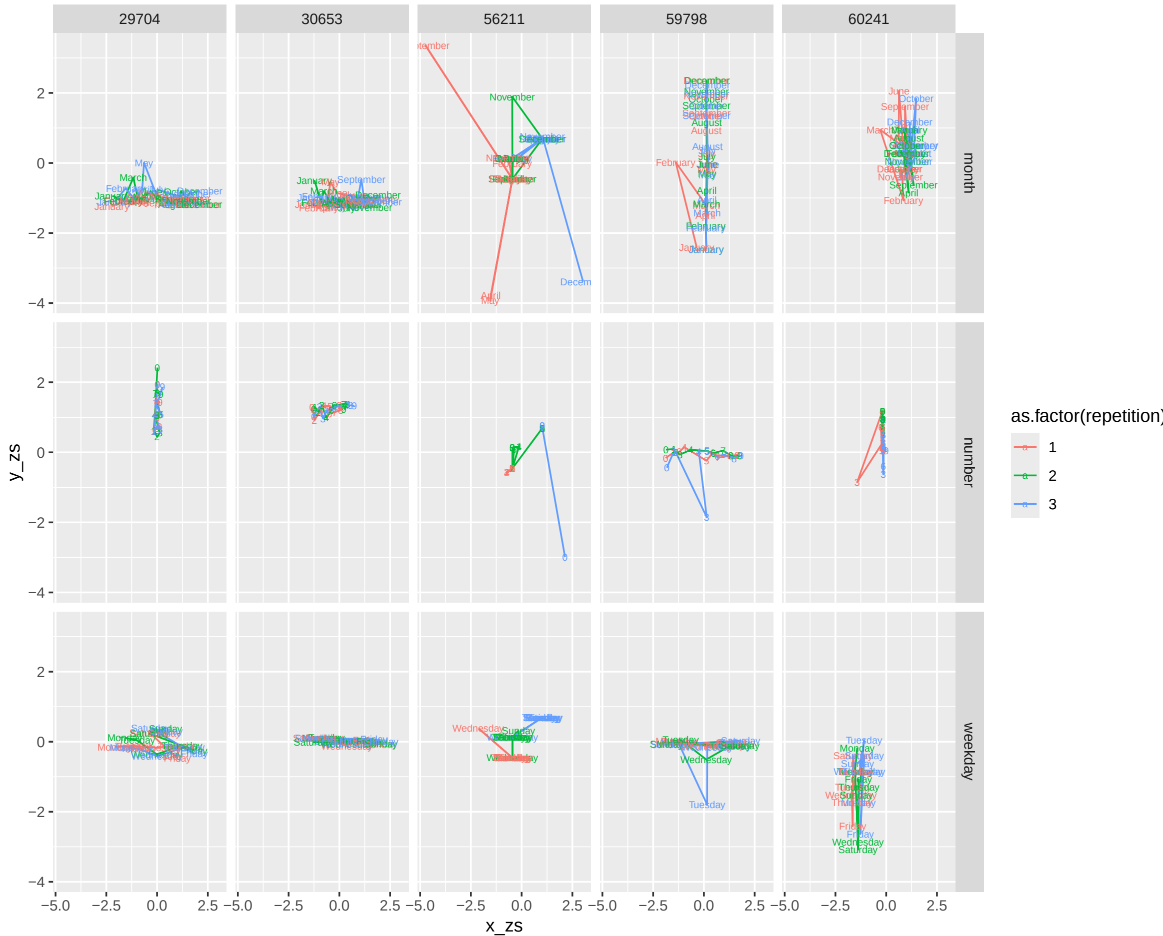


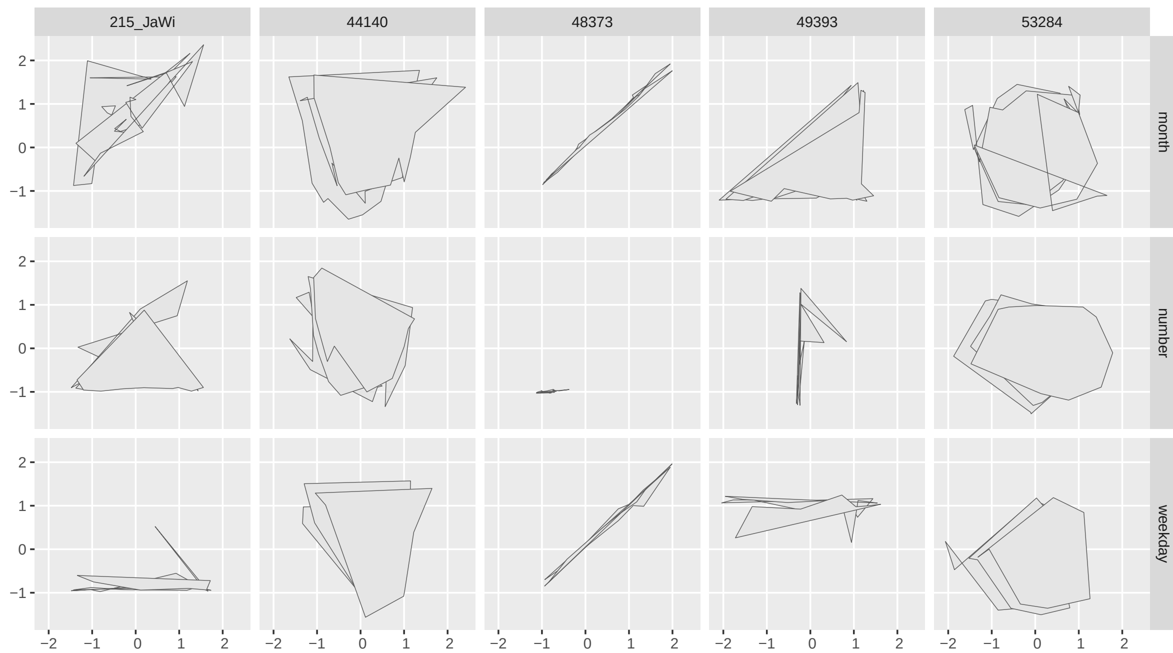

```
$ROC_properties
          Feature     AUC threshold sensitivity specificity      ppv      npv
1 GA_isValidStruct 77.2289 0.1666667    70.03367    74.72527 75.09025 69.62457
   ci_low  ci_high
1 73.47085 80.98703

$Coningency_table

      Ctl           Syn
  Ctl "204 (74.7%)" "69 (25.3%)"
  Syn "89 (30%)"    "208 (70%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

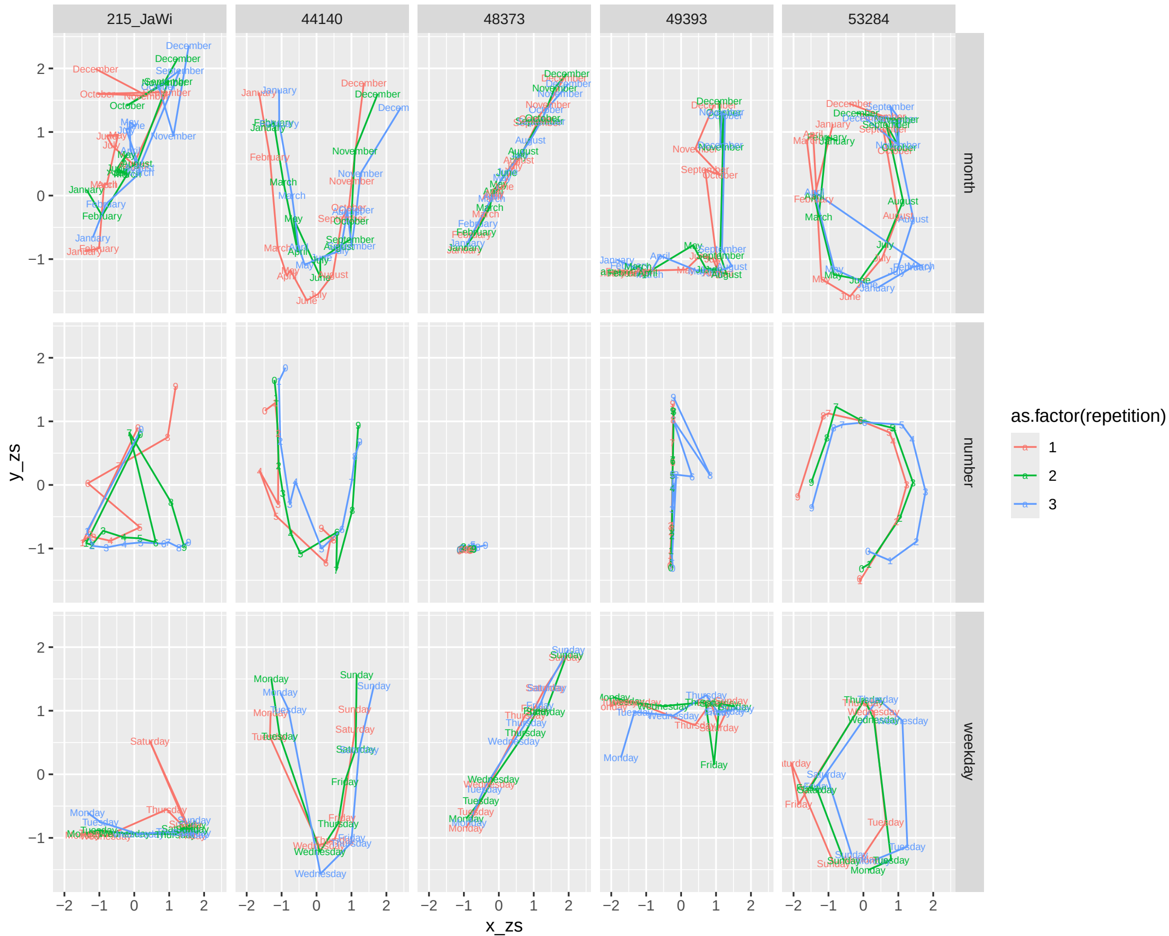
```
  <fct> <int> <dbl> <dbl>
1 Ctl     273 0.120 0.183
2 Syn     297 0.363 0.271

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 77.23%
95% CI: 73.47%-80.99% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

13.2.1 Extra: Visualize false pos and neg




29704
30653
56211
59798
60241
month
number
weekday
y_zs
x_zs
as.factor(repetition)
1
2
3


215_JaWi
44140
48373
49393
53284
month
number
weekday
2
1
0
−1
−2
−1
0
1
2



14 Topological DE-9IM

See: https://r-spatial.org/book/03-Geometries.html#sec-opgeom See: https://en.wikipedia.org/wiki/DE-9IM

DE-9IM is a standard for several topological model's features. It is called by st_relate. It returns a 3 x 3 matrix (DE9IM) for each relations:

$$\mathrm{DE9IM}(a,b) = \begin{bmatrix} \dim(I(a)\cap I(b)) & \dim(I(a)\cap B(b)) & \dim(I(a)\cap E(b)) \\ \dim(B(a)\cap I(b)) & \dim(B(a)\cap B(b)) & \dim(B(a)\cap E(b)) \\ \dim(E(a)\cap I(b)) & \dim(E(a)\cap B(b)) & \dim(E(a)\cap E(b)) \end{bmatrix}$$

dim is the dimension of the intersection ( ) of the interior (I), boundary (B), and exterior (E) of geometries a and b.

Hence it returns a *spatial predicate* wdefined with mas domains:

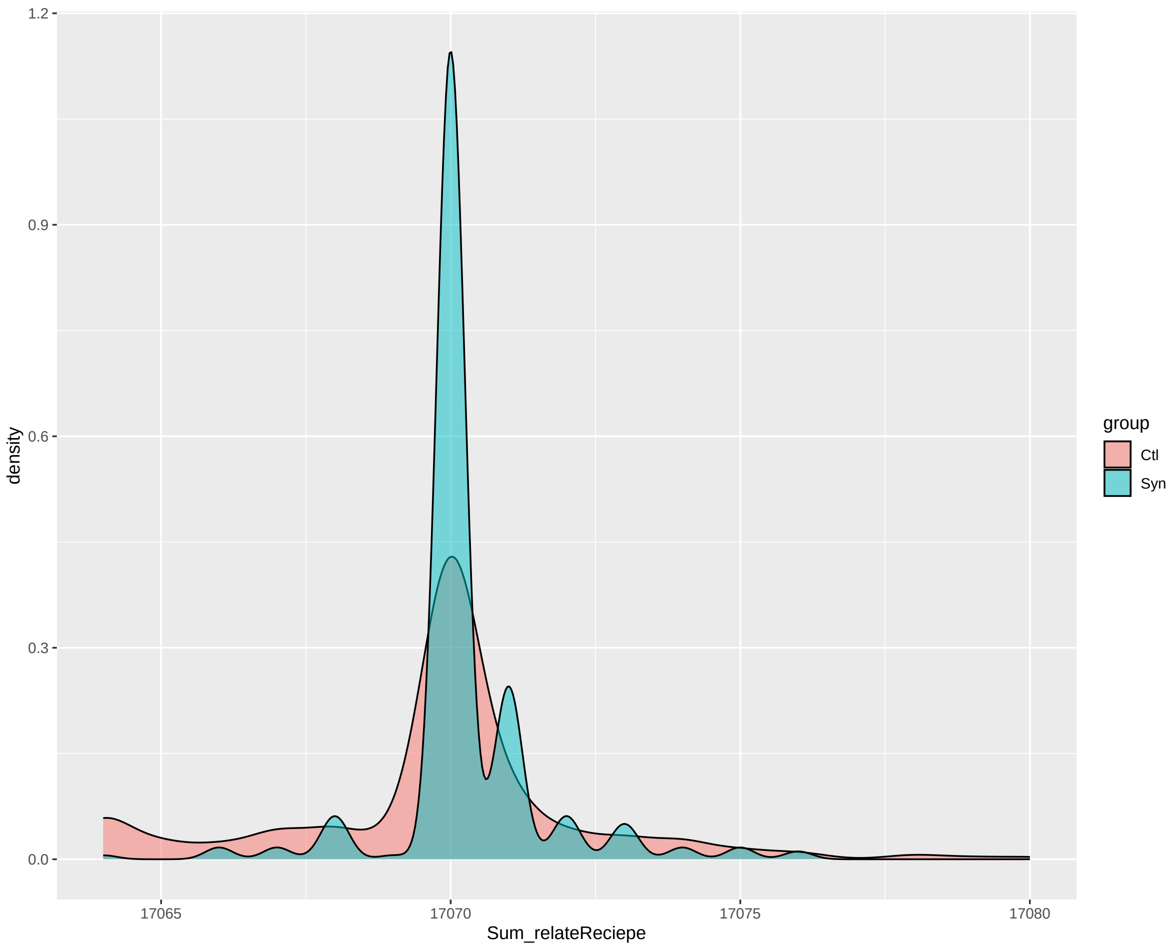
	Ctl	Syn	Subs
212101212	954	1567	0.6088066
2121012F2	2	2	1.0000000
21210F212	0	1	0.0000000
212111212	4	8	0.5000000
212F01212	18	8	2.2500000
212F01FF2	15	2	7.5000000
212F0F212	12	2	6.0000000
212F0FFF2	5	1	5.0000000
212F1FFF2	1	0	Inf
2F2101212	12	1	12.0000000
2F21012F2	10	3	3.3333333
2F2F01212	16	2	8.0000000
2F2F01FF2	41	5	8.2000000
2FF10F212	12	1	12.0000000
2FF10F2F2	3	0	Inf
2FF11F212	1	0	Inf
2FFF0F212	32	3	10.6666667
2FFF0FFF2	138	19	7.2631579
2FFF1FFF2	626	853	0.7338804
FF2F01212	7	0	Inf
FF2F01FF2	27	6	4.5000000
FF2F11212	1	0	Inf
FF2FF1212	292	125	2.3360000
FFFF0F212	26	2	13.0000000
FFFF0FFF2	202	62	3.2580645

	Ctl	Syn	Subs
212101212	1010	1627	0.6207744
2121012F2	0	2	0.0000000
21210F212	2	3	0.6666667
21210F2F2	1	0	Inf
212111212	12	9	1.3333333
212F01212	11	2	5.5000000
212F01FF2	11	0	Inf
212F0F212	12	1	12.0000000
212F0FFF2	5	1	5.0000000
212F11212	1	0	Inf
212F11FF2	1	0	Inf
2F2101212	13	7	1.8571429
2F21012F2	27	6	4.5000000
2F2111212	1	0	Inf

2F2F01212	20	1	20.0000000
2F2F01FF2	44	10	4.4000000
2FF10F212	21	2	10.5000000
2FF10F2F2	8	1	8.0000000
2FF11F2F2	1	0	Inf
2FFF0F212	59	7	8.4285714
2FFF0FFF2	173	31	5.5806452
2FFF1FFF2	591	846	0.6985816
FF2F01212	5	2	2.5000000
FF2F01FF2	40	4	10.0000000
FF2F11212	3	0	Inf
FF2FF1212	98	40	2.4500000
FFFF0F212	31	8	3.8750000
FFFF0FFF2	256	63	4.0634921

	Ctl	Syn	Subs
212101212	942	1568	0.6007653
2121012F2	0	2	0.0000000
21210F212	0	2	0.0000000
21210F2F2	1	0	Inf
212111212	8	7	1.1428571
212F01212	12	3	4.0000000
212F01FF2	14	1	14.0000000
212F0F212	23	7	3.2857143
212F0FFF2	5	0	Inf
212F11212	1	0	Inf
2F2101212	18	5	3.6000000
2F21012F2	11	1	11.0000000
2F2111212	1	0	Inf
2F2F01212	12	1	12.0000000
2F2F01FF2	44	3	14.6666667
2FF10F212	7	0	Inf
2FF10F2F2	4	1	4.0000000
2FFF0F212	39	8	4.8750000
2FFF0FFF2	147	27	5.4444444
2FFF1FFF2	611	853	0.7162954
FF2F01212	6	2	3.0000000
FF2F01FF2	30	4	7.5000000
FF2F11212	2	0	Inf
FF2FF1212	276	125	2.2080000
FFFF0F212	40	4	10.0000000
FFFF0FFF2	203	49	4.1428571

	Ctl	Syn	Subs
2121012122	942	1568	0.6007653
212101212	954	1567	0.6088066
2121012121	1010	1627	0.6207744
2FFF1FFF21	591	846	0.6985816
2FFF1FFF22	611	853	0.7162954
2FFF1FFF2	626	853	0.7338804
FF2FF12122	276	125	2.2080000
FF2FF1212	292	125	2.3360000



2FFF1FFF2: S1 Interior vs. S2 Interior: The interiors intersect in 2 dimensions (2). S1 Interior vs. S2 Boundary: No intersection (F). S1 Interior vs. S2 Exterior: No intersection (F). S1 Boundary vs. S2 Interior: No intersection (F). S1 Boundary vs. S2 Boundary: A 1-dimensional intersection occurs (e.g., they share a common line segment) (1). S1 Boundary vs. S2 Exterior: No intersection (F). S1 Exterior vs. S2 Interior: No intersection (F). S1

Exterior vs. S2 Boundary: No intersection (F). S1 Exterior vs. S2 Exterior: The exteriors intersect in 2 dimensions (2).

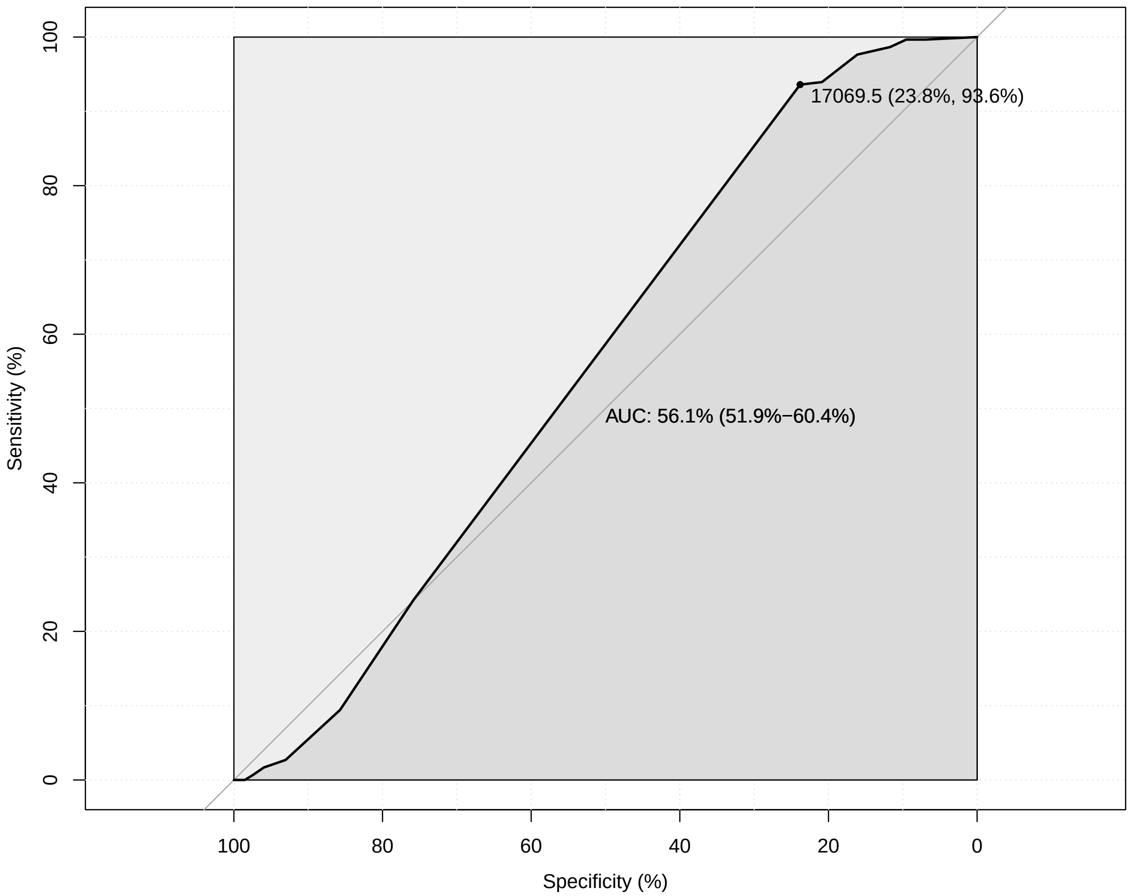
2FFF0FFF2: 2: The intersection of the first geometry's interior and the second geometry's interior creates a polygon (a two-dimensional intersection). F: The interior of the first geometry does not intersect the boundary of the second. F: The interior of the first geometry does not intersect the exterior of the second. F: The boundary of the first geometry does not intersect the interior of the second. 0: The boundary of the first geometry intersects the boundary of the second geometry at a point (a zero-dimensional intersection). F: The boundary of the first geometry does not intersect the exterior of the second. F: The exterior of the first geometry does not intersect the interior of the second. F: The exterior of the first geometry does not intersect the boundary of the second. 2: The exterior of the first geometry intersects the exterior of the second geometry, creating a polygon (a two-dimensional intersection).

FFFF0FFF2: F (False): The intersection of the interior of the first geometry with the interior of the second geometry is empty. F (False): The intersection of the interior of the first geometry with the boundary of the second geometry is empty. F (False): The intersection of the interior of the first geometry with the exterior of the second geometry is empty. F (False): The intersection of the boundary of the first geometry with the interior of the second geometry is empty. 0 (Zero-Dimensional): The intersection of the boundary of the first geometry with the boundary of the second geometry is a point (0-dimensional). F (False): The intersection of the boundary of the first geometry with the exterior of the second geometry is empty. F (False): The intersection of the exterior of the first geometry with the interior of the second geometry is empty. F (False): The intersection of the exterior of the first geometry with the boundary of the second geometry is empty. 2 (Two-Dimensional): The intersection of the exterior of the first geometry with the exterior of the second geometry is a 2-dimensional area.

14.1 Example

14.2 ROC

```
Setting levels: control = Ctl, case = Syn
```

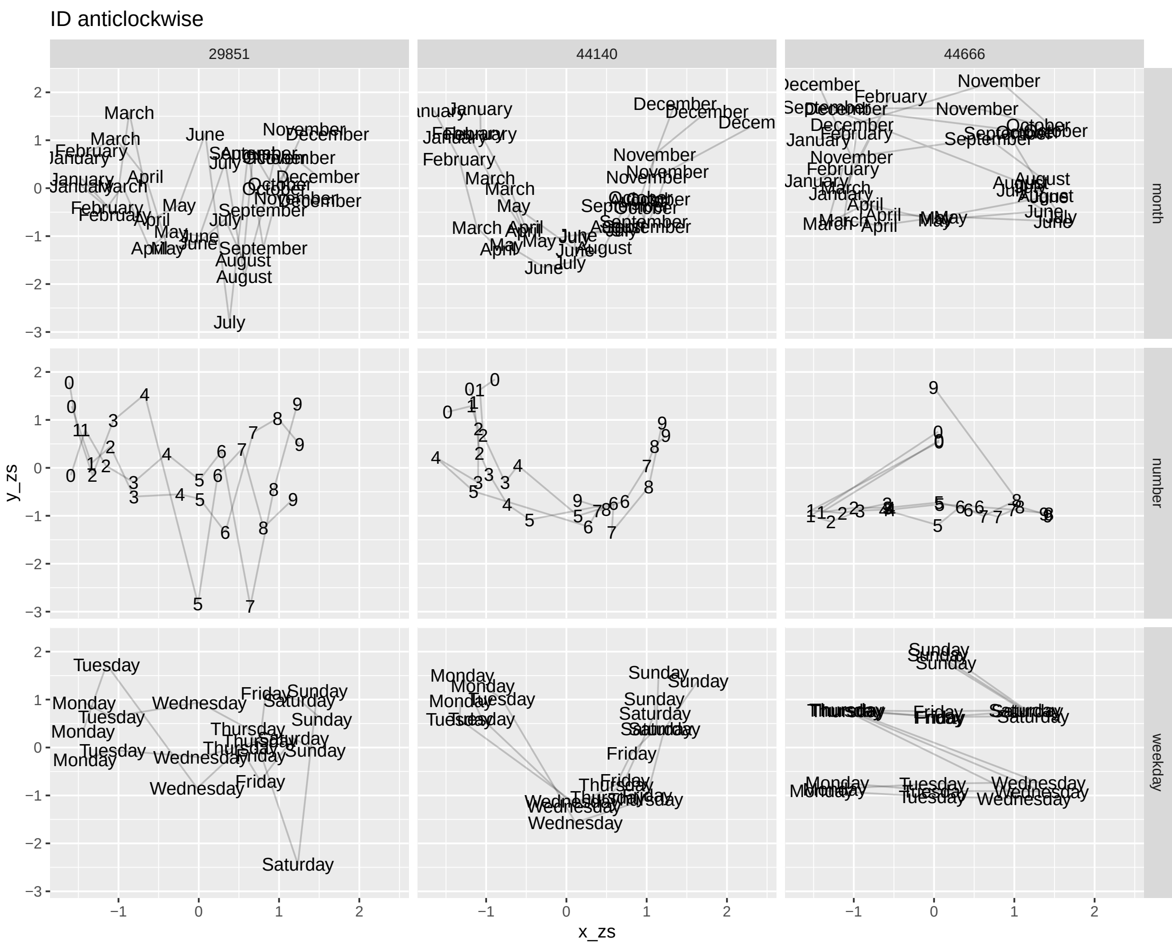


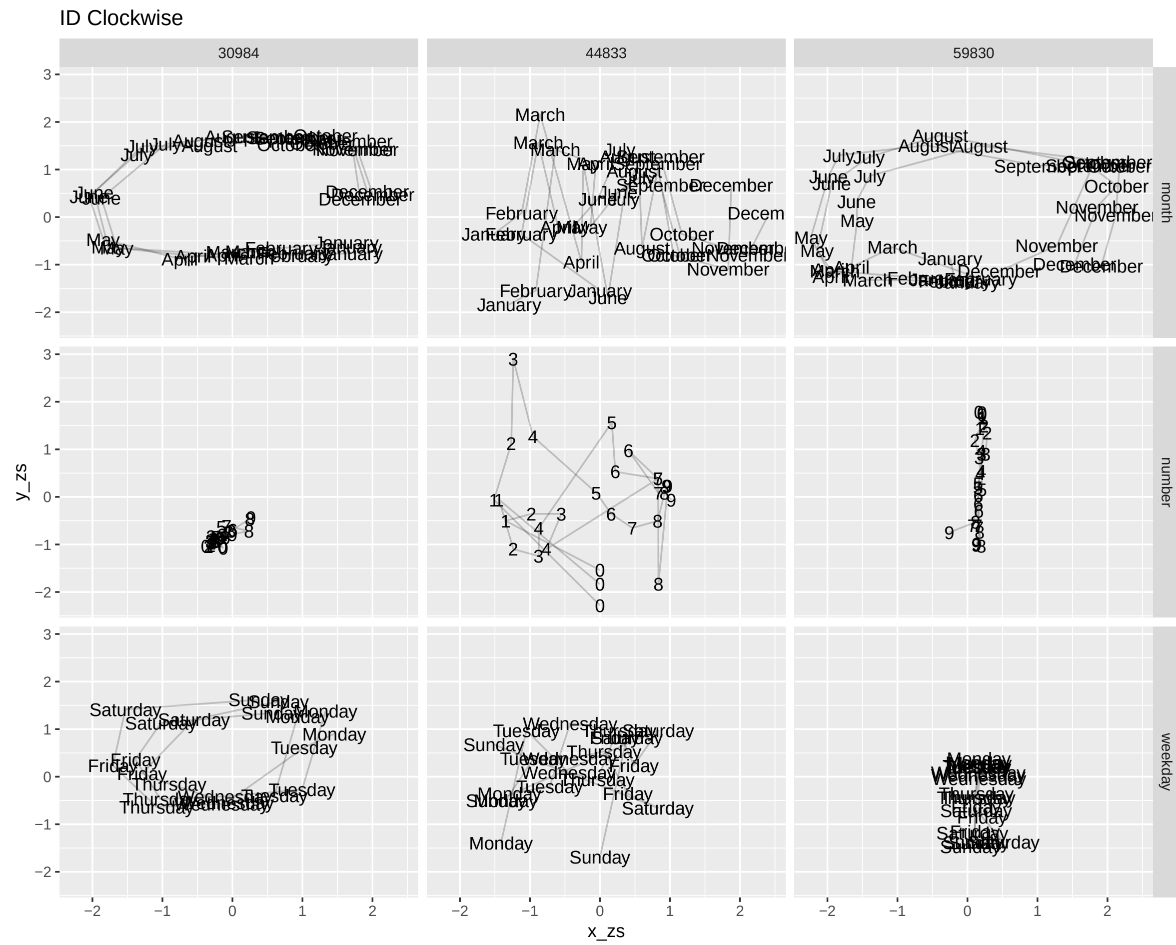

```
$ROC_properties
            Feature     AUC threshold sensitivity specificity      ppv      npv
1 Sum_relateReciepe 56.1488   17069.5    93.60269    23.80952 57.20165 77.38095
    ci_low  ci_high
1 51.89826 60.39932

$Coningency_table

      Ctl           Syn
  Ctl "65 (23.8%)" "208 (76.2%)"
  Syn "19 (6.4%)"  "278 (93.6%)"

$Descr_table
# A tibble: 2 x 4
  group     n   Mean    SD
```

```
  <fct> <int>  <dbl> <dbl>
1 Ctl     273 17070.  2.65
2 Syn     297 17070.  1.30

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 56.15%
95% CI: 51.9%-60.4% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

15 is clockwise (?)

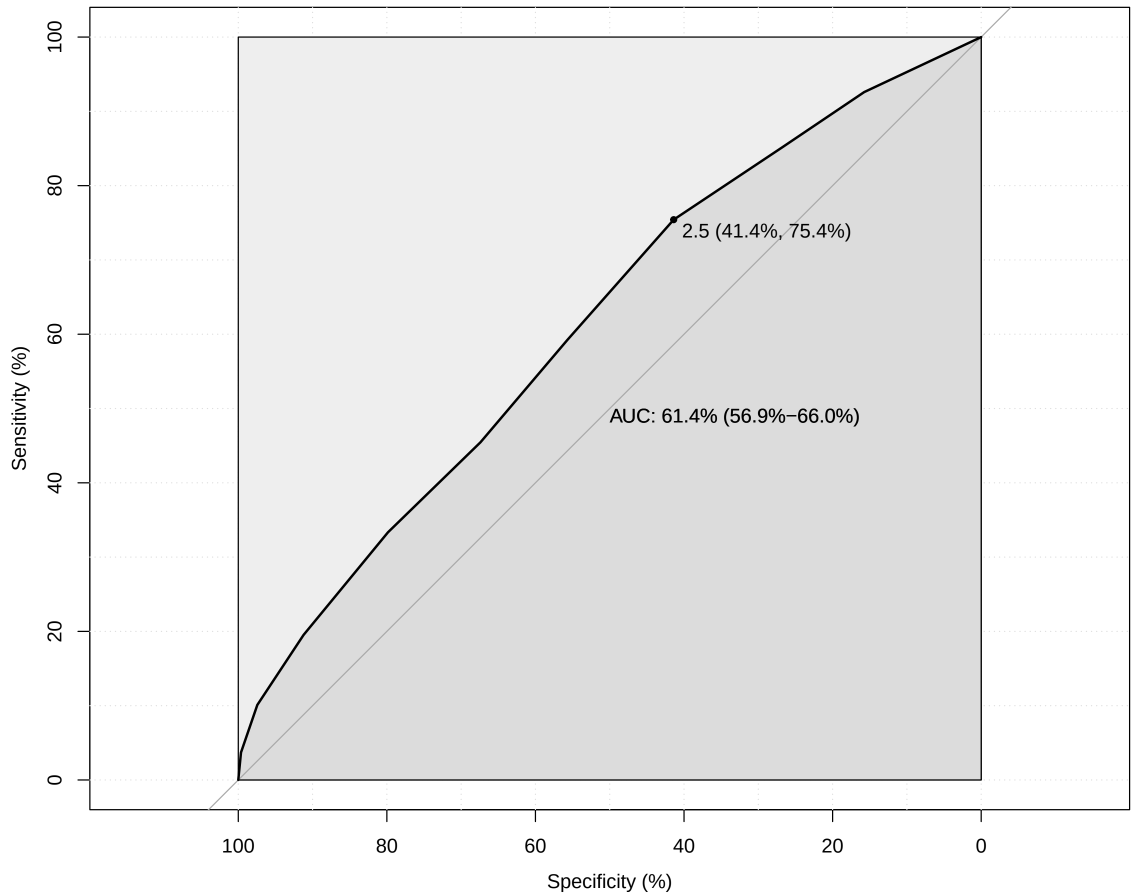
15.1 Example





15.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
          Feature     AUC threshold sensitivity specificity      ppv      npv
1 Sum_isClockwise 61.4472       2.5    75.42088    41.39194 58.33333 60.75269
    ci_low  ci_high
1 56.89624 65.99815

$Coningency_table

      Ctl            Syn
  Ctl "113 (41.4%)" "160 (58.6%)"
  Syn "73 (24.6%)"  "224 (75.4%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

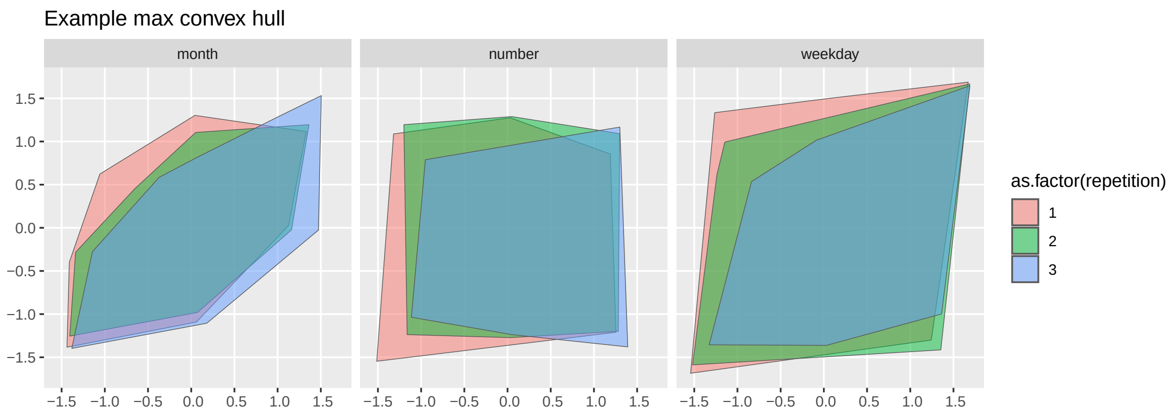
```
  <fct> <int> <dbl> <dbl>
1 Ctl     273  3.25  2.30
2 Syn     297  4.25  2.42

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 61.45%
95% CI: 56.9%-66% (DeLong)

Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

16 Poly to circle

17 Convex hull

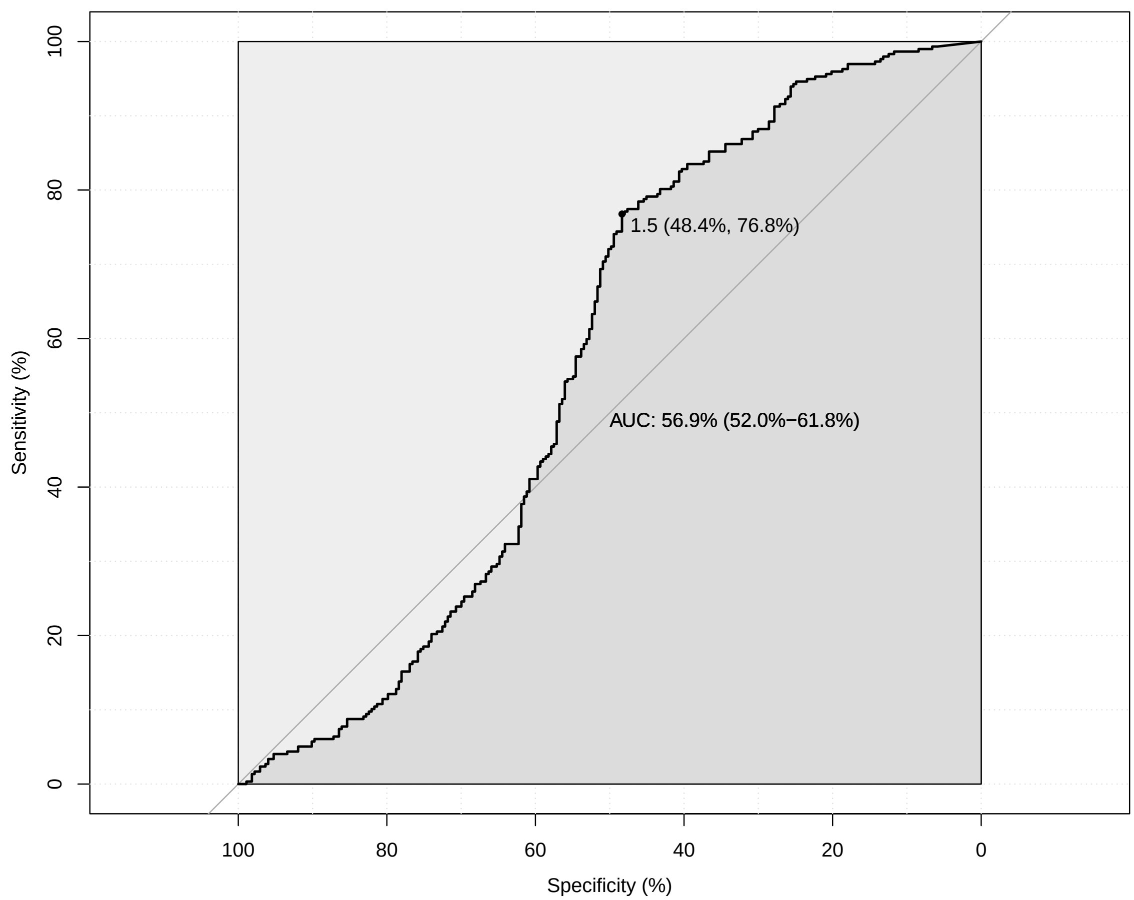
18 Convex Hull Area

18.1 Example



18.2 ROC

```
Setting levels: control = Ctl, case = Syn
```




```
$ROC_properties
           Feature     AUC threshold sensitivity specificity      ppv      npv
1 GA_area_convhull 56.8777  1.491143    76.76768    48.35165 61.78862 65.67164
   ci_low  ci_high
1 51.95812 61.79726

$Coningency_table

      Ctl            Syn
  Ctl "132 (48.4%)" "141 (51.6%)"
  Syn "69 (23.2%)"  "228 (76.8%)"

$Descr_table
# A tibble: 2 x 4
  group     n  Mean    SD
```

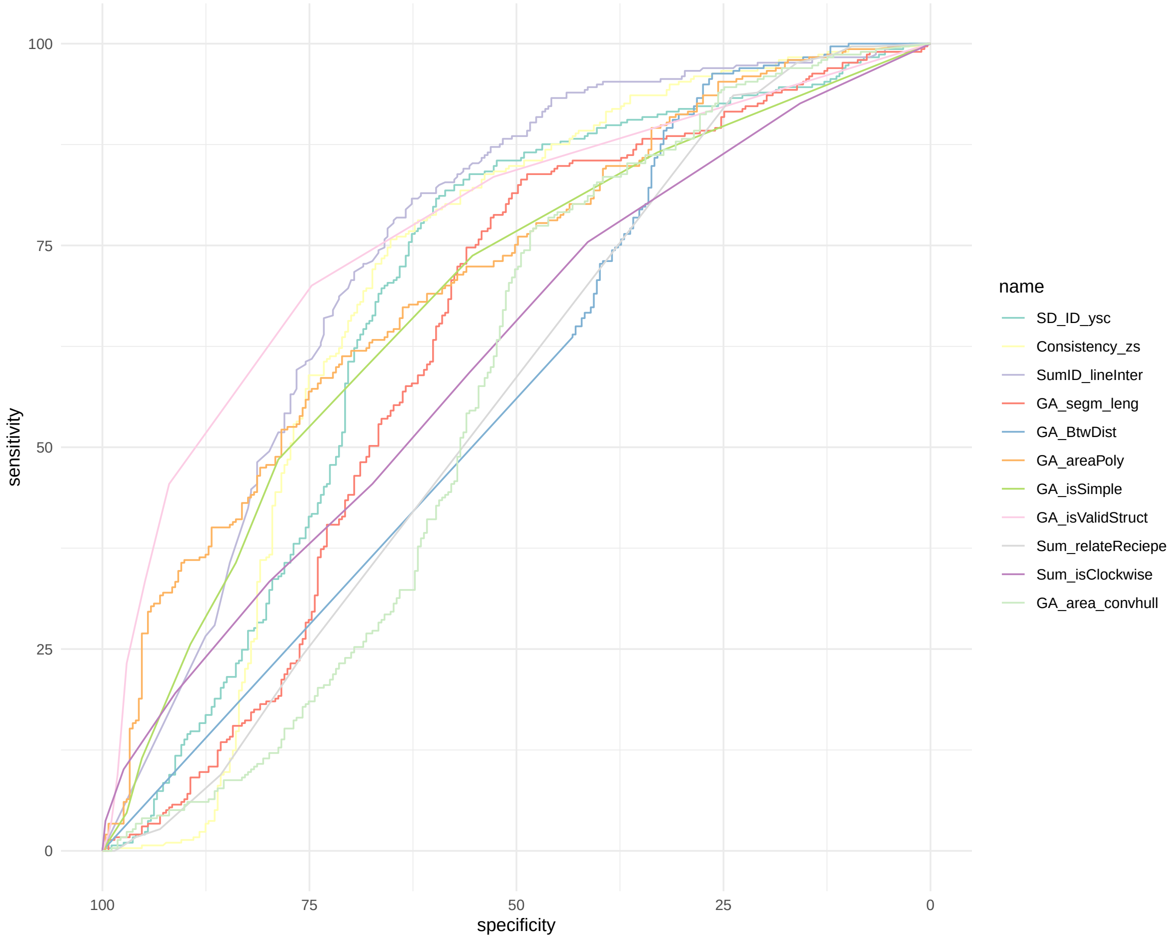
```
  <fct> <int> <dbl> <dbl>
1 Ctl     273  2.29  1.88
2 Syn     297  2.62  1.42

$ROC_curve

Call:
roc.formula(formula = data[[group_col]] ~ data[[feature]], data = data,     direction = Dirhe

Data: data[[feature]] in 273 controls (data[[group_col]] Ctl) < 297 cases (data[[group_col]]
Area under the curve: 56.88%
95% CI: 51.96%-61.8% (DeLong)


Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
Setting levels: control = Ctl, case = Syn
```

19 Compare all features:

19.1 Summary table:

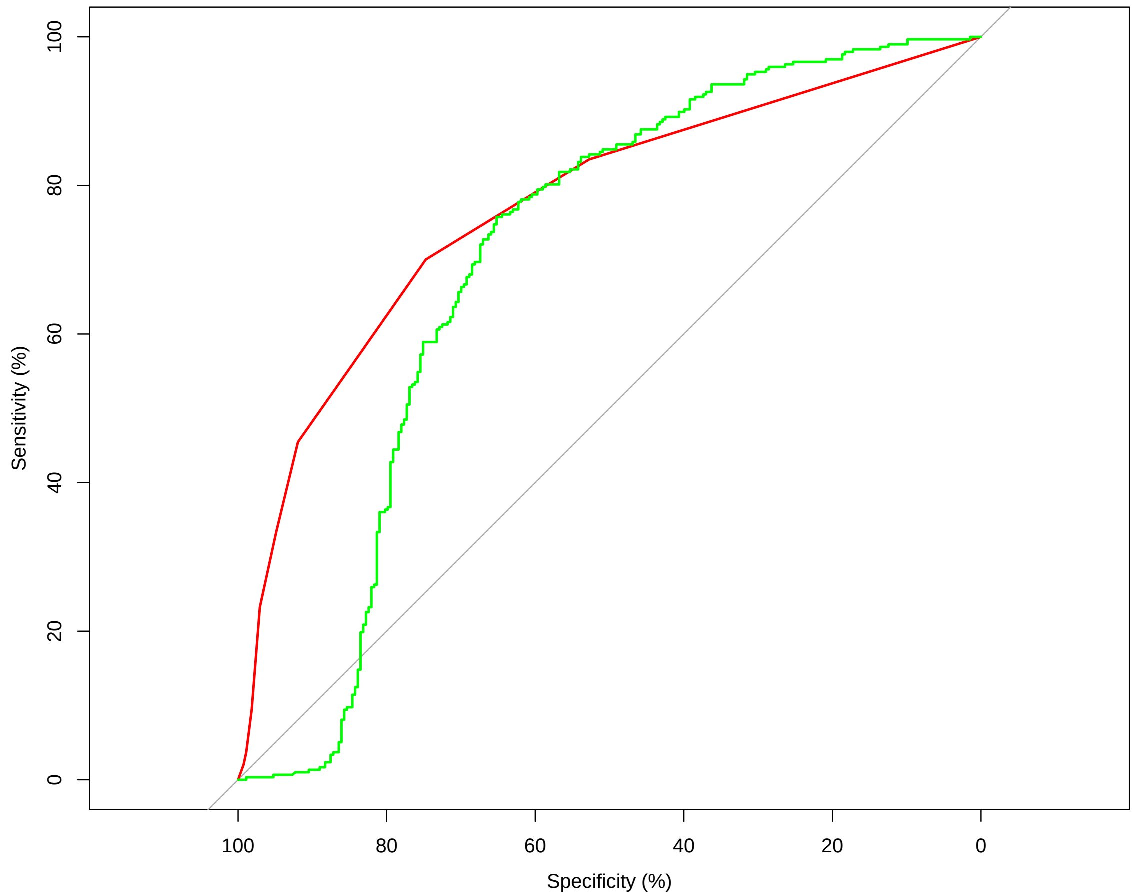
	Feature	AUC	threshold	sensitivity	specificity	ppv	npv	ci_low	ci_high
10	GA_isValidStruct	77.23	0.17	70.03	74.73	75.09	69.62	73.47	80.99
5	SumID_lineInter	74.98	110.50	80.81	62.64	70.18	75.00	70.88	79.08
8	GA_areaPoly	71.03	1.29	61.28	71.06	69.73	62.78	66.84	75.23
4	Consistency_zs	69.98	2.27	75.76	65.20	70.31	71.20	65.40	74.56
2	SD_ID_ysc	68.67	0.12	80.81	59.71	68.57	74.09	64.13	73.21
9	GA_isSimple	68.13	0.17	73.74	55.31	64.22	65.94	63.79	72.47
6	GA_segm_leng	63.10	7.76	83.16	49.45	64.16	72.97	58.37	67.84
1	SD_ID_xsc	61.68	0.12	88.55	44.69	63.53	78.21	56.86	66.50
12	Sum_isClockwise	61.45	2.50	75.42	41.39	58.33	60.75	56.90	66.00
7	GA_BtwDist	57.67	0.03	96.30	26.37	58.73	86.75	53.45	61.88
13	GA_area_convhull	56.88	1.49	76.77	48.35	61.79	65.67	51.96	61.80
11	Sum_relateReciepe	56.15	17069.50	93.60	23.81	57.20	77.38	51.90	60.40
3	Consistency	48.71	0.06	78.23	38.43	57.91	61.97	43.20	54.23

19.2 Summary plot



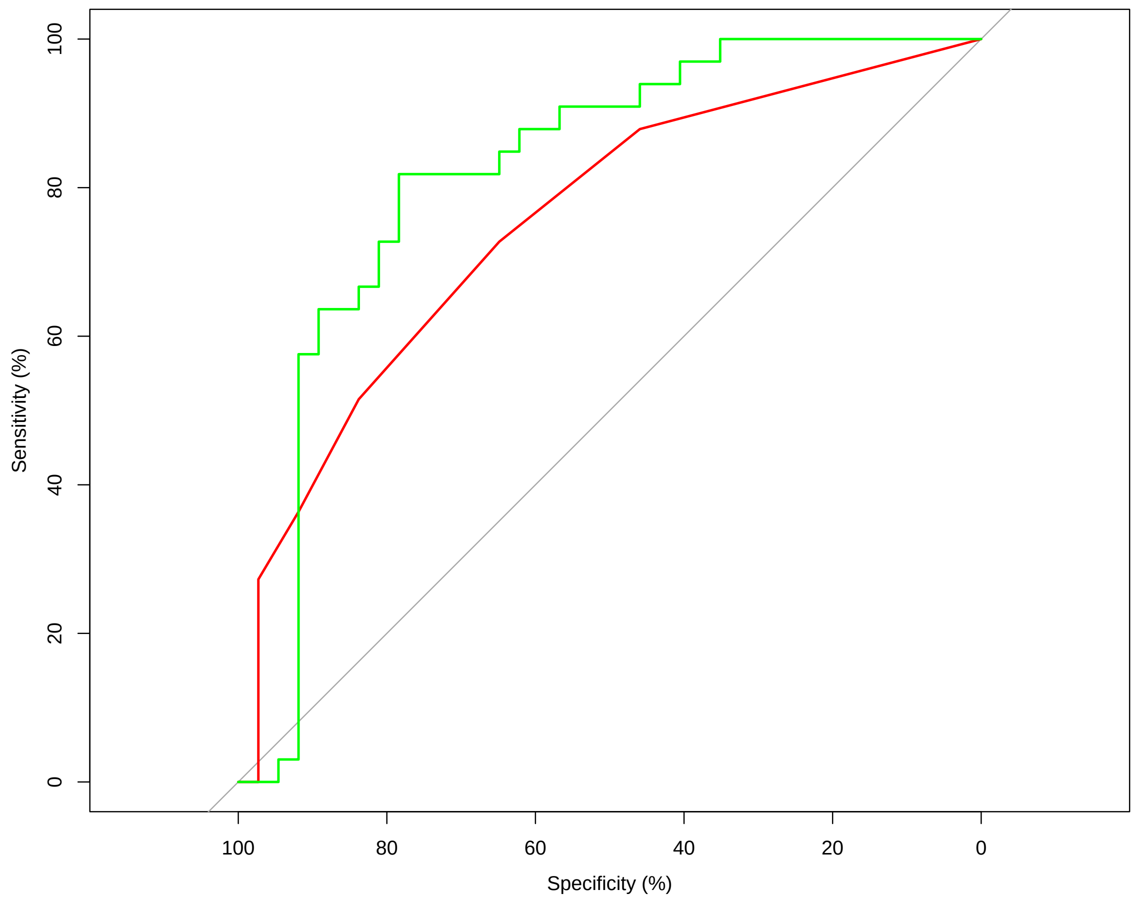
20 Could compare each feature singularly:

```
Setting levels: control = Ctl, case = Syn

Setting direction: controls < cases

Setting levels: control = Ctl, case = Syn

Setting direction: controls > cases
```




```
    Bootstrap test for two correlated ROC curves

data:  ROC_Valid and ROC_Cons
D = 2.6475, boot.n = 2000, boot.stratified = 1, p-value = 0.008108
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
   77.22894    69.98125

Setting levels: control = Ctl, case = Syn

Setting direction: controls < cases

Setting levels: control = Ctl, case = Syn
```

```
Setting direction: controls > cases
```




```
    Bootstrap test for two correlated ROC curves

data:  ROC_Valid and ROC_Cons
D = -0.96881, boot.n = 2000, boot.stratified = 1, p-value = 0.3326
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
AUC of roc1 AUC of roc2
   75.38903    82.22768
```

20.1 *Phase II* Methods

20.2 *Phase II* Materials:

Materials are desribed here https://osf.io/pjb6e/?view_only=d467ebf4c1f94076ae4ac61298255065.

20.3 *Phase II* Planned population

https://osf.io/6h8dx

21 Discussion

From the different features we extracted, topological validity across the repetitions appeared to be the one leading to the largest Area Under the Curve. The optimal cutoff was exactly 1.5, leading to a sensitivity () and specificity ().

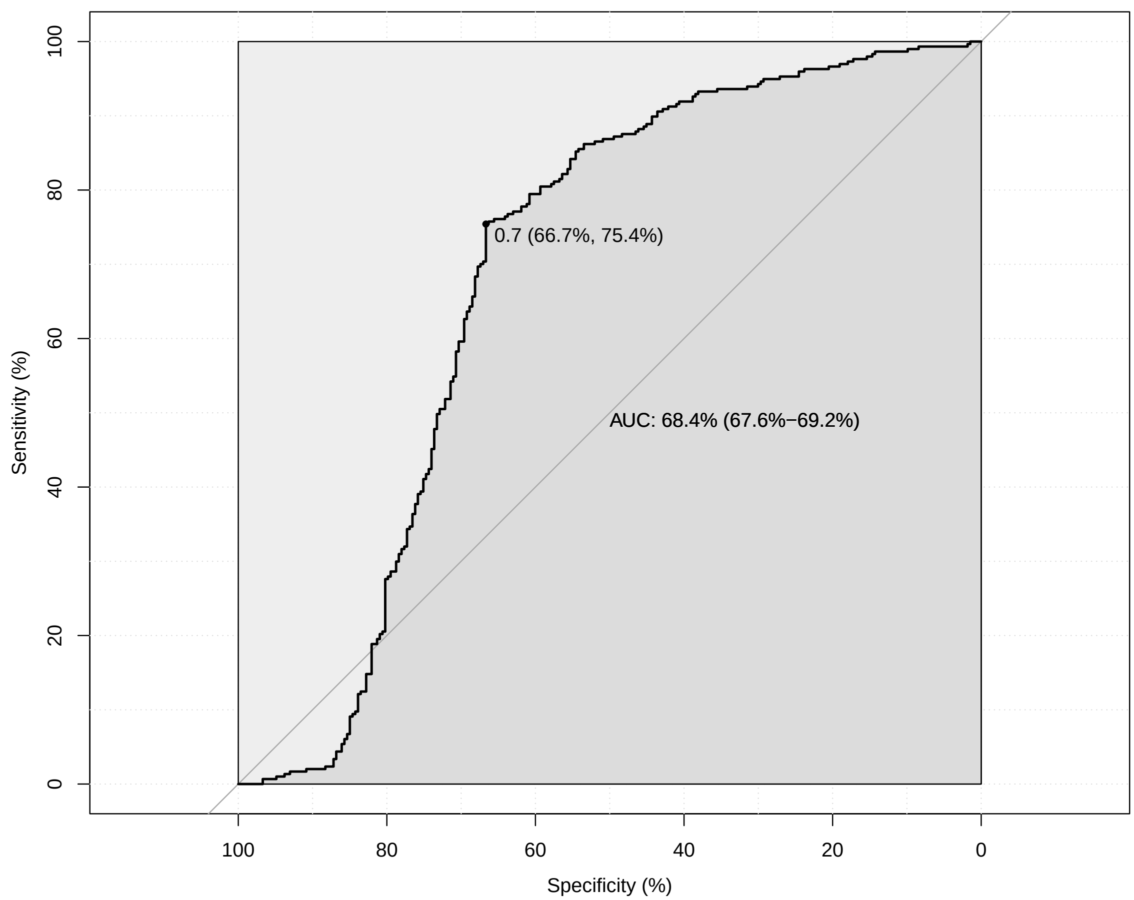
The optimal criterion ineeds to be informed about the order between inducers (i.e. to construct the polygons) and interestingly suggests that synthetic inducer are structurally mapped following topological rules analogous to geographical space structures. Hence suggesting a spatial nature for the synthetic forms of space sequence synesthetes.

22 Supplementary: Conditions

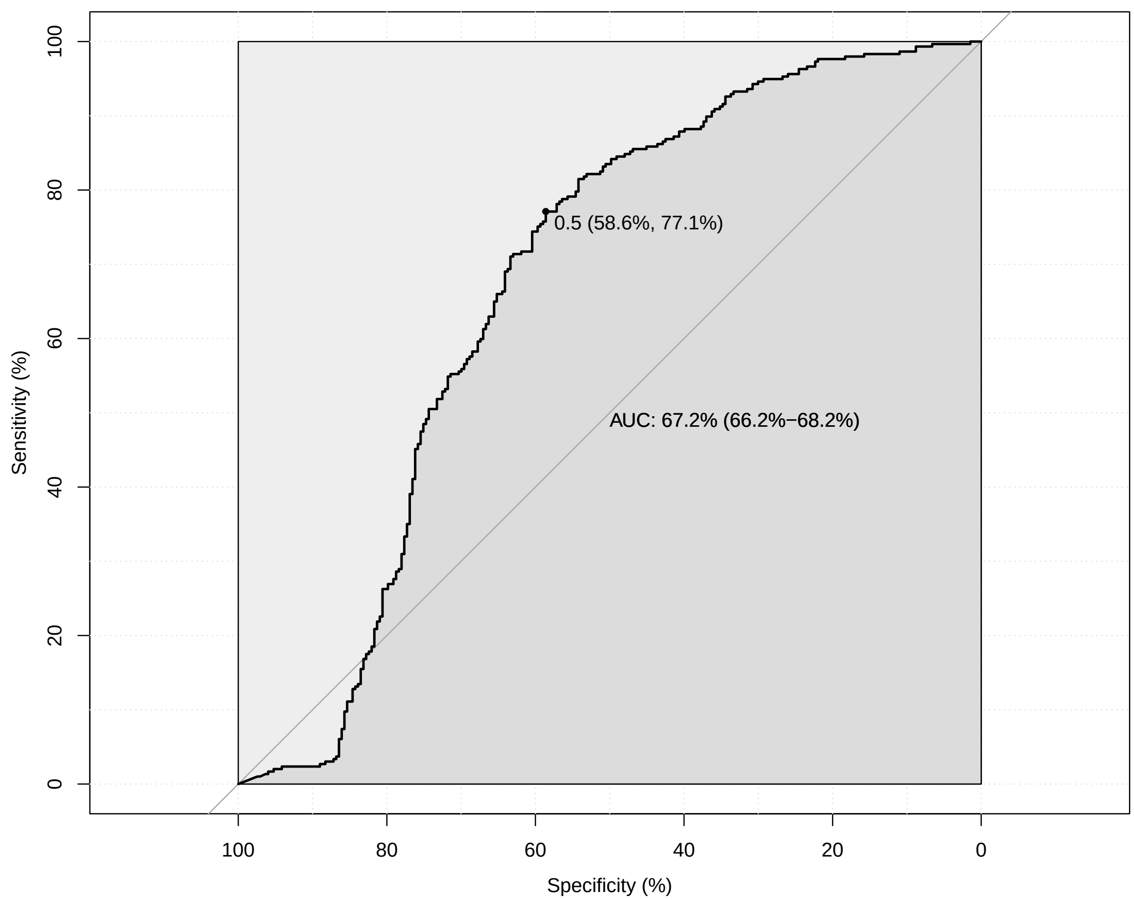
22.1 Per Conditions

```
Warning in rm(ROC_curvesCond): object 'ROC_curvesCond' not found
```

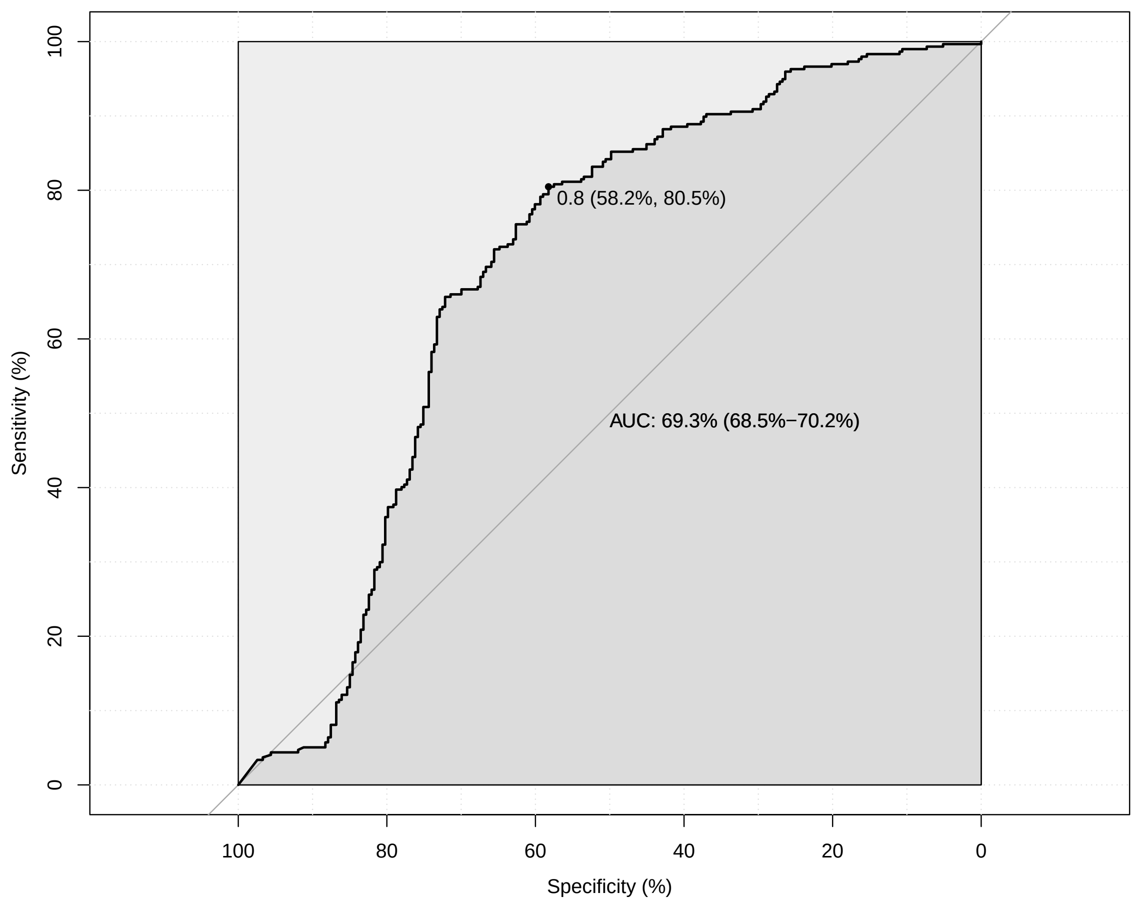
```
Setting levels: control = Ctl, case = Syn
```



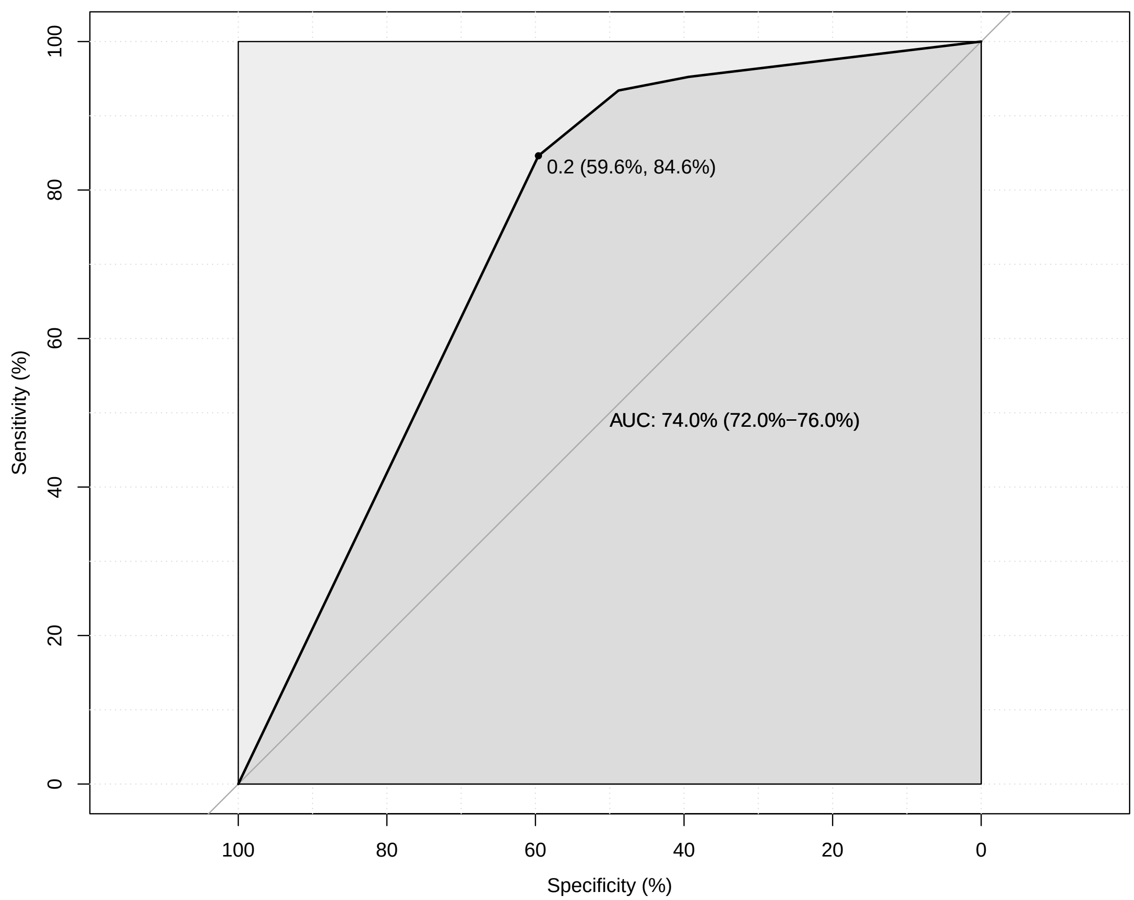

```
Setting levels: control = Ctl, case = Syn
```



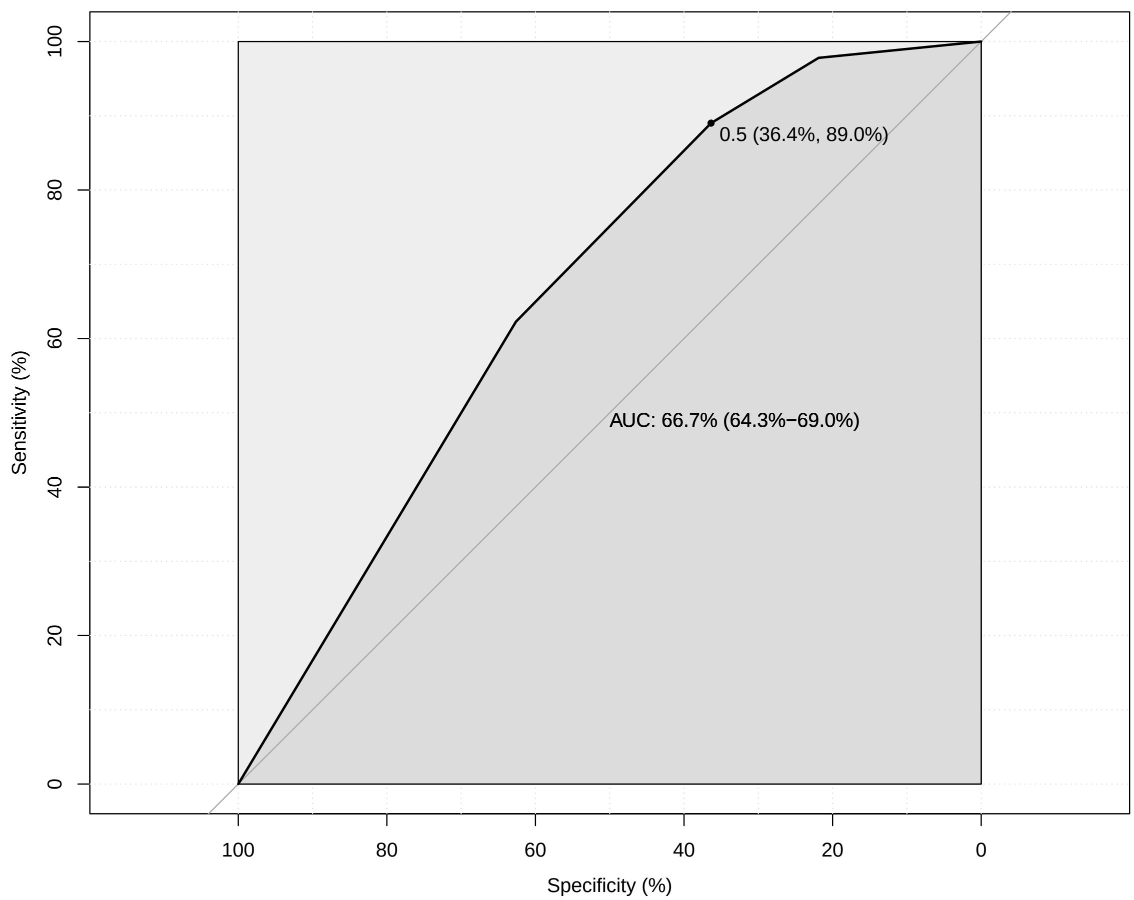

```
Setting levels: control = Ctl, case = Syn
```



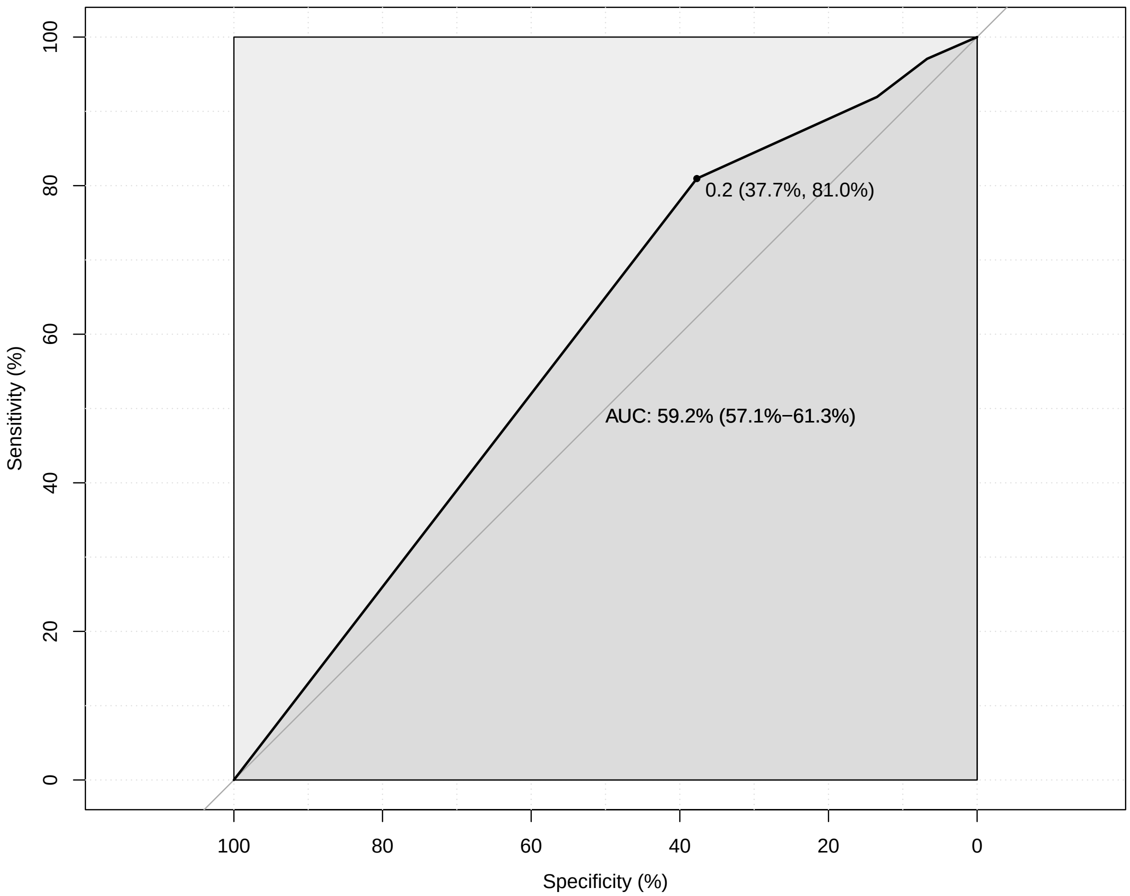

```
Setting levels: control = Syn, case = Ctl
```



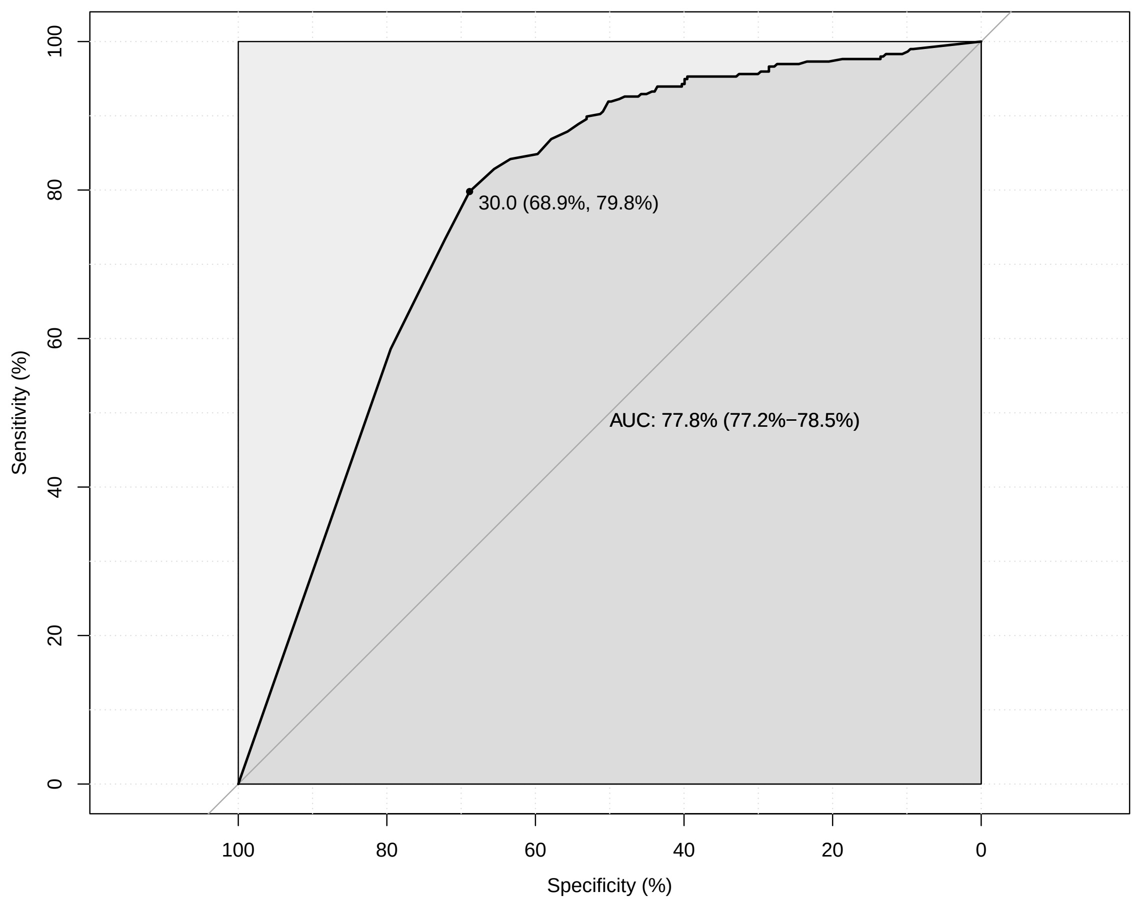

```
Setting levels: control = Syn, case = Ctl
```



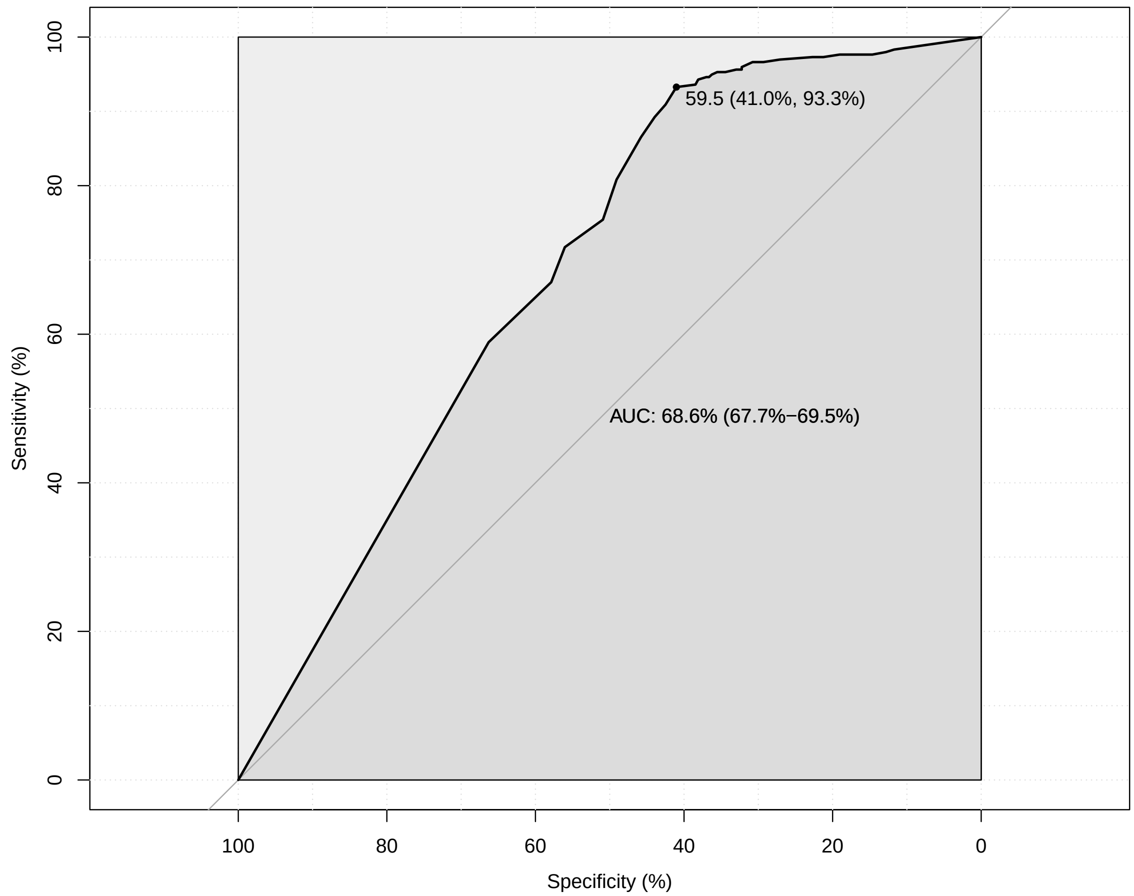

```
Setting levels: control = Syn, case = Ctl
```



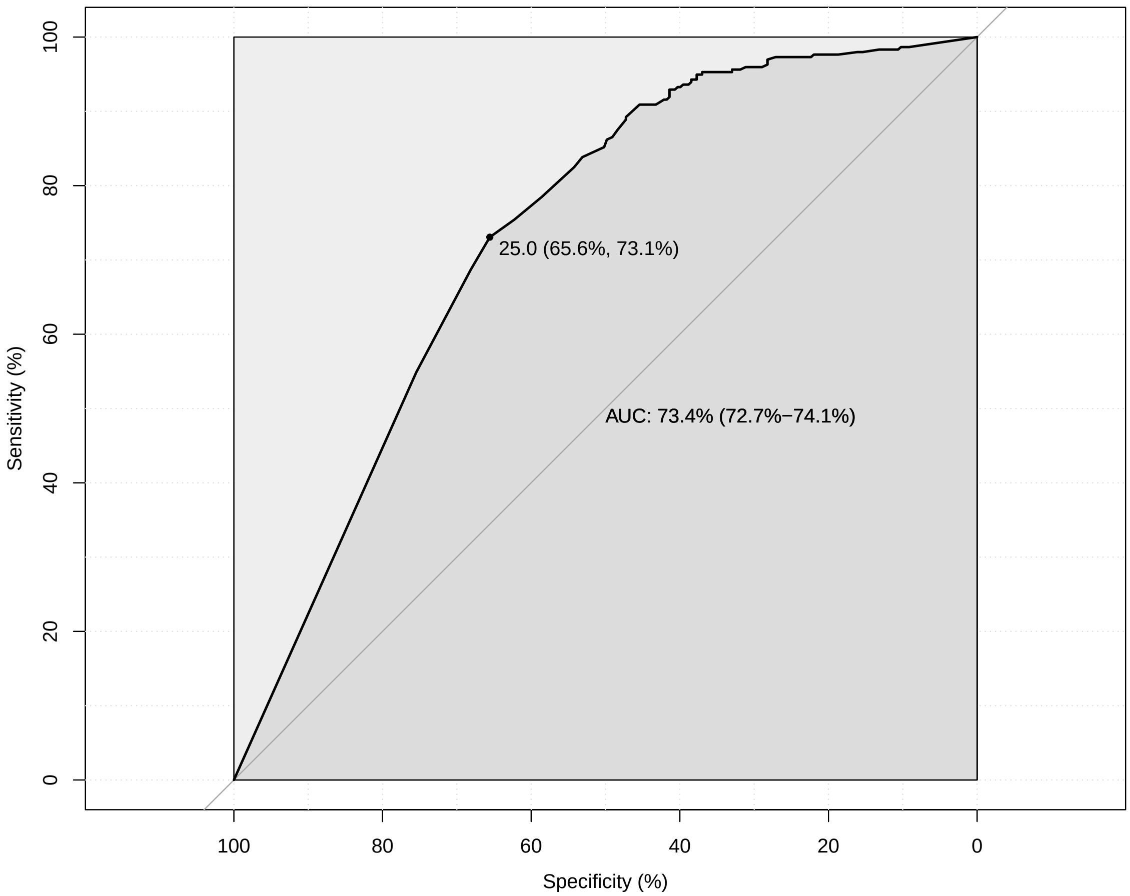

```
Setting levels: control = Ctl, case = Syn
```

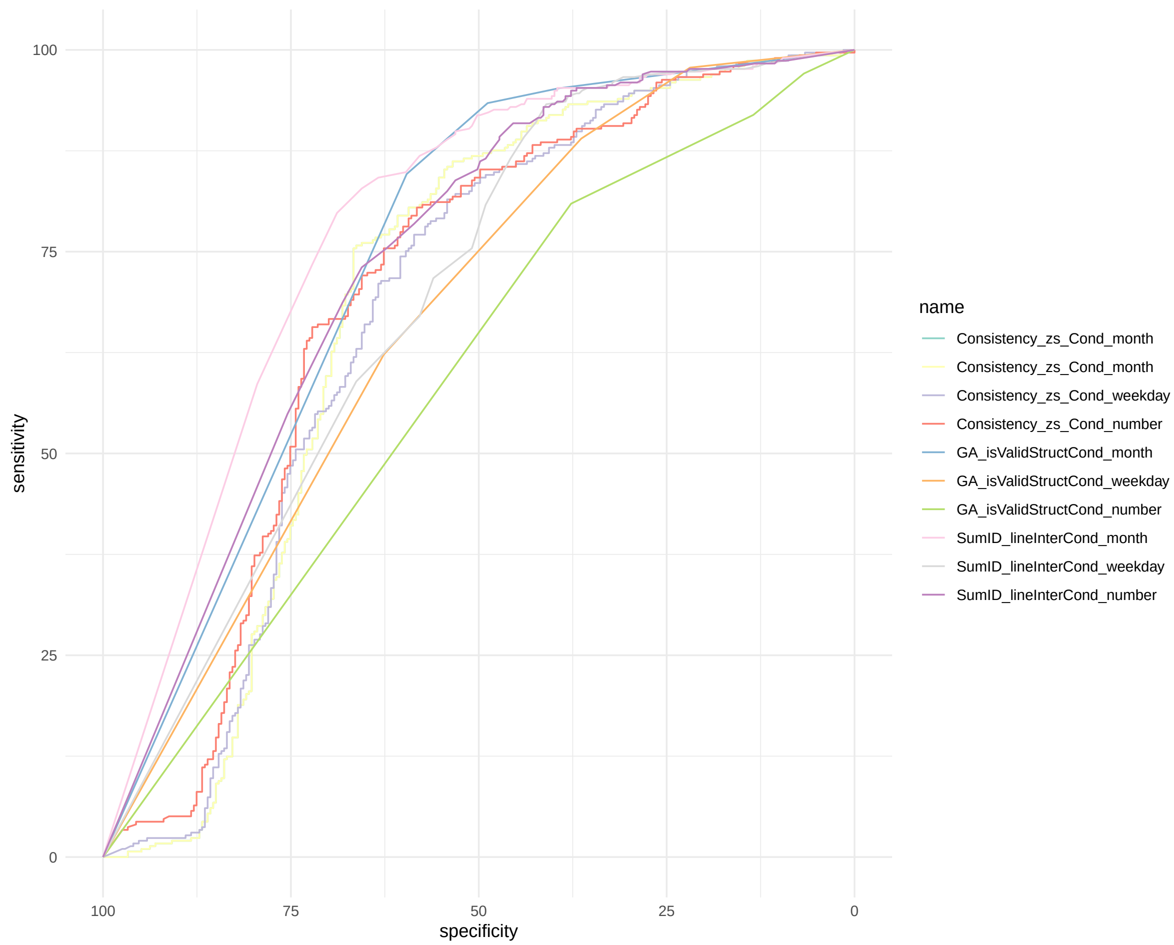



```
Setting levels: control = Ctl, case = Syn
```




```
Setting levels: control = Ctl, case = Syn
```


25.0 (65.6%, 73.1%)
AUC: 73.4% (72.7%−74.1%)
Sensitivity (%)
Specificity (%)
100
80
60
40
20
0



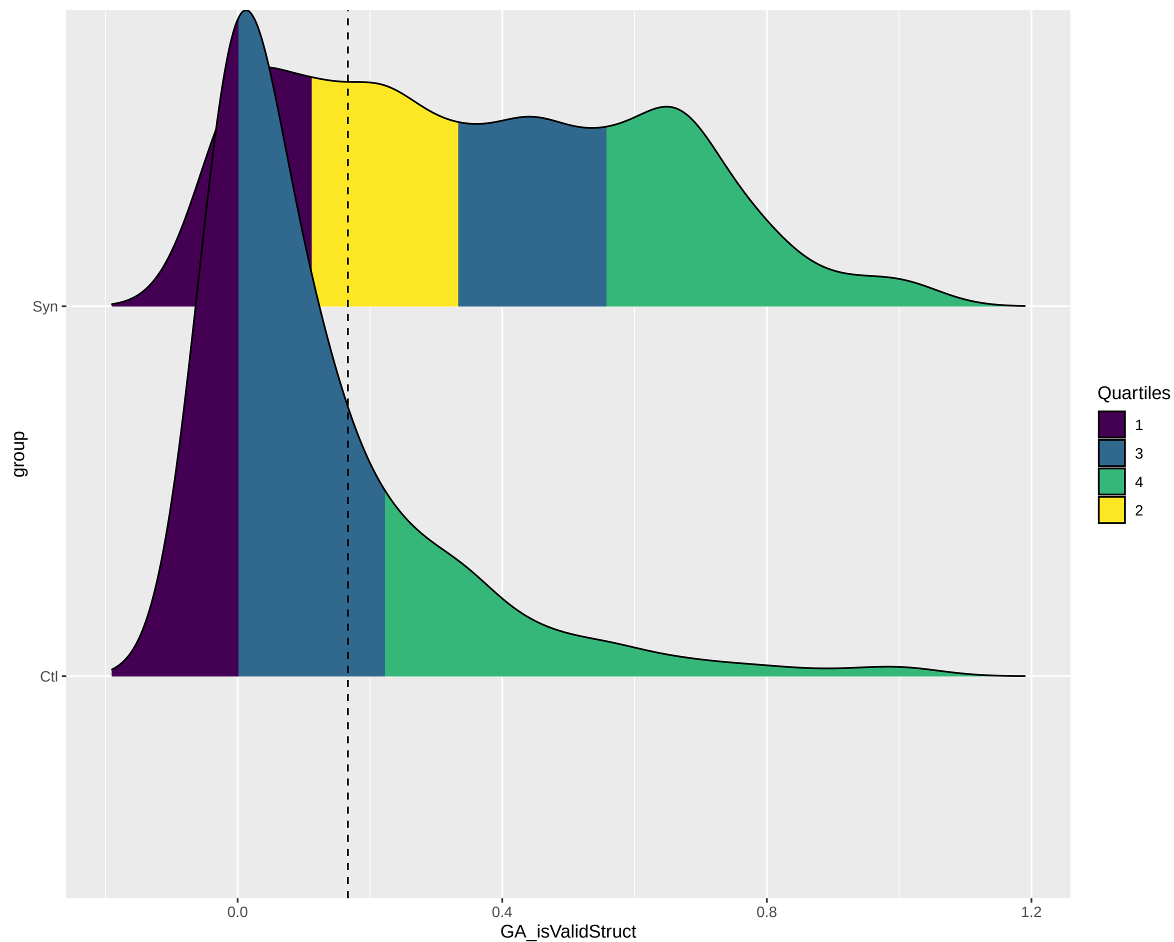
22.2 Summary table Per cond

	Feature	AUC	threshold	sensitivity	specificity	ppv	npv	ci_low	ci_high
7	SumID_lineInterCond_month	77.83823	10.0000000	79.79798	68.86447	73.60248	75.80645	77.21193	78.46449
4	GA_isValidStructCond_month	74.03280	0.1666667	84.61538	59.59596	65.81197	80.82192	72.03161	76.03390
9	SumID_lineInterCond_number	73.39511	25.0000000	73.06397	65.56777	69.77492	69.11197	72.66268	74.12757
3	Consistency_zs_Cond_number	69.32760	0.8323637	80.47138	58.24176	67.70538	73.27189	68.50053	70.15464
8	SumID_lineInterCond_weekday	68.63515	59.5000000	93.26599	41.02564	63.24201	84.84848	67.72618	69.54396
1	Consistency_zs_Cond_month	68.41800	0.7376572	75.42088	66.66667	71.11111	71.37255	67.63694	69.19906
2	Consistency_zs_Cond_weekday	67.21120	0.5401378	77.10438	58.60806	66.95906	70.17544	66.19750	68.22487
5	GA_isValidStructCond_weekday	66.67040	0.5000000	89.01099	36.36364	56.25000	78.26087	64.33453	69.00621
6	GA_isValidStructCond_number	59.16860	0.1666667	80.95238	37.71044	54.43350	68.29268	57.05240	61.28481

23 Test GLM

24 Graph

```
Warning: `stat(quantile)` was deprecated in ggplot2 3.4.0.
i Please use `after_stat(quantile)` instead.

Picking joint bandwidth of 0.0633
```



References

Brang, David, Luke E. Miller, Marguerite McQuire, V. S. Ramachandran, and Seana Coulson. 2013. "Enhanced Mental Rotation Ability in Time-Space Synesthesia." *Cognitive*

Processing 14 (4): 429–34. https://doi.org/10.1007/s10339-013-0561-5.

Deroy, Ophelia, and Charles Spence. 2013. "Why We Are Not All Synesthetes (Not Even Weakly So)." *Psychonomic Bulletin & Review* 20 (4): 643–64. https://doi.org/10.3758/s13423-013-0387-2.

Galton, Francis. 1880. "Visualised Numerals." *Nature* 21 (533): 252–56. https://doi.org/10.1038/021252a0.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

Piazza, M., P. Pinel, and S. Dehaene. 2006. "Objective Correlates of an Unusual Subjective Experience: A Single-Case Study of Number–form Synaesthesia." *Cognitive Neuropsychology* 23 (8): 1162–73. https://doi.org/10.1080/02643290600780080.

Root, Nicholas, Michiko Asano, Helena Melero, Chai-Youn Kim, Anton V. Sidoroff-Dorso, Argiro Vatakis, Kazuhiko Yokosawa, Vilayanur Ramachandran, and Romke Rouw. 2021. "Do the Colors of Your Letters Depend on Your Language? Language-Dependent and Universal Influences on Grapheme-Color Synesthesia in Seven Languages." *Consciousness and Cognition* 95 (October): 103192. https://doi.org/10.1016/j.concog.2021.103192.

Rothen, Nicolas, Kristin Jünemann, Andy D. Mealor, Vera Burckhardt, and Jamie Ward. 2016. "The Sensitivity and Specificity of a Diagnostic Test of Sequence-Space Synesthesia." *Behavior Research Methods* 48 (4): 1476–81. https://doi.org/10.3758/s13428-015-0656-2.

Rothen, Nicolas, Anil K. Seth, Christoph Witzel, and Jamie Ward. 2013. "Diagnosing Synaesthesia with Online Colour Pickers: Maximising Sensitivity and Specificity." *Journal of Neuroscience Methods* 215 (1): 156–60. https://doi.org/10.1016/j.jneumeth.2013.02.009.

Sagiv, N, J Simner, J Collins, B Butterworth, and J Ward. 2006. "What Is the Relationship Between Synaesthesia and Visuo-Spatial Number Forms?" *Cognition* 101 (1): 114–28. https://doi.org/10.1016/j.cognition.2005.09.004.

Seron, Xavier, Mauro Pesenti, Marie-Pascale Noël, Gérard Deloche, and Jacques-André Cornet. 1992. "Images of numbers, or "when 98 is upper left and 6 sky blue"." *Cognition*, Numerical Cognition, 44 (1): 159–96. https://doi.org/10.1016/0010-0277(92)90053-K.

Van Petersen, Eline, Mareike Altgassen, Rob Van Lier, and Tessa M. Van Leeuwen. 2020b. "Enhanced Spatial Navigation Skills in Sequence-Space Synesthetes." *Cortex* 130 (September): 49–63. https://doi.org/10.1016/j.cortex.2020.04.034.

———. 2020a. "Enhanced Spatial Navigation Skills in Sequence-Space Synesthetes." *Cortex* 130 (September): 49–63. https://doi.org/10.1016/j.cortex.2020.04.034.

Ward, Jamie. n.d. "Optimizing a Measure of Consistency for Sequence-Space Synaesthesia." https://doi.org/10.31234/osf.io/5cnr7.

Ward, Jamie, Alberta Ipser, Eva Phanvanova, Paris Brown, Iris Bunte, and Julia Simner. 2018. "The Prevalence and Cognitive Profile of Sequence-Space Synaesthesia." *Consciousness and Cognition* 61 (May): 79–93. https://doi.org/10.1016/j.concog.2018.03.012.