

KSFE virtuell
13. April 2021

Tipps & Tricks für SAS User

Erzeugung einer Grafik aus öffentlich verfügbaren Dateien am Beispiel von Covid19-Daten des RKI

Ralf Sigmund, Clinical Biostatistics – BI Pharma GmbH & Co. KG

Übersicht

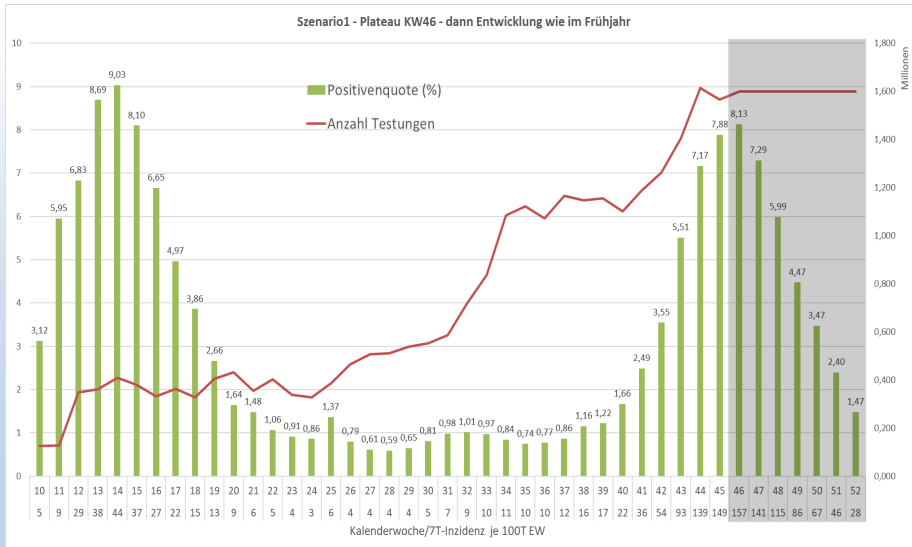


- Motivation – was wollten wir uns anschauen
- Tägliche Situationsberichte – mittwochs: Bundesweite Testhäufigkeit (Quellen)
- Das SAS Programm
 - Einlesen der Excel Datei
 - Verarbeitung der Excel Datei
 - Erstellung der Grafik mit SGPLOT
- Wie sah die Grafik von letzter Woche aus?
- Fragen/Diskussion
- Anhang

Motivation – was wollten wir uns anschauen

Rückblick in den November 2. Welle – Lockdown light

- Wie könnte die 7-Tageinzidenz um Weihnachten aussehen, wenn man annimmt jetzt geht es wieder so runter wie an Ostern 2020, also nach der ersten Welle?



Wir wissen alle, dass es anders kam ...

Das Bild war aber der Start, dass man sich die Entwicklung öfter anschauen wollte.

Ein schneller Lösungsansatz in Excel → Datei runterladen → Daten kopieren und einfügen → Datenbereich anpassen

Wie kann dies mit SAS automatisiert werden?
Input Programming (R. Minkenberg)

Anmerkung: Daten werden immer mittwochs in Tabellenform veröffentlicht

Tägliche Situationsberichte – mittwochs: Bundesweite Testhäufigkeit

Robert Koch Institut -> Coronavirus SARS-CoV-2 (COVID-19)

- URL: [RKI - Coronavirus SARS-CoV-2](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/nCoV_node.html;jsessionid=F3A910CAB7E15210EF51634F66AADCBA.internet122)

https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/nCoV_node.html;jsessionid=F3A910CAB7E15210EF51634F66AADCBA.internet122

- URL: [RKI - Coronavirus SARS-CoV-2 - Tabellen zu Testzahlen, Testkapazitäten und Probenrückstau pro Woche \(24.3.2021\)](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Testzahlen-gesamt.html;jsessionid=9BD437CBD1ABC78D03F93CEBE7937C4D.internet111?nn=13490888)

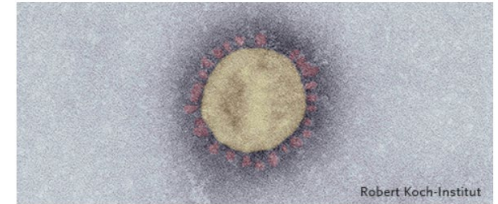
https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Testzahlen-gesamt.html;jsessionid=9BD437CBD1ABC78D03F93CEBE7937C4D.internet111?nn=13490888

- Erscheinungsdatum dd. Monat 202x » [xlsx \(23 KB, Datei ist nicht barrierefrei\)](#)

- URL zur xlsx Datei:

https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Testzahlen-gesamt.xlsx?__blob=publicationFile

COVID-19 (Coronavirus SARS-CoV-2)



Daten zum Download

- Dashboard-Daten
- Gesamtübersicht der pro Tag ans RKI übermittelten Fälle, Todesfälle und 7-Tages-Inzidenzen nach Bundesland und Landkreis
- Aktuelle Ergebnisse des Nowcasting und der R-Schätzung
- Fälle nach Altersgruppe und Meldewoche (Tabelle wird dienstags aktualisiert)
- Fälle nach Meldewoche und Geschlecht sowie Anteile mit für COVID-19 relevanten Symptomen, Anteile Hospitalisierter und Verstorbener (Tabelle wird dienstags aktualisiert)
- Fälle, die einem Ausbruch zugeordnet werden, nach Meldewoche und Infektionsumfeld (Tabelle wird dienstags aktualisiert)
- Tabellen zu Testzahlen, Testkapazitäten und Probenrückstau pro Woche (Tabelle wird mittwochs aktualisiert)
- Todesfälle nach Sterbedatum (Tabelle wird freitags aktualisiert)
- SurvStat - individuelle Abfrage von Meldedaten

SAS - Programm

SAS Programm – Einlesen der Excel Datei

```
filename getdata url 'https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/  
Daten/Testzahlen-gesamt.xlsx?__blob=publicationFile'  
connect proxy=&proxy;  
  
data _NULL_;  
  n = -1;  
  infile getdata recfm=s nbyte=n length=len lrecl=32767;  
  input;  
  file 'u:\rkiTestzahlen.xlsx' recfm=n lrecl=32767;  
  put _infile_ $varying32767. len;  
run;  
  
ods html close; ods listing;  
libname inxls excel 'u:\rkiTestzahlen.xlsx' getnames=yes mixed=yes;
```

- URL Access Method ermöglicht Remote-Zugriff (&proxy individuell)
- Mit dem record-format 'S' werden Daten im 'stream record format' eingelesen
- Excel libname (inxls) bewirkt, dass die Excel Datei ganz normal prozessiert werden kann

SAS Programm – Verarbeitung der Excel Datei

```
data figuredat0(rename=(POSITIVENANTEIL____=positivenquote));
  set inxls.'1_Testzahlerfassung$`n
            (dbSasType=("Positivenanteil (%) "n=numeric));
  if _n_ eq 1 then do;
    kalenderwoche = '<=10/2020';
    POSITIVENANTEIL____ = positiv_getestet / anzahl_testungen * 100;
  end;
  if substr(kalenderwoche,1,1) eq '*' then call symput('foot1', kalenderwoche);
  anzahl_testungen = anzahl_testungen / 1e6;
  if indexc(kalenderwoche, '/') eq 0 then delete;
  format POSITIVENANTEIL____ 4.1;
run;
```

- Mit dem SAS Name literal „...\$“n können Sonderzeichen verarbeitet werden
- dbSasType=(... =numeric): Variable-Type soll ‚numerisch‘ sein
- Fallstricke: Variablen werden plötzlich umbenannt (hier Quote/Anteil)

SAS Programm – Erstellung der Grafik mit SGPLOT

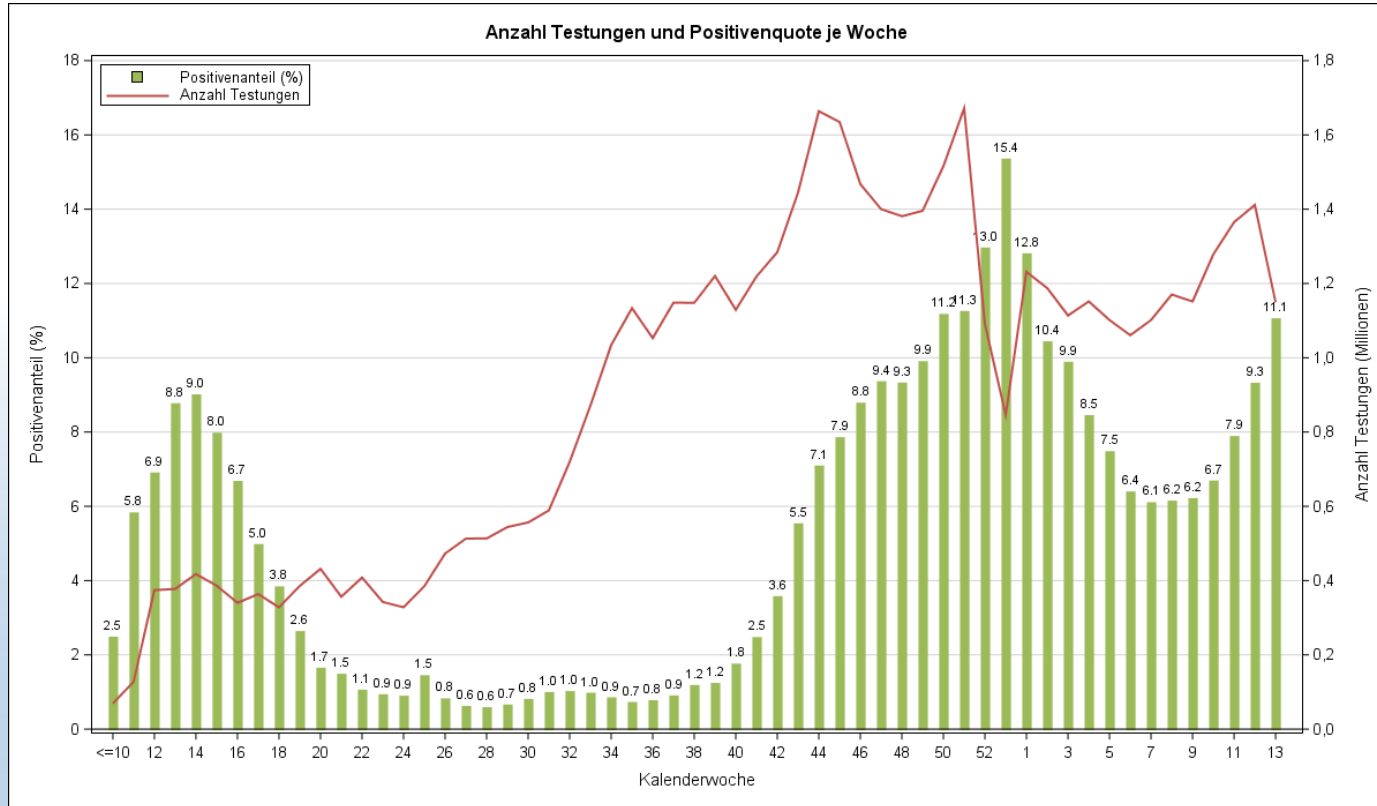
```
ods graphics / reset imagename='rkiTestzahlen' imagefmt=png width=1200px height=700px;

proc sgplot data= figuredat0 tmpout='u:\temp1.sas';
  title 'Anzahl Testungen und Positivenquote je Woche';
  *footnote "&foot1.";

  xaxis type=discrete valuesformat=$kw. fitpolicy=thin;
  yaxis type=linear min=0 max=18 values=(0 to 18 by 2) grid gridattrs=(color=lightgrey);
  y2axis type=linear min=0 max=1.8 values=(0 to 1.8 by 0.2) valuesformat=numx3.1
  label='Anzahl Testungen (Millionen)';
  vbarparm category=kalenderwoche response=positivenquote
    / fillattrs=(color=cx9bbb59) filltype=solid nooutline barwidth=0.4
    datalabel datalabelfitpolicy=none;
  series x=kalenderwoche y=anzahl_testungen
    / lineattrs=(color=cxc0504d pattern=solid thickness=2pt) y2axis;
  keylegend / location=inside position=topleft down=2;
run;
quit;
```

- vbarparm: Säulendiagramm der Positivquoten über die KWen
 - datalabelfitpolicy=none: per default prüft SAS ob die Label überlappen. Wenn ja, dann werden die Label um -90° gedreht, was man unterdrücken kann
- series: Linienplot der Anzahl Testungen über die KWen
- color=CXrrggbb (RGB Anteile im Hexadezimalsystem - ff0000=Rot)

Wie sah die Grafik von letzter Woche aus? (erstellt am 08.04.2021)



Diskussion

Fragen?

Anhang 1: Format für KW / Template - SGRENDER / RGB Beispiel

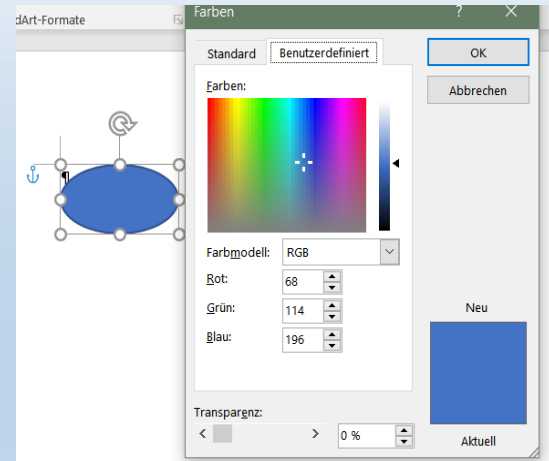
Format für Kalenderwoche

```
data _fmt_kw(drop=kalenderwoche);      set figuredat0(keep=kalenderwoche);
  length fmtname $10;
  retain fmtname '$kw';
  start = kalenderwoche;
  label = scan(kalenderwoche,1,'/') !! ifc(indexc(kalenderwoche,'*') gt 0, '*', '');
run;
proc format cntlin=_fmt_kw;
run;
```

Das template ‚temp1.sas‘ kann via
SGRENDER benutzt werden

```
%include "U:\temp1.sas";
PROC SGRENDER data=figuredat0 template=sgplot;
RUN;
```

RGB – Ein schneller Blick
z.Bsp. in ein WordArt



Anhang 2: Das gesamte Programm (1 von 2)

```
%LET proxy='http://<your proxy server>';

filename getdata url 'https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Testzahlen-gesamt.xlsx?__blob=publicationFile'
    connect proxy=&proxy;

data _NULL_;
    n = -1;
    infile getdata recfm=s nbyte=n length=len lrecl=32767;
    input;
    file 'u:\rkiTestzahlen.xlsx' recfm=n lrecl=32767;
    put _infile_ $varying32767. len;
run;

ods html close; ods listing;
libname inxls excel 'u:\rkiTestzahlen.xlsx' getnames=yes mixed=yes;

data figuredat0(rename=(POSITIVENANTEIL____=positivenquote));
    set inxls.'1_Testzahlerfassung$'n(dbSasType=("Positivenanteil (%) "n=numeric));
    if _n_ eq 1 then do;
        kalenderwoche = '<=10/2020';
        POSITIVENANTEIL____ = positiv_getestet / anzahl_testungen * 100;
    end;
    if substr(kalenderwoche,1,1) eq '*' then call symput('foot1', kalenderwoche); *might be needed;
    anzahl_testungen = anzahl_testungen / 1e6;
    if index(kalenderwoche, '/') eq 0 then delete;
    format POSITIVENANTEIL____ 4.1;
run;
```

Anhang 2: Das gesamte Programm (2 von 2)

```
data _fmt_kw(drop=kalenderwoche); * define format for Kalenderwoche;
  set figuredat0(keep=kalenderwoche);
  length fmtname $10;
  retain fmtname '$kw';
  start = kalenderwoche;
  label = scan(kalenderwoche,1,'/') !! ifc(indexc(kalenderwoche,'*') gt 0, '*', '');
run;

proc format cntlin=_fmt_kw;
run;

libname inxls clear; *close the inxls connection;

ods graphics / reset imagename='rkiTestzahlen' imagefmt=png width=1200px height=700px;

proc sgplot data=figuredat0 tplout='u:\templ.sas';
  title 'Anzahl Testungen und Positivenquote je Woche';
  *footnote "&foot1.";
  xaxis type=discrete valuesformat=$kw. fitpolicy=thin;
  yaxis type=linear min=0 max=18 values=(0 to 18 by 2) grid gridattrs=(color=lightgrey);
  y2axis type=linear min=0 max=1.8 values=(0 to 1.8 by 0.2) valuesformat=numx3.1 label='Anzahl Testungen (Millionen)';
  vbarparm category=kalenderwoche response=positivenquote / fillattrs=(color=cx9bbb59) filltype=solid nooutline barwidth=0.4
  datalabel datalabelfitpolicy=none;
  series x=kalenderwoche y=anzahl_testungen / lineattrs=(color=cxc0504d pattern=solid thickness=2pt) y2axis;
  keylegend / location=inside position=topleft down=2;
run;
quit;
```

Anhang 3: Weitere Quellen

- [024-31: De-Mystifying the SAS® LIBNAME Engine in Microsoft Excel: A Practical Guide](https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/02431.pdf)
<https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/02431.pdf>
- [020-31: So, Your Data are in Excel! \(sas.com\)](https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/02031.pdf)
<https://support.sas.com/resources/papers/proceedings/proceedings/sugi31/02031.pdf>
- [SAS Help Center: LIBNAME Statement: Access and Excel Engines](https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=acpcref&docsetTarget=n1wvmggexroxgyn17rp61jml3cvn.htm&locale=en)
https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=acpcref&docsetTarget=n1wvmggexroxgyn17rp61jml3cvn.htm&locale=en