

Landkarten in SAS

Thomas Rüdiger
 AXA Konzern AG
 Colonia-Allee 19-21
 51067 Köln
 thomas.ruediger@axa.de

Zusammenfassung

Neben der altbewährten Prozedur PROC GMAP sind Landkarten ab der Version SAS 9.4 auch über die neue Prozedur PROC SGPLOT möglich. Beide Prozeduren verwenden unterschiedliche Datenquellen über die SAS-eigenen Libraries MAPS (gmap) und MAPSGFK. Als Vorteil stellt sich die Option dataskin dar, durch die Grafiken aufgrund Schattierung einen modernen und präsentablen Eindruck erzeugen. Zudem erleichtern SAS-eigene Macros wie %annomac und %centroid(<Map_Data>,<Ausgabe>,ID) die Erstellung der annotate-Tabelle.

Schlüsselwörter: Graphik, PROC GMAP, PROC SGPLOT

1 Einleitung

1.1 Verschiedene Landkarten-Prozeduren in SAS

SAS bietet 2 Prozeduren zu Landkartenauswertungen: PROC GMAP und PROC SGPLOT. Nicht nur in der Datenaufbereitung unterscheiden sich beide Prozeduren, sondern bereits in den von SAS zur Verfügung gestellten Daten. PROC GMAP arbeitet mit der SAS-Library MAPS: in dieser Library sind die Daten abgelegt entsprechend der ursprünglichen Vorgabe für Tabellennamen. Mehr als 8 Zeichen für den Dateinamen waren zu Anfang der Software SAS nicht erlaubt. Zu einer Tabelle mit Koordinaten (Beispiel MAPS.GERMANY) werden beschreibende Daten in einer Tabelle abgelegt, deren Tabellennamen mit den Koordinatentabellennamen ähnlich ist und in der Regel mit der Ziffer 2 endet (Beispiel MAPS.GERMANY2). In der für SGPLOT konzipierten Library MAPSGFK gilt die Erweiterung auf 32 Zeichen im Tabellennamen. Zu einer Koordinatentabellen (Beispiel MAPS.GERMANY) gibt es eine beschreibende Attribut-Tabelle (Beispiel MAPS.GERMANY_ATTR).

Tabelle 1: Anfangsfragestellungen zu beiden Prozeduren

Frage	PROC GMAP	PROC SGPLOT
In welcher Library liegen die Daten für Landkarten und Beschreibungen (Annotate-Daten)?	MAPS	MAPSGFK (Geodaten der Unternehmensgruppe GfK)

Wie alt ist die Prozedur?	Uralt (Tabellennamen für Geodaten in Library MAPS sind max. 8 Zeichen lang). Sie gehört zu SAS/GRAPH	Die Prozedur gibt es seit SAS 9.4 und gehört kostengünstig zu SAS/Base
GERMANY-Tabellen basieren auf sog. Landkreisen. Das sind 5-stellige Kreisgemeindegemeinschaften (KGS, Feldname ID). Gibt es Unterschiede?	413 KGS (11/2008 waren es 439 KGS) ID ist seit Juni 2013 eine laufende Nummer, davor analog SGPLOT	402(=amtliche Zahl) KGS ID entsprechend Aufbau des KGS (inkl. Bundesland, Regierungsbezirk)
Gibt es weitere Unterschiede bei Länderdaten?	Daten zu Kuba und Haiti gibt es nur in MAPS.	MAPSGFK enthält gegenüber MAPS mehr afrikanische Länder, z.B. Elfenbeinküste, Uganda
Wie lassen sich Grafik-Optionen einstellen?	Per GOPTIONS: goptions reset=all cback=white border htitle=12pt htext=10pt;	Per ODS (verwendbar z.B. für pdf): ods graphics / reset width=600px height=400px imagefmt=png imagename='Map' imagemap=on tipmax=4000;

1.2 Verwendete Daten

Der graphische Output zum Vergleich der beiden Prozeduren bezieht sich auf die Anzahl Großstädte pro Land. Die Datenquelle ist: <https://www.worlddata.info/>



Abbildung 1: Ausschnitt der Daten für die beiden SAS-Prozeduren

2 SAS-Landkarten-Prozeduren im Überblick

2.1 PROC GMAP

PROC GMAP gehört zum Programmpaket von SAS/GRAPH. Die finale Graphik ist in Abbildung 2 zu sehen. Über mehrere pattern-Statements lassen sich Häufigkeiten oder Summen abbilden.

```
goptions reset=all cback=white border htitle=12pt htext=10pt;
PATTERN1 v=s c=cxeeee6ef;
PATTERN2 v=s c=cxcab0cf;
PATTERN3 v=s c=cx9c6c92;
Pattern4 v=s c=cx80408e;
pattern5 v=s c=red;
```

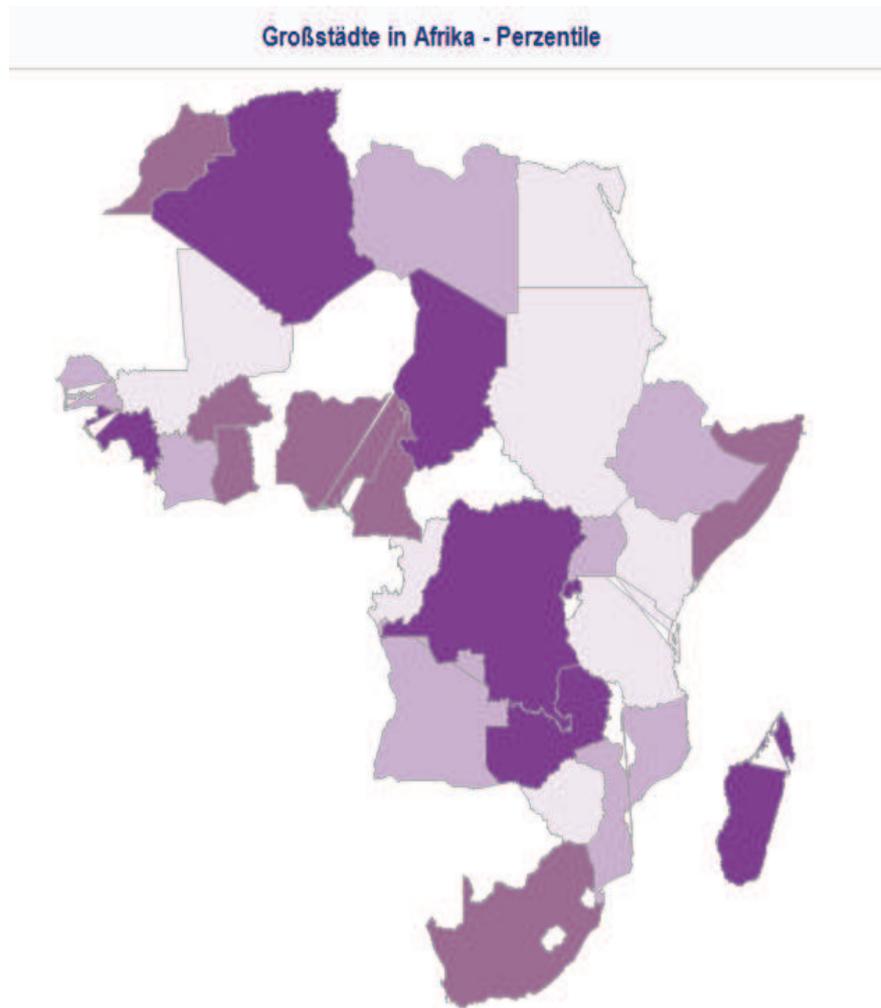


Abbildung 2: Darstellung von Häufigkeiten für den Kontinent Afrika

Die Syntax von PROC GMAP erlaubt eine flächige Darstellung der Landkartenelemente (choro-Statement). Zudem wird eine ID-Variable erwartet, die ein Land oder für Deutschland den Landkreis repräsentiert. Ein gmap-Beispielprogramm findet sich unter <http://support.sas.com/kb/24/906.html>

```
title "Großstädte in Afrika - &Variable_Response_Label.";
proc gmap data=GMAP(keep=ID CONT) map=GMAP(keep=ID CONT X Y);
  id cont id;
  choro id / nolegend outline=grayaa cempty=grgy;
run;
quit;
```

2.2 PROC SGPLOT

PROC SGPLOT gehört zum Programmpaket von SAS/BASE. Die finale Graphik ist in Abbildung 3 zu sehen.

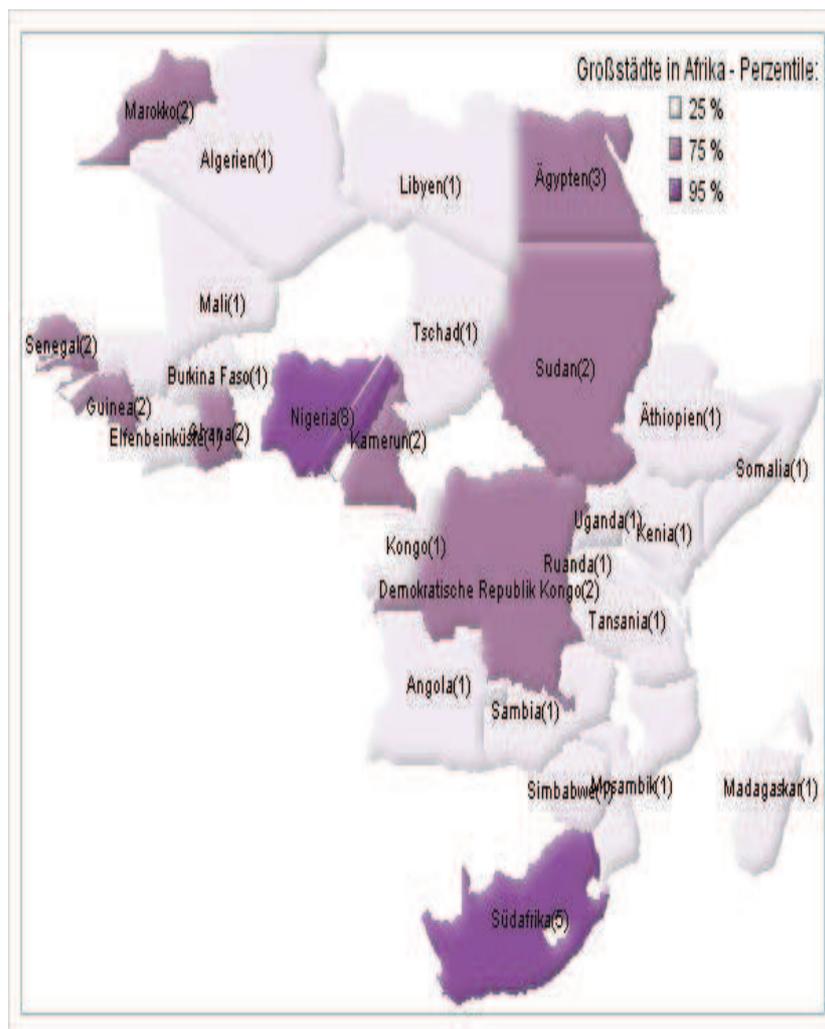


Abbildung 3: Darstellung von Häufigkeiten für den Kontinent Afrika

Die Syntax von PROC SGPLOT erlaubt eine optisch ansprechende Darstellung der Landkartenelemente (polygon-Statement). Über entsprechende benutzerdefinierte Formate lassen sich farbliche Einstufungen von Häufigkeiten hinterlegen. Texte in der Grafik lassen sich über SAS-eigene Macros (annomac/centroid) einbinden. Ggf. sind bei sich in der Grafik überlappenden Texten Verschiebungen der x- und y-Koordinaten in der Tabelle notwendig, die über das Macro centroid erstellt wird. Ein sgplot-Beispielprogramm findet sich unter <http://support.sas.com/kb/53/367.html>

```

/* Definition Tabellennamen / Response-Variable*/
%let Output_Response=WORK.Grossstaedte;
%let Map_Data=MAPSGFK.WORLD;
%let Map_Attr=&Map_Data._ATTR;
%let Map_Continent=MAPSGFK.WORLD_ATTR;

%let Variable_Response=Anzahl_GrossStaedte;
%let Variable_Response_Label=Perzentile;

/* SAS/GRAPH annotate macros kompilieren */
options nomprint fmtsearch=(WORK.Formats);
%annomac;
/* Polygon-Texte */
%centroid(&Map_Data.,DATA0,ID);

%macro Kontinent_Graphik(Cont,Bezeichnung);
title "Großstädte in &Bezeichnung.";
proc sgplot data=&Output_Response.(where=(CONT="&Cont.))
datrmap=attrmap ;
  format &Variable_Response. RSPFMT.;
  polygon
    x=x
    y=y
    id=id /
    group=&Variable_Response. attrid=maparea
    /*outline*/ /*Grenzlinien*/
    lineattrs=(color=gray99)
    fill
    fillattrs=(transparency=0.1)
    dataSkin=matte
    name='poly'
    tip=(ID &Variable_Response.);
  scatter
    x=xcen
    y=ycen /
    markerchar=IDNAME
    tip=(id);
  keylegend 'poly' /
  title="Großstädte in &Bezeichnung. -
&Variable_Response_Label.:"
  position=topright noborder location=inside down=8
  ;
  xaxis
  offsetmin=0.01
  offsetmax=0
  display=none;
  yaxis
  offsetmin=0.01
  offsetmax=0
  display=none;
run;

```

T. Rüdiger

```
%mend Kontinent_Graphik;  
%Kontinent_Graphik(AF,Afrika);
```

Leider gibt es bei der Verwendung von MAPSGFK-Daten in einigen Grafiken optische Verzerrungen. In den Grafiken zu Europa und anderen Kontinenten finden sich „künstlerische“ Verzerrungen, innerhalb der Europa-Karten speziell für Spanien, Italien, Großbritannien und Deutschland. s. Abbildung 4.

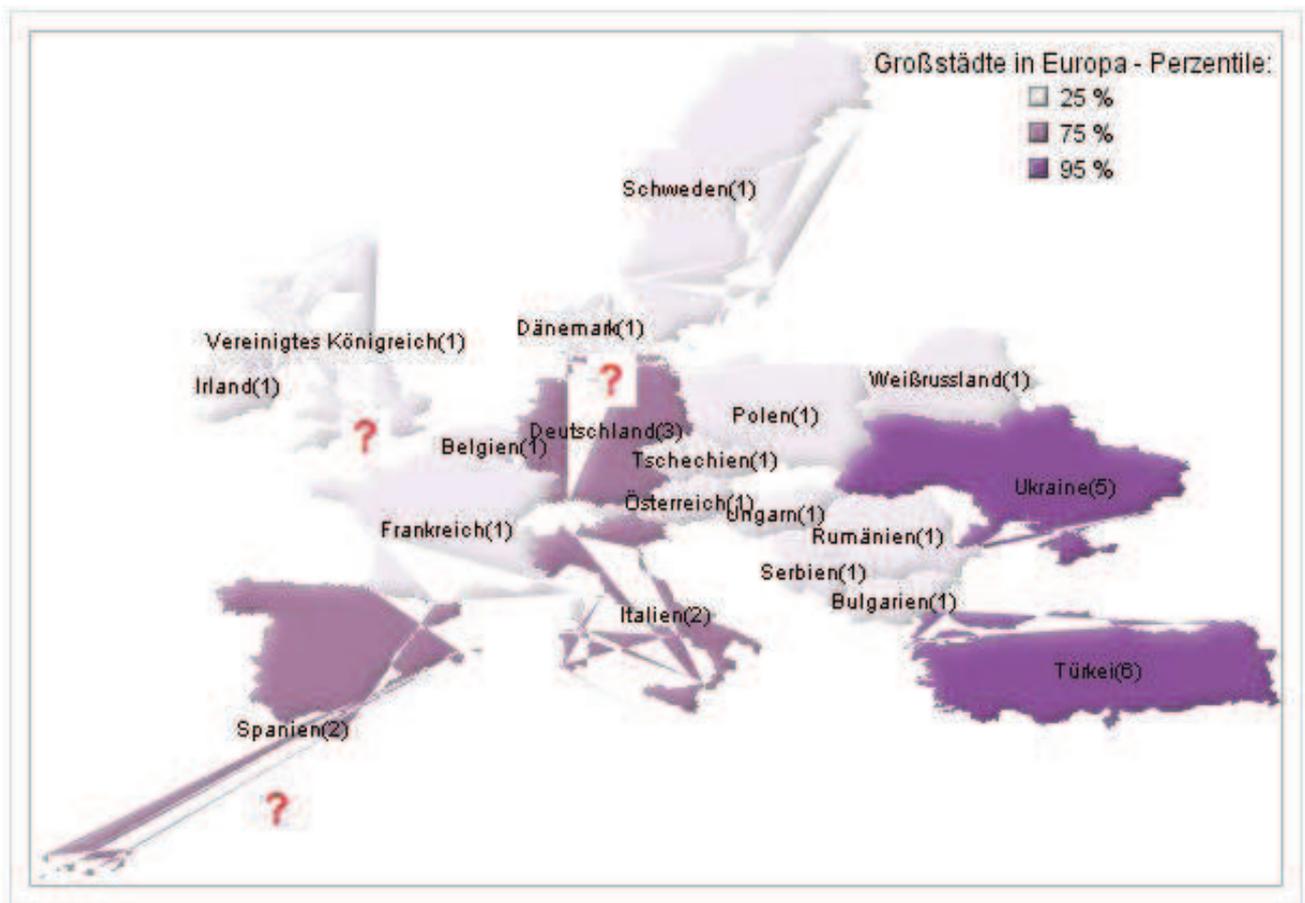


Abbildung 4: Darstellung von Häufigkeiten für den Kontinent Europa

3 Response- und Annotate-Tabellen zu den SAS-Landkarten-Prozeduren

3.1 PROC GMAP

Die Responsetabelle zu GMAP enthält ID-Daten (Länder, Landkreise), Responsedaten (z.B. Anzahl_GrossStaedte), X- und Y-Koordinaten zum Zeichnen der Ländergrenzen (Polygone).

x	y	CONT	ID
-1832.753792	546.42979418	91	61
-1832.516244	546.32344502	91	61
-1832.406261	546.40514334	91	61
-1832.413215	546.61790386	91	61
-1832.546616	546.7617275	91	61
-1832.710053	546.75438865	91	61
-1832.796733	546.64364815	91	61
-1643.143828	546.01796781	91	72
-1643.082756	545.50529495	91	72
-1642.79777	544.64808029	91	72

Abbildung 5: Responsetabelle für PROC GMAP

Die Annotate-Tabelle für GMAP benötigt zusätzlich aufwändige X- und Y-Koordinaten sowie SAS-kodierte Darstellungen z.B. zur Position (2=oberhalb Text/Symbol, 8=unterhalb, ...)

color	text	style	x	y	xsy	ysy	function	size	position
blue	M	special	2844.2391393	1227.0848931	2	2			
black	Beijing	'Albany AMT'	2844.2391393	1227.0848931	2	2			
red	M	special	208.05428658	355.147331	2	2			
black	Lagos	'Albany AMT'	208.05428658	355.147331	2	2			
red	M	special	243.86590852	1705.0061338	2	2			
black	Hamburg	'Albany AMT'	243.86590852	1705.0061338	2	2			
blue	M	special	-1809.708314	135.71470523	2	2			
black	Bogota	'Albany AMT'	-1809.708314	135.71470523	2	2			
purple	M	special	-1924.909368	1090.8474597	2	2			
black	Cary	'Albany AMT'	-1924.909368	1090.8474597	2	2			
purple	M	special	1885.8507601	860.7627566	2	2			
black	New Delhi	'Albany AMT'	1885.8507601	860.7627566	2	2			
purple	M	special	3270.5098098	+712.0672253	2	2			
black	Alice Springs	'Albany AMT'	3270.5098098	+712.0672253	2	2			
red	M	special	-2772.999417	1705.6350979	2	2			
black	Edmonton	'Albany AMT'	-2772.999417	1705.6350979	2	2			
green	M	special	-1428.556379	-1051.794735	2	2			
black	Buenos Aires	'Albany AMT'	-1428.556379	-1051.794735	2	2			
green	M	special	2270.5936289	1796.6467768	2	2			
black	Krasnoyarsk	'Albany AMT'	2270.5936289	1796.6467768	2	2			

POSITION='2'

POSITION='5'

POSITION='8'

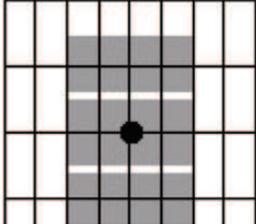


Abbildung 6: Annotate-Tabelle für PROC GMAP

3.2 PROC SGPLOT

Die Responsetabelle zu sgplot enthält ID-Daten (Länder, Landkreise), Responsedaten (z.B. Anzahl_GrossStaedte), X- und Y-Koordinaten zum Zeichnen der Ländergrenzen (Polygone) sowie X- und Y-Koordinaten (xcen, ycen im Beispiel) zum Platzieren von Texten oder Objekten).

CONT	ID	IDNAME	IDNAME_SAS	MAPSGFK	MAPS	Anzahl_GrossStaedte	X	Y	xcen	ycen
EU	DE	Deutschland(3)	Germany	GERMANY	GERMANY	3	-17665.82255	5951.1654077	-17599.69854	5794.8318724
EU	DE	Deutschland(3)	Germany	GERMANY	GERMANY	3	-17662.70434	5950.3575163	-17599.69854	5794.8318724
EU	DE	Deutschland(3)	Germany	GERMANY	GERMANY	3	-17655.10997	5950.7836504	-17599.69854	5794.8318724
EU	DE	Deutschland(3)	Germany	GERMANY	GERMANY	3	-17651.6285	5949.4756295	-17599.69854	5794.8318724
EU	DE	Deutschland(3)	Germany	GERMANY	GERMANY	3	-17647.99348	5949.4490854	-17599.69854	5794.8318724
EU	DE	Deutschland(3)	Germany	GERMANY	GERMANY	3	-17645.01238	5948.5279162	-17599.69854	5794.8318724

Abbildung 7: Responsetabelle für PROC SGPLOT

Die Annotate-Tabelle für sgplot enthält Daten wie Text- oder Füllfarben etc.

	id	textcolor	Value	fillcolor
1	maparea	black	Keine Daten	white
2	maparea	black	25 %	cxeee6ef
3	maparea	black	5 %	cxcab0cf
4	maparea	black	75 %	cx9c6c92
5	maparea	black	95 %	cx80408e
6	maparea	black	> 95 %	red

Abbildung 8: Annotate-Tabelle für PROC SGPLOT