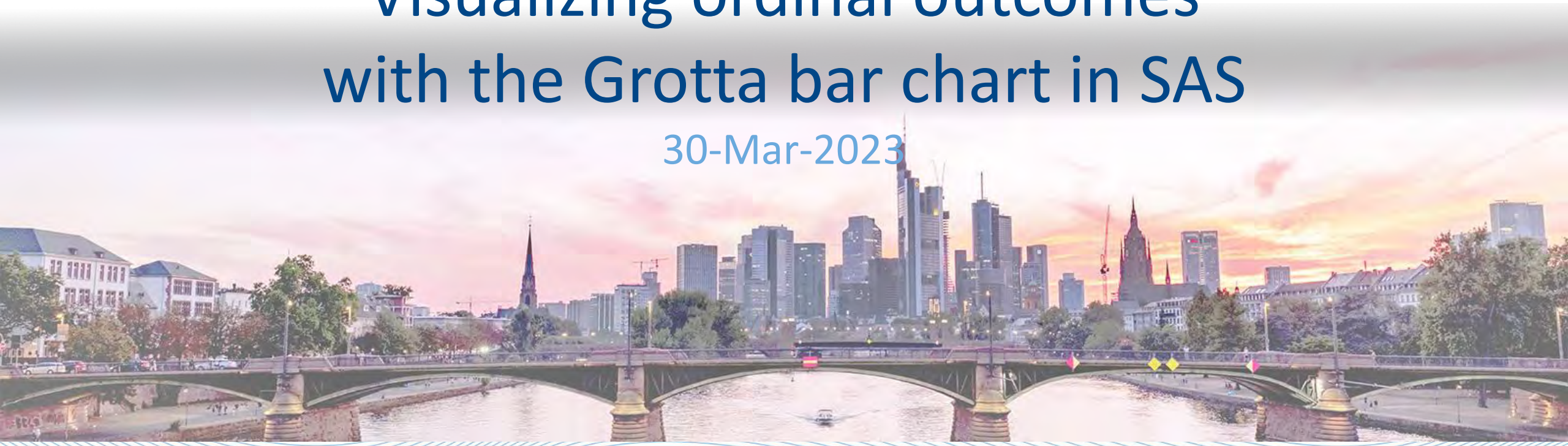
main analytics

Visualizing ordinal outcomes with the Grotta bar chart in SAS

30-Mar-2023



www.mainanalytics.de

Yann Féat
yann.feat@mainanalytics.de

Contents

Introduction

Background, example use cases and presentation of the Grotta plot

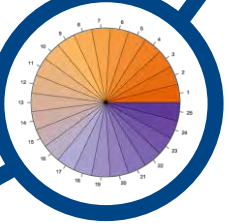
Overview: creating a Grotta bar chart with R

General process description, from preprocessing to plotting

Displaying a Grotta bar chart using SAS

Data processing and plotting with SAS procedures
FREQ, SGPLOT and SGPANEL

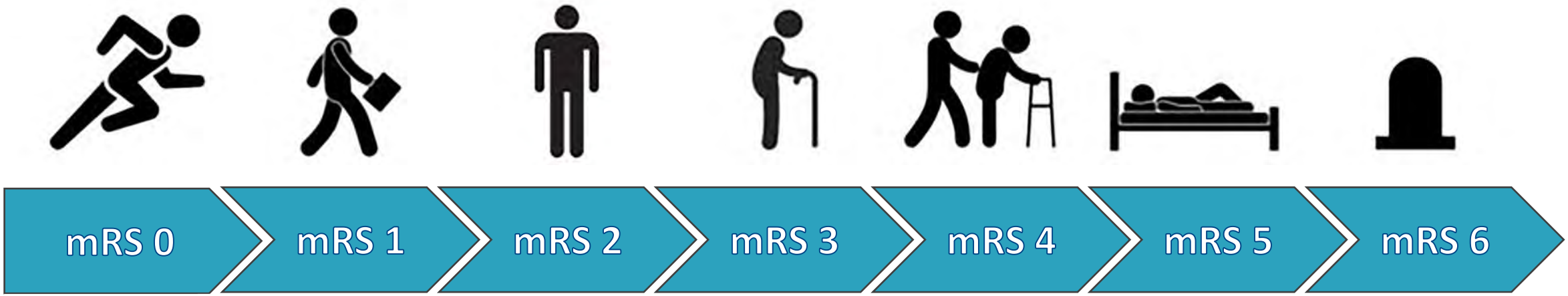
Color palettes



Presentation of an approach to handle color schemes for R and SAS

Introduction

Background: modified Rankin Scale



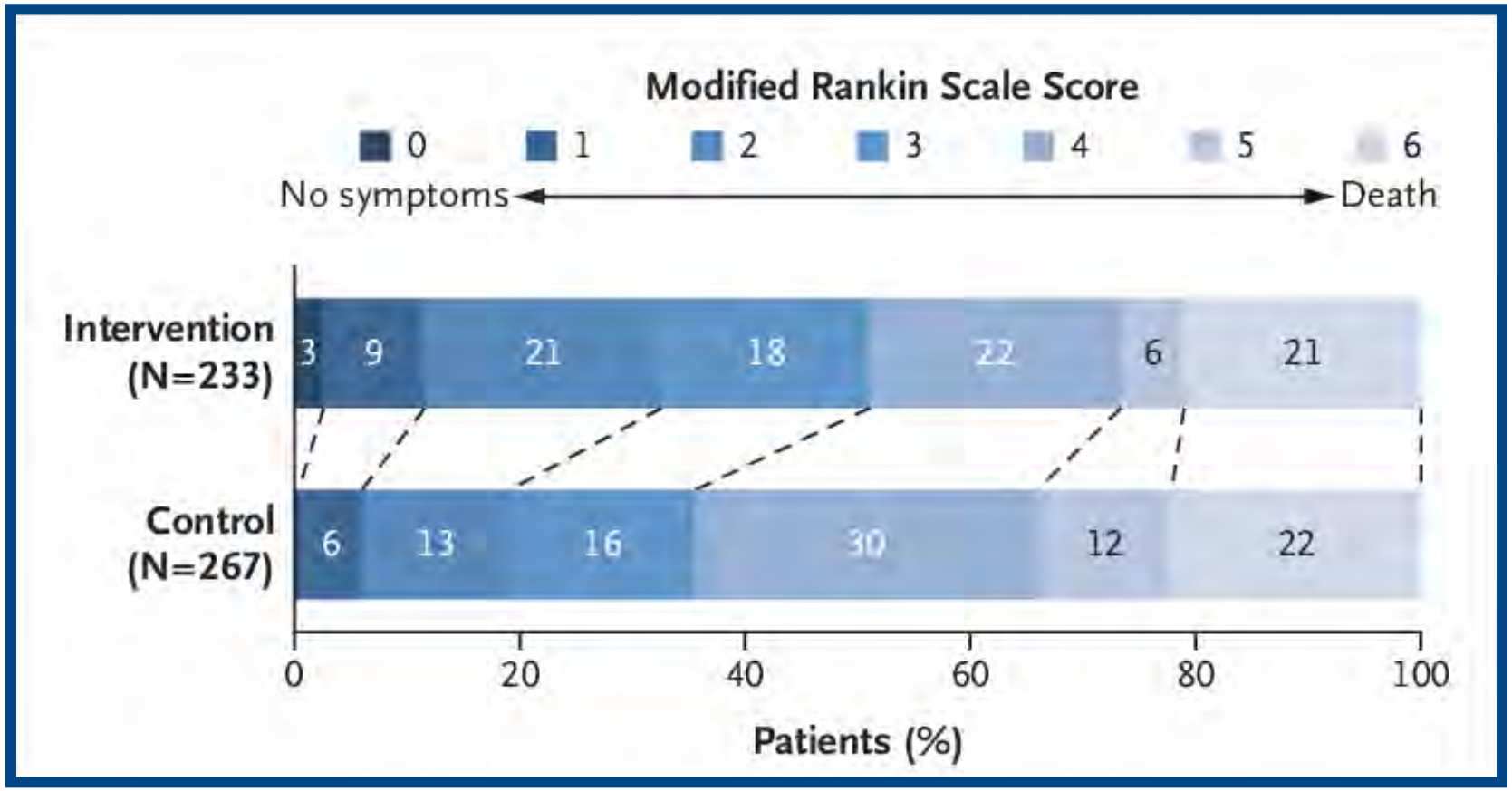
- Degree of disability or dependence in the daily activities of people who have suffered a stroke or other causes of neurological disability

Introduction

Pre-stroke mRS— no. (%)	Intervention (N = 233)	Control (N = 267)
0	7 (3.0)	2 (0.7)
1	21 (9.0)	16 (6.0)
2	49 (21.0)	35 (13.1)
3	42 (18.0)	43 (16.1)
4	51 (21.9)	80 (30.0)
5	14 (6.0)	32 (12.0)
6	49 (21.0)	59 (22.1)

A randomized trial of intraarterial treatment for acute ischemic stroke. (2015). *New England Journal of Medicine, 372*(1), 11–20

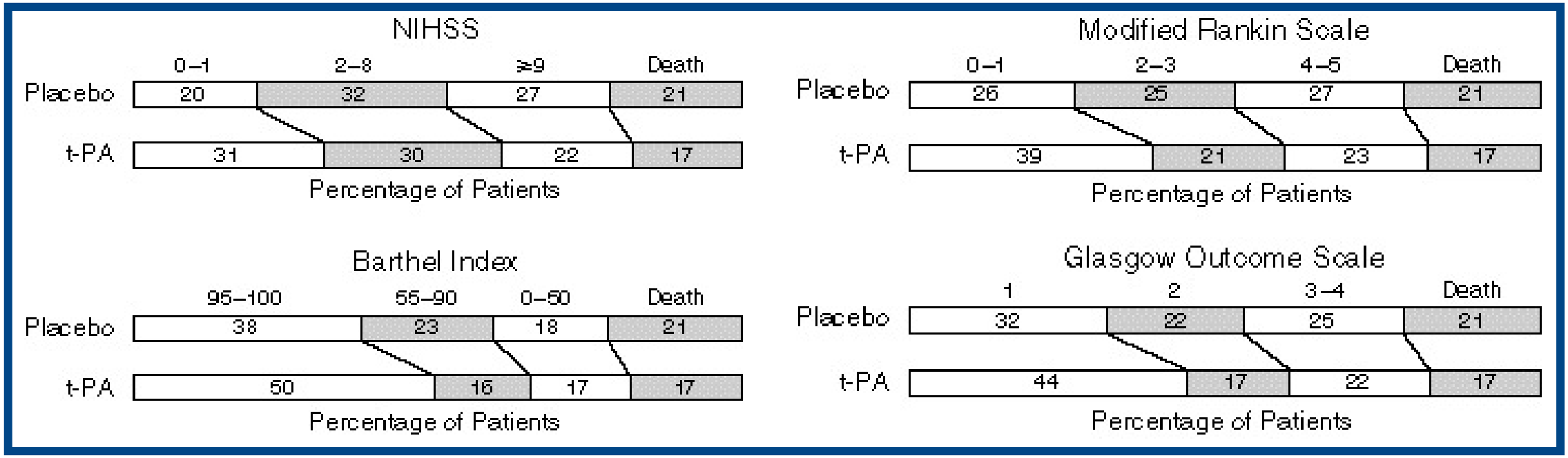
Introduction



A randomized trial of intraarterial treatment for acute ischemic stroke. (2015). *New England Journal of Medicine, 372*(1), 11–20

Introduction

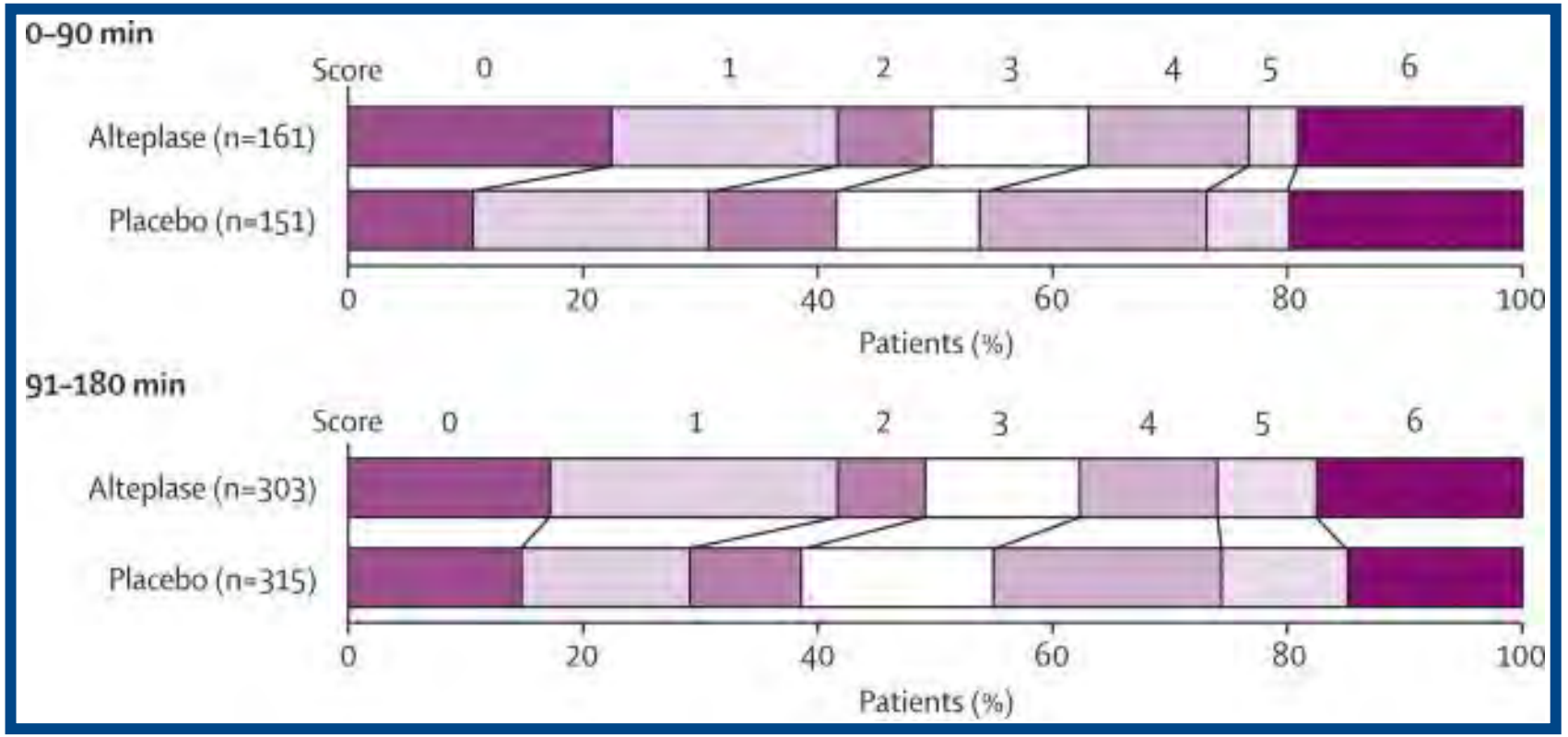
„Original“ Grotta bar chart – with several outcomes



Tissue plasminogen activator for acute ischemic stroke. (1995).
New England Journal of Medicine, 333(24), 1581–1588

Introduction

Working example – stratified



Time to treatment with intravenous alteplase and outcome in stroke: An updated pooled analysis of ECASS, Atlantis, ninds, and epithet trials. (2010). *The Lancet, 375*(9727), 1695–1703.

Overview: creating a Grotta bar chart with R

Useful libraries

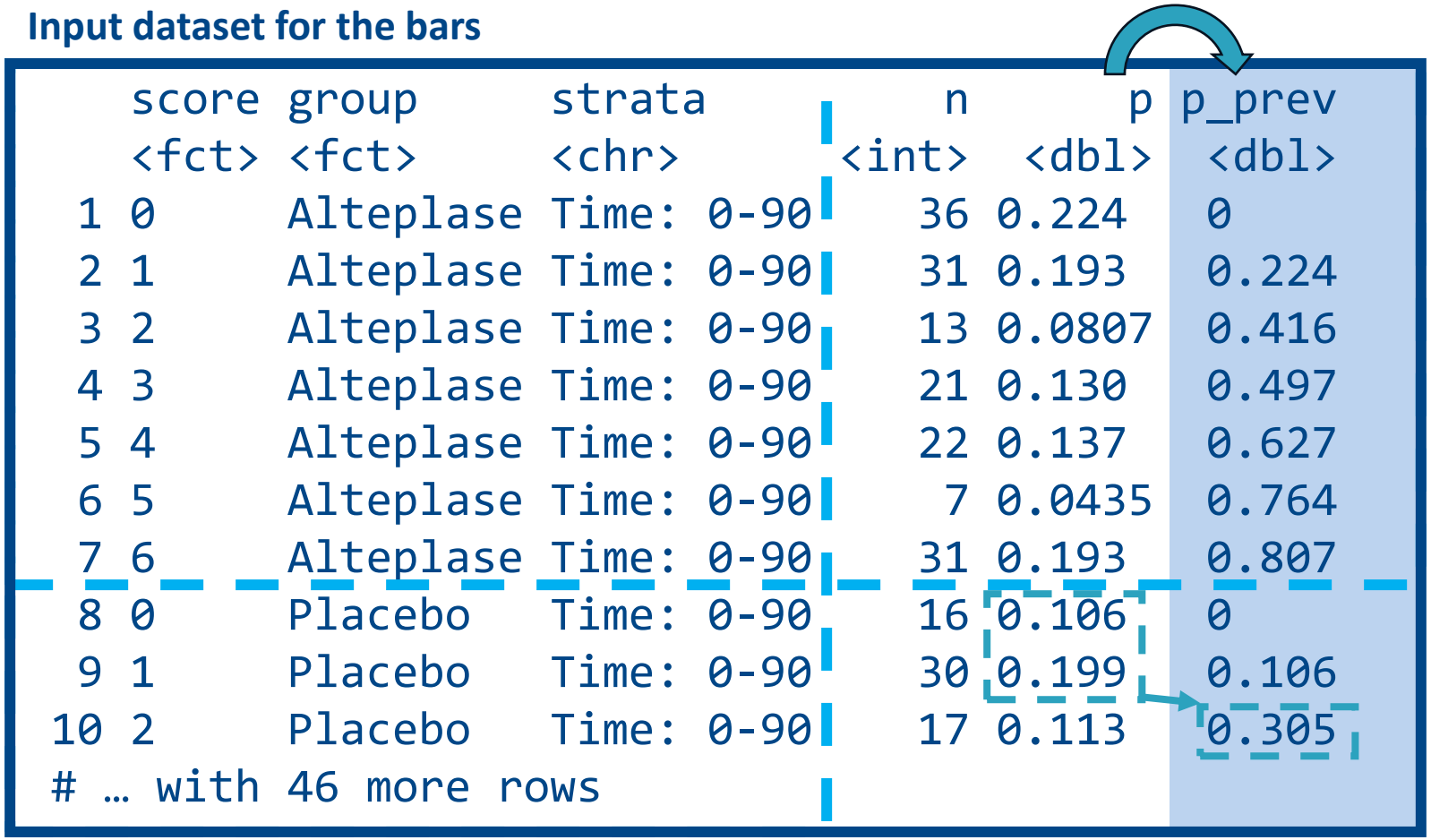
- dplyr
- purrr
- tidyverse
- ggplot2

Data

```
  time        treat          mRS
  <fct>       <fct>          <fct>
1 0-90        Alteplase      0
2 271-360     Alteplase      0
3 0-90        Placebo        3
# … with 3666 more rows
```

Overview: creating a Grotta bar chart with R

Input dataset for the bars



	score	group	strata	n	p	p_prev
	<fct>	<fct>	<chr>	<int>	<dbl>	<dbl>
1	0	Alteplase	Time: 0-90	36	0.224	0
2	1	Alteplase	Time: 0-90	31	0.193	0.224
3	2	Alteplase	Time: 0-90	13	0.0807	0.416
4	3	Alteplase	Time: 0-90	21	0.130	0.497
5	4	Alteplase	Time: 0-90	22	0.137	0.627
6	5	Alteplase	Time: 0-90	7	0.0435	0.764
7	6	Alteplase	Time: 0-90	31	0.193	0.807
8	0	Placebo	Time: 0-90	16	0.106	0
9	1	Placebo	Time: 0-90	30	0.199	0.106
10	2	Placebo	Time: 0-90	17	0.113	0.305

… with 46 more rows

Overview: creating a Grotta bar chart with R

Input dataset for the connecting lines

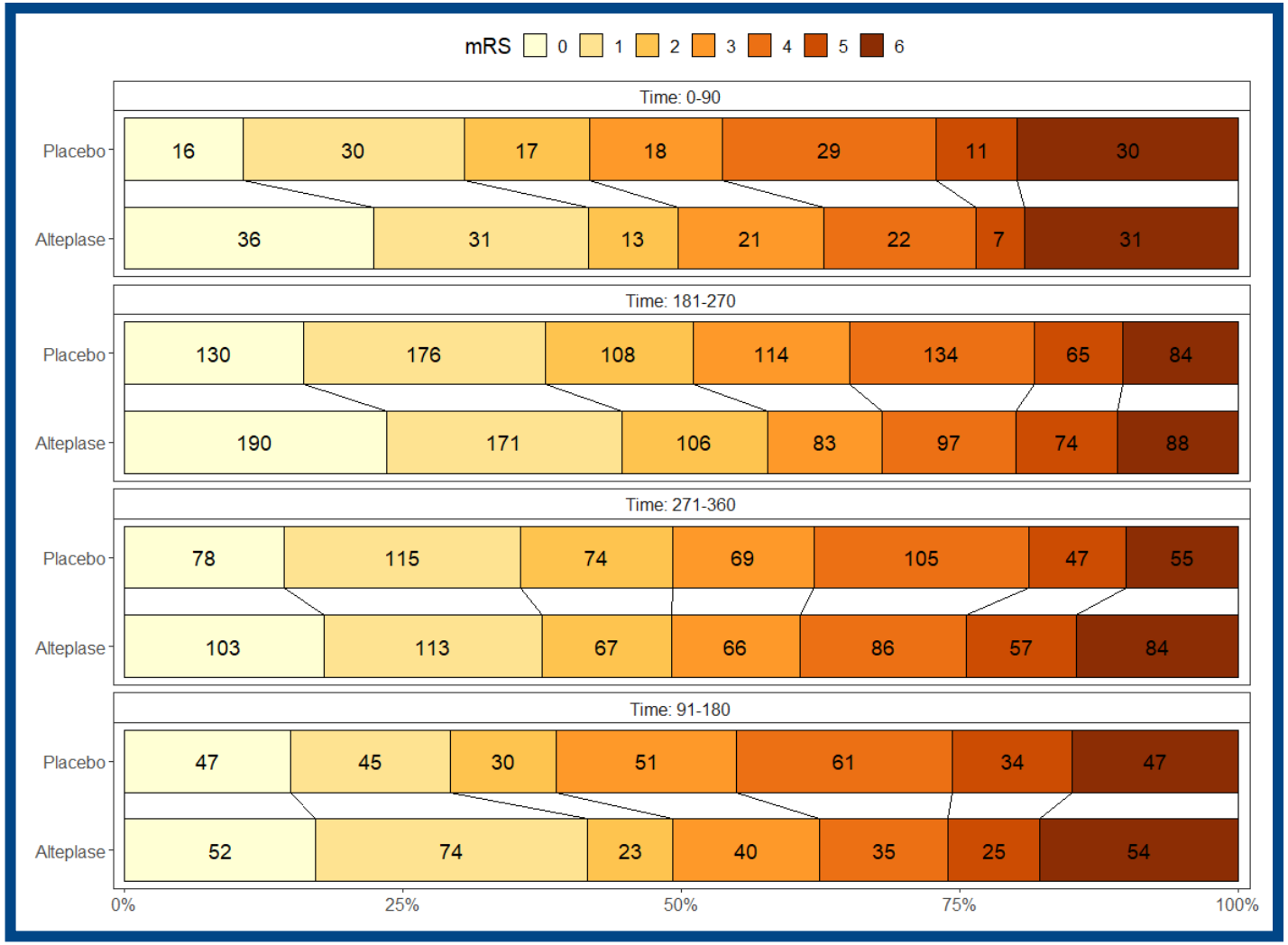
	strata	score	group	n	p	p_prev	line_id
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	Time: 0-90	0	1.45	0	0	0	1
2	Time: 0-90	1	1.45	36	0.224	0	2
3	Time: 0-90	2	1.45	31	0.193	0.224	3
4	Time: 0-90	3	1.45	13	0.0807	0.416	4
5	Time: 0-90	4	1.45	21	0.130	0.497	5
6	Time: 0-90	5	1.45	22	0.137	0.627	6
7	Time: 0-90	6	1.45	7	0.0435	0.764	7
8	Time: 0-90	7	1.45	31	0.193	0.807	8
9	Time: 0-90	0	1.55	0	0	0	1
10	Time: 0-90	1	1.55	16	0.106	0	2
# … with 54 more rows							

y_1 x_1 y_2 x_2

One line

Vertical space between bars: WIDTH = y_2 - y_1

Overview: creating a Grotta bar chart with R

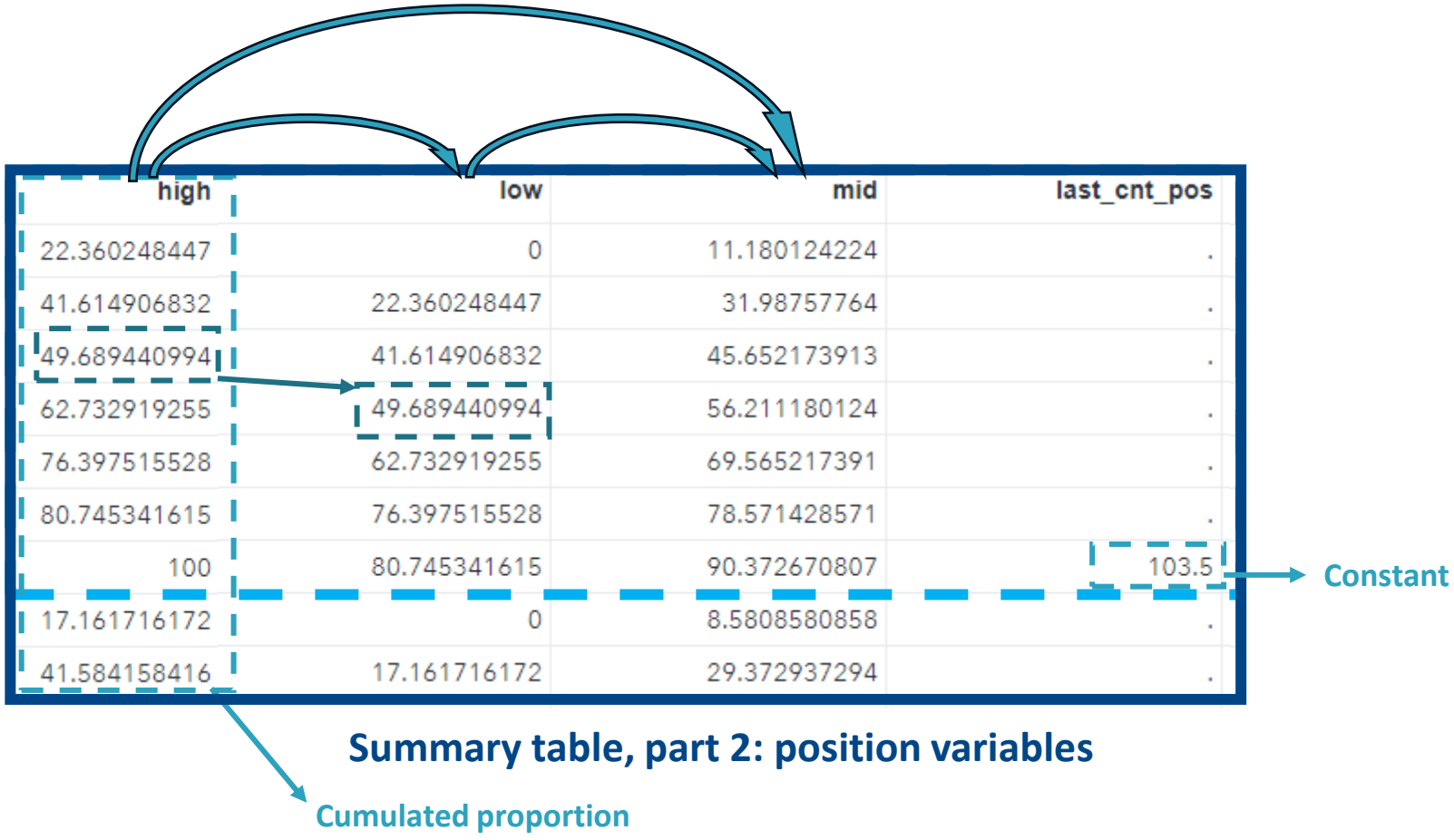


Displaying a Grotta bar chart using SAS

class1_rank	class2_rank	outcome_rank	count	percent
1	1	0	36	22.360248447
1	1	1	31	19.254658385
1	1	2	13	8.0745341615
1	1	3	21	13.043478261
1	1	4	22	13.664596273
1	1	5	7	4.347826087
1	1	6	31	19.254658385
1	2	0	52	17.161716172
1	2	1	74	24.422442244

Summary table, part 1: PROC FREQ output

Displaying a Grotta bar chart using SAS



high	low	mid	last_cnt_pos
22.360248447	0	11.180124224	.
41.614906832	22.360248447	31.98757764	.
49.689440994	41.614906832	45.652173913	.
62.732919255	49.689440994	56.211180124	.
76.397515528	62.732919255	69.565217391	.
80.745341615	76.397515528	78.571428571	.
100	80.745341615	90.372670807	103.5
17.161716172	0	8.5808580858	.
41.584158416	17.161716172	29.372937294	.

Summary table, part 2: position variables

Displaying a Grotta bar chart using SAS

highcnt	last_cnt_label	class1_name	class2_name	outcome_na...
36		Alteplase	0-90	mRS 0
67		Alteplase	0-90	mRS 1
80		Alteplase	0-90	mRS 2
101		Alteplase	0-90	mRS 3
123		Alteplase	0-90	mRS 4
130		Alteplase	0-90	mRS 5
161	N=161	Alteplase	0-90	mRS 6
52		Alteplase	91-180	mRS 0
126		Alteplase	91-180	mRS 1

Retrieved from the numeric values of the grouping variables

Summary table, part 3: labelling variables

Displaying a Grotta bar chart using SAS

Summary table: Step 1

```
* Formatted contingency table;

proc freq data=ds_&randint;
  table class1_rank*class2_rank*outcome_rank / out=summary_data_&randint outpct;
run;

proc sort data=summary_data_&randint out=summary_data_&randint;
  by class1_rank class2_rank outcome_rank;
run;

data summary_data_&randint(keep=class1_rank class2_rank outcome_rank percent cnt high highcnt);
  set summary_data_&randint(drop=percent);
  rename pct_row=percent count=cnt;
  length high highcnt 8;
  by class1_rank class2_rank;
  if first.class2_rank then do;
    high = pct_row;
    highcnt = count;
  end;
  else do;
    high + pct_row;
    highcnt + count;
  end;
run;
```

Stratification (time) Group (treatment) Outcome (mRS score)

Displaying a Grotta bar chart using SAS

Summary table: Step 2

```
data summary_data_&randint;
  set summary_data_&randint;
  label high='Percentage';
  length low highcnt mid 8 last_cnt_pos 3 last_cnt_label $ 18;
  low = lag(high);
  by class1_rank class2_rank;
  if first.class2_rank then low = 0;
  mid = (high+low)/2;
  if last.class2_rank then do;
    last_cnt_label = compress(cat("N=", put(highcnt, 8.)));
    last_cnt_pos = 100 + &total_cnt_offset;
  end;
  if cnt=0 then cnt = .;
run;
```

Lag function applied to a calculated variable by group: requires a second data step

Displaying a Grotta bar chart using SAS

Summary table:

Step 3

```
data summary_data_&randint;
  merge summary_data_&randint dic_class1_&randint;
  by class1_rank;
run;

proc sort data=summary_data_&randint;
  by class2_rank;
run;

data summary_data_&randint;
  merge summary_data_&randint dic_class2_&randint;
  by class2_rank;
run;

proc sort data=summary_data_&randint;
  by outcome_rank;
run;
```

```
data summary_data_&randint;
  rename cnt=count;
  retain outcome_id -1;
  merge summary_data_&randint dic_outcome_&randint;
  by outcome_rank;
  if first.outcome_rank then outcome_id + 1;
run;

proc sort data=summary_data_&randint;
  label outcome_name="&var_outcome_label";
  by class1_rank class2_rank outcome_rank;
run;
```

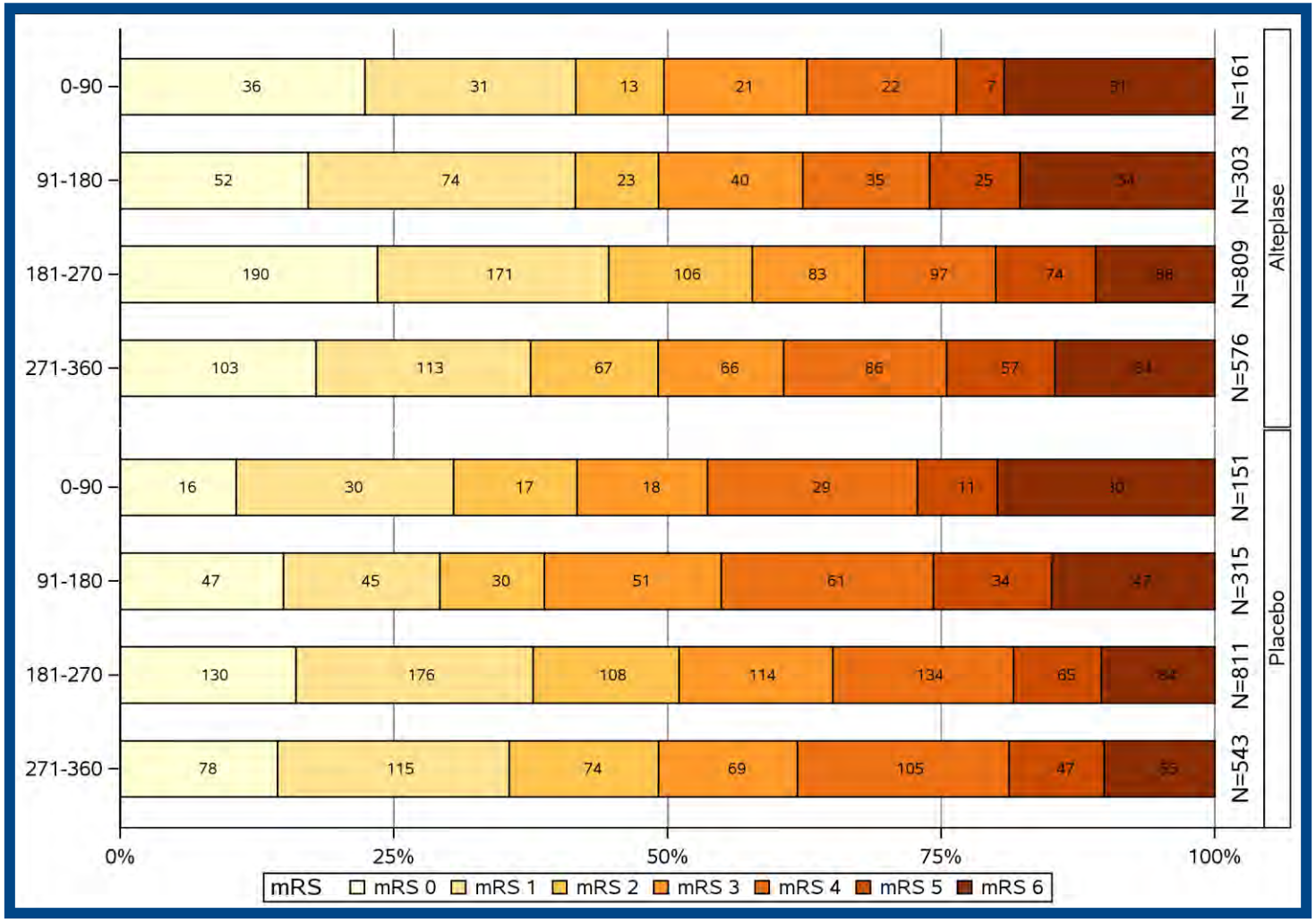
class2_name	class2_rank
0-90	1
91-180	2
181-270	3
271-360	4

class1_name	class1_rank
Alteplase	1
Placebo	2

outcome_rank	outcome_na...
0	mRS 0
1	mRS 1
2	mRS 2
3	mRS 3

Displaying a Grotta bar chart using SAS

Two-way graph



Displaying a Grotta bar chart using SAS

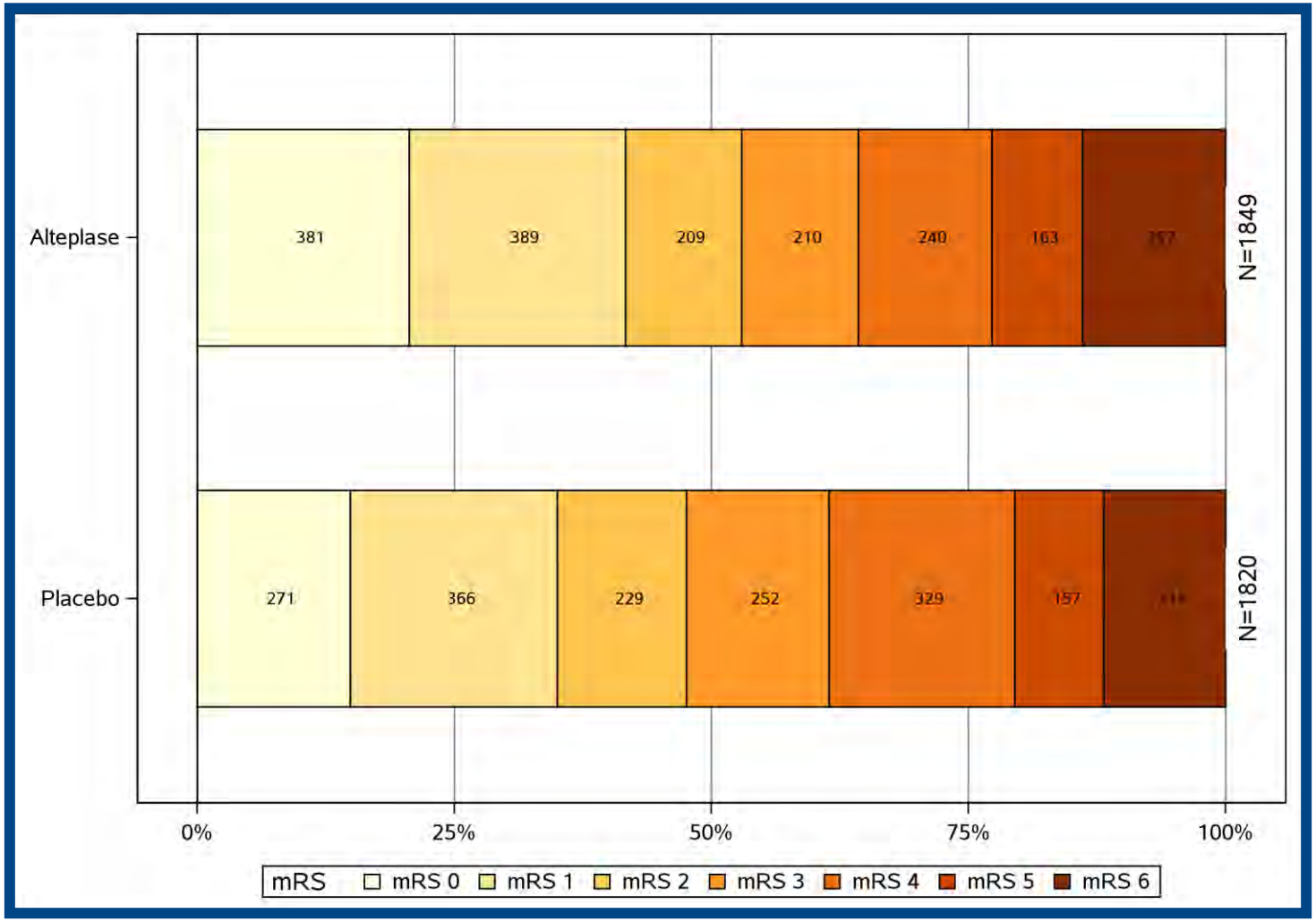
```
proc sgpanel data=summary_data_&randint;
  styleattrs datacolors=(&colors_outcome);
  rowaxis display=(nolabel);
  colaxis display=(nolabel) values=(0 25 50 75 100) valuesdisplay=("0%" "25%" "50%" "75%" "100%") grid
          gridattrs=(color=grey pattern=solid thickness=1);

  panelby class1_name / layout=rowlattice novarname noborder colheaderpos=bottom spacing=&panel_spacing;

  hbarparm category=class2_name response=percent /
    group=outcome_name groupdisplay=stack outlineattrs=(color=black) barwidth=&bar_width;
  scatter y=class2_name x=mid / markerchar=count;
  text y=class2_name x=last_cnt_pos text=last_cnt_label /
    rotate=90 position=top backfill fillattrs=(color=white) textattrs=(size=10pt);
run;
```

Displaying a Grotta bar chart using SAS

One-way graph



Displaying a Grotta bar chart using SAS

Two-way graph

```
proc sgpanel data=summary_data_&randint;
  styleattrs datacolors=(&colors_outcome);
  rowaxis display=(nolabel);
  colaxis display=(nolabel) values=(0 25 50 75 100) valuesdisplay=("0%" "25%" "50%" "75%" "100%") grid
          gridattrs=(color=grey pattern=solid thickness=1);

  panelby class1_name / layout=rowlattice novarname noborder colheaderpos=bottom spacing=&panel_spacing;

  hbarparm category=class2_name response=percent /
    group=outcome_name groupdisplay=stack outlineattrs=(color=black) barwidth=&bar_width;
  scatter y=class2_name x=mid / markerchar=count;
  text y=class2_name x=last_cnt_pos text=last_cnt_label /
    rotate=90 position=top backfill fillattrs=(color=white) textattrs=(size=10pt);
run;
```

One-way graph

```
proc sgplot data=summary_data_&randint;
  styleattrs datacolors=(&colors_outcome);
  xaxis display=(nolabel) values=(0 25 50 75 100) valuesdisplay=("0%" "25%" "50%" "75%" "100%") grid
        gridattrs=(color=grey pattern=solid thickness=1);
  yaxis display=(nolabel);

  hbarparm category=class2_name response=percent /
    group=outcome_name groupdisplay=stack outlineattrs=(color=black) barwidth=&bar_width;
  scatter y=class2_name x=mid / markerchar=count;
  text y=class2_name x=last_cnt_pos text=last_cnt_label /
    rotate=90 position=top backfill fillattrs=(color=white) textattrs=(size=10pt);
run;
```

R and SAS comparison

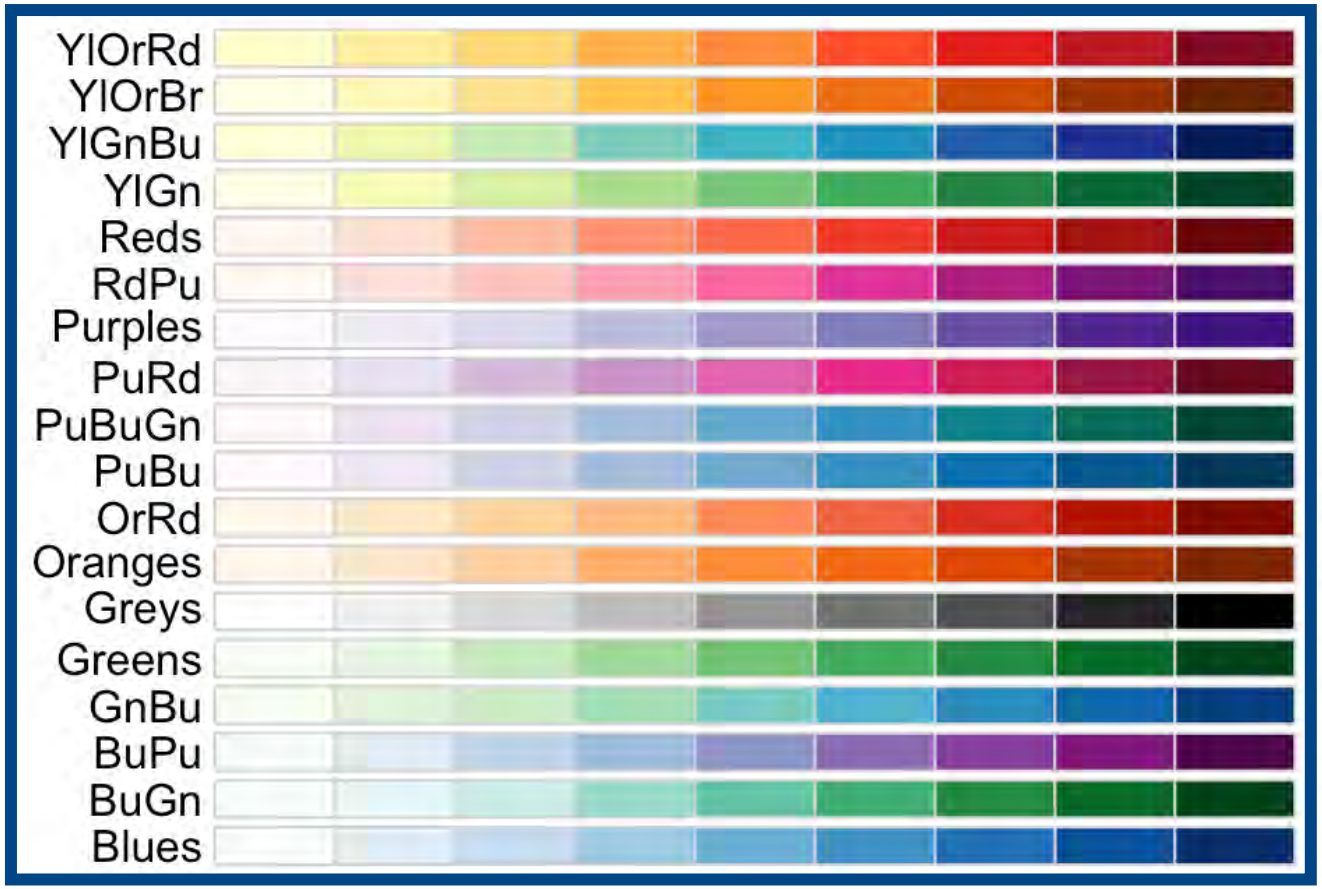
R code

- Column oriented
- Possibility of using different datasets to draw shapes and overlay them
- Gives more control over the output
- Relies on external packages for efficient programming (e.g., tidyverse, ggplot2)
- Easier to make reusable

SAS code

- Row oriented
- Procedures SGPANEL and SGPLOT restricted to one input dataset and thus involves more preprocessing
- Requires changing fewer graphic options to get a “clean” result
- Involves more preprocessing

Color palettes



Sequential palettes: suited to data that progresses from low to high, i.e., quantitative or ordinal

Color palettes in R

Possibility to use the *RColorBrewer* package directly

```
> library(RColorBrewer)
> brewer.pal(7, "YlOrBr")
[1] "#FFFFD4" "#FEE391" "#FEC44F" "#FE9929" "#EC7014" "#CC4C02" "#8C2D04"
```

***RColorBrewer* integrated into *ggplot2* functions for more convenience**

```
# Plot
ggplot(tbl) +
  geom_rect(color = "black",
            aes(xmin=group-WIDTH/2, xmax=group+WIDTH/2,
                ymin=p_prev, ymax=p_prev+p, fill=score)) +
  geom_line(data = y, aes(x=group, y=p+p_prev, group=line_id)) +
  geom_text(data = tbl[which(tbl$n>0),], size = 5,
            aes(x=group, y=p_prev+0.5*p,
                color=as.numeric(score) > 0, label=sprintf("%d", n))) +
  scale_fill_brewer(palette = "YlOrBr", direction = 1)+
```

Color palettes in SAS

- BREWERPAL macro modeled on the *RColorBrewer* R package:
  York University. *SAS Macro Programs: Brewerpal*. brewerpal - Generate Brewer color palette(s), euclid.psych.yorku.ca/datavis/sasmac/brewerpal.html
  - Relies on a dataset containing color names provided by the website
  - Sample call: `%brewerpal(n=7, palette=YlOrBr)`

References

- A randomized trial of intraarterial treatment for acute ischemic stroke. (2015). *New England Journal of Medicine, 372*(1), 11–20
- Time to treatment with intravenous alteplase and outcome in stroke: An updated pooled analysis of ECASS, Atlantis, ninds, and epithet trials. (2010). *The Lancet, 375*(9727), 1695–1703.
- Tissue plasminogen activator for acute ischemic stroke. (1995). *New England Journal of Medicine, 333*(24), 1581–1588
- Saver, J. L et al. (2021). Standardized nomenclature for modified rankin scale global disability outcomes: Consensus recommendations from stroke therapy academic industry roundtable XI. *Stroke*, 52(9), 3054–3062

References

- York University. (2005) SAS Macro Programs: Brewerpal. brewerpal - Generate Brewer color palette(s). Retrieved March 20, 2023, from euclid.psych.yorku.ca/datavis/sasmac/brewerpal.html
- Johns, H. (2013) rankinPlot: Convenient Plotting for the Modified Rankin Scale and Other Ordinal Outcome Data. CRAN
- Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes (R Package Version 1.1-3)

Thank you for your attention!

yann.feat@mainanalytics.de