



R
—
H

**Rechnungshof
Österreich**

Unabhängig und objektiv für Sie.

Datenanalyse

- Organisation
- Prüfprozess
- Fallbeispiele

gartner@rechnungshof.gv.at

glinitzer@rechnungshof.gv.at

steinbichler@rechnungshof.gv.at

Datenanalyse

- Seit 2013 Ausbau der Datenanalyse im RH
- Aktuell 12 Datenanalyse-Experten
- 15 MA machen Onlinekurse auf datacamp.com im Bereich der Datenanalyse (R, Python, MySQL, ...)
- Fokussierung auf folgende Werkzeuge
 - R-Statistiksoftware
 - QGIS für GIS-Analysen
 - Shiny, D3, Html, CSS, ...
 - Excel
 - Gitlab für die Zusammenarbeit und Versionierung der Programmierprojekte

Organisation

- zentrale Weiterentwicklung der Datenanalyse ist ein Aufgabenbereich der IT-Abteilung
- In der IT-Abteilung: 5 MA mit Expertenwissen im Bereich der Datenanalyse
- In den Prüfabteilungen gibt es 7 Personen mit Expertenwissen im Bereich der Datenanalyse
- Weitere 20 Personen mit vertieften Know-How im Bereich der Datenanalyse, diese MA sind quasi die „Schnittstellen“ zwischen den Prüfabteilungen und den Datenanalysten

Learnings

Gute Datenanalysen brauchen Zeit

Datenanalysen sind bereits bei den Prüfungsgesprächen sehr hilfreich – oftmals ein Überraschungseffekt der geprüften Stelle *„Was der RH bereits alles weiß“* – allein aufgrund der Datenauswertungen können *„zielgerichtete“* Fragen bereits bei den ersten Prüfungsgesprächen gestellt werden!

Der RH hilft den geprüften Stellen beim
„Heben ihrer Datenschätze“

Qualitätssicherung, Sicherstellung des 4-Augen Prinzips

R-Markdown Beispiel – ein Vorteil von vielen von R gegenüber Excel

2.4 Daten kombinieren

```
epi <- bind_rows(epi_timeline, epi_epikurve)
```

2.5 Daten für Vergleich zusammenfassen

```
erkrankte_neu <- epi %>%
  group_by(publikationstag, gruppe) %>%
  filter(datum == max(datum)) %>%
  select(publikationstag, erkrankte_neu = erkrankte, gruppe)

# Summe aller Fälle bis zu diesem Publikationstag und Summe bis zum Vortag
epi_summe <- epi %>%
  group_by(gruppe, publikationstag) %>%
  summarise(erkrankte_summe = sum(erkrankte)) %>%
  mutate(erkrankte_summe_vortag = lag(erkrankte_summe)) %>%
  left_join(erkrankte_neu, by = c("publikationstag", "gruppe"))
```



Prüfprozess

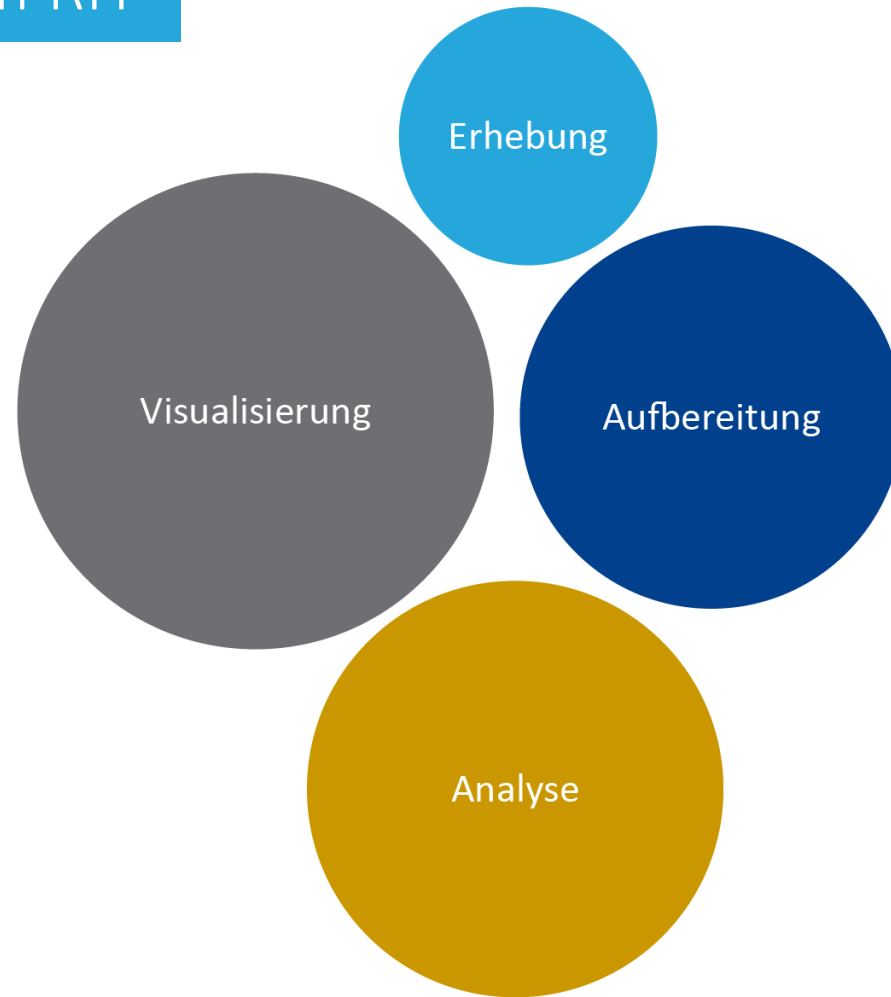
R
—
H



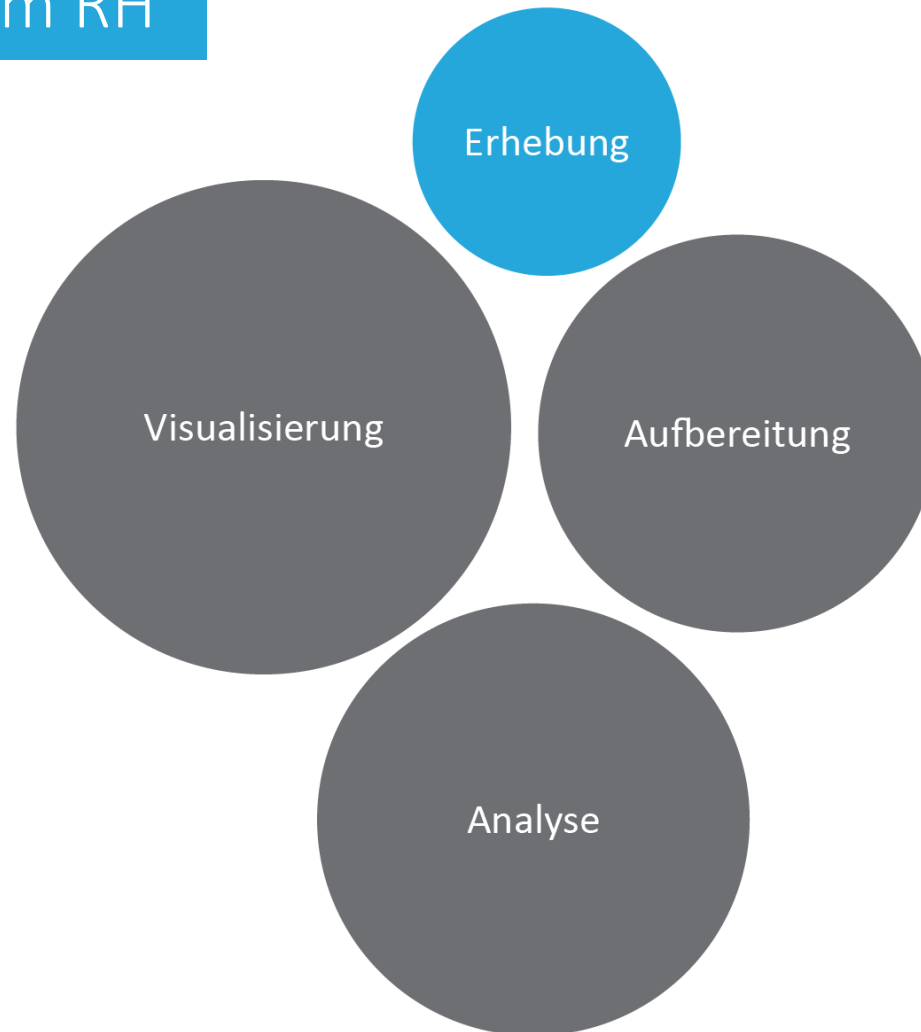
Rechnungshof
Österreich

Unabhängig und objektiv für Sie.

Datenanalyse im RH



Datenanalyse im RH




Datenerhebung

- Fragestellungen einer Prüfung in quantitativ überprüfbare Hypothesen zu übersetzen
 - *„Wie viele Impfstoffdosen wäre zur Verfügung gestanden, wenn bei allen Verträgen mit Impfstoffherstellern gemäß Bevölkerungsschlüssel bestellt worden wäre?“*
- Prüfen der Datenanforderungen für die Beantwortung der Fragestellung
 - Anforderung der Lieferzeittabellen, verabreichte Impfdosen, Rahmenbedingungen

Datenerhebung

- Abfrage der Datenbank einer geprüften Stelle
- Unterstützung/Schulung bei Erstellung/Konzeption von Umfragen
- Scraping öffentlich zugänglicher Informationen
- Spezielle Abfragen via Statcube / Open Government Data

Datenerhebung

 **statistik** / **coronaDAT** Public


Data-Repository (Quelle: <https://info.gesundheitsministerium.at> und <https://covid19-dashboard.ages.at/>)

☆ 37 stars 🍴 6 forks

☆ Star 🔔 Notifications

[<> Code](#) [🔍 Issues](#) [🔗 Pull requests](#) [▶ Actions](#) ⋮

🔗 master Go to file

 **bernhard-da** data update (only raw dat... ⋮ 4 hours ago 🕒 8,407

[View code](#)

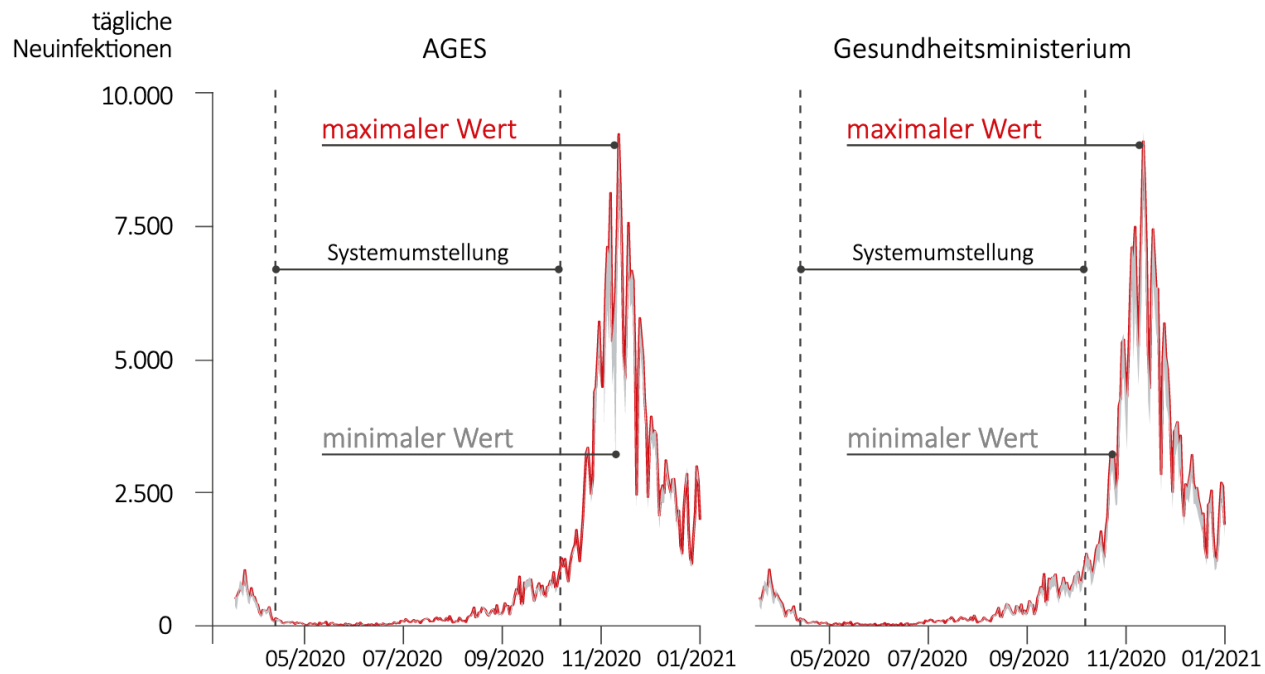
Datenerhebung

Österreich

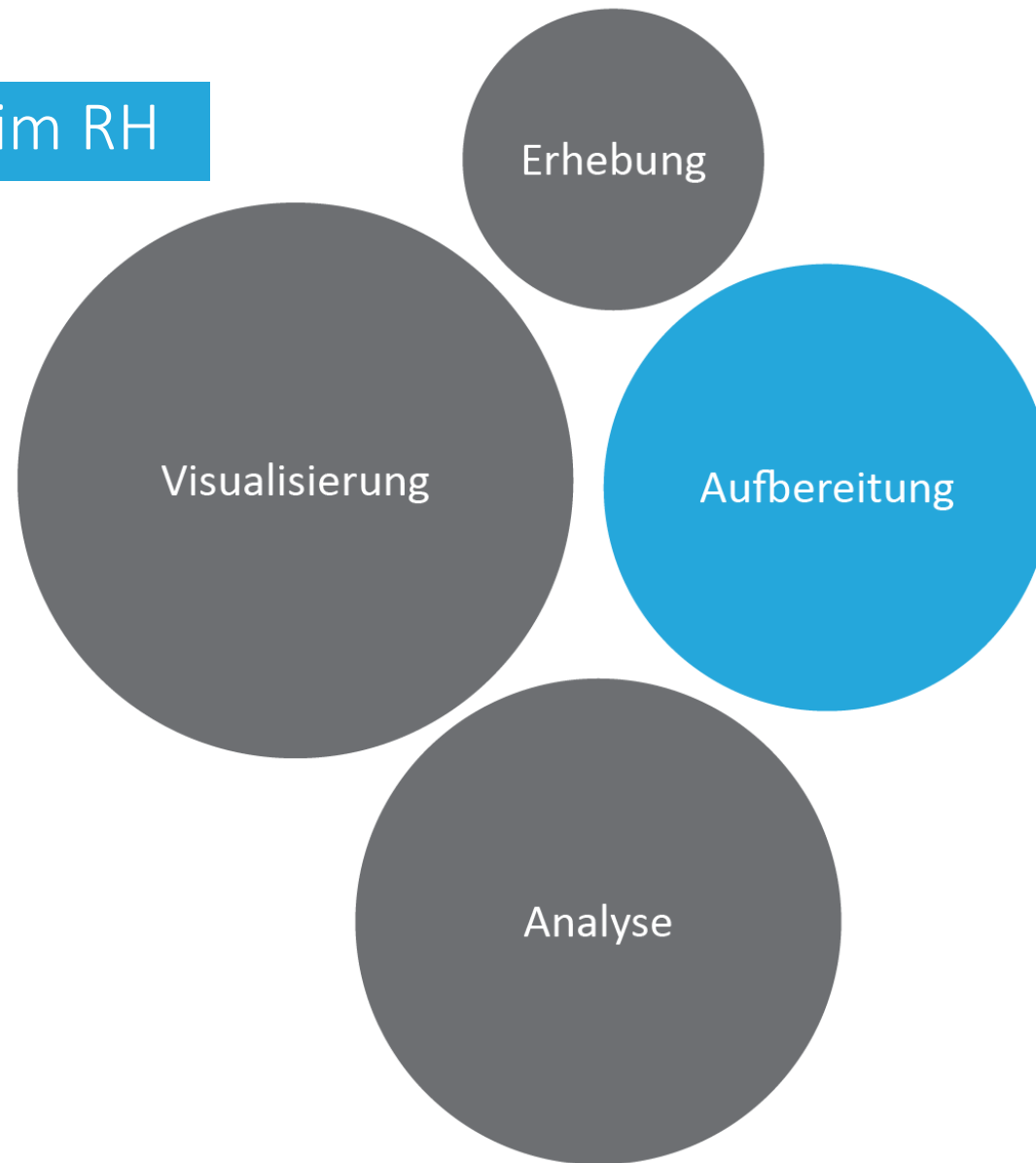
EPIDEMIOLOGISCHE KURVE



Quellen: Repositories; Darstellung: RH



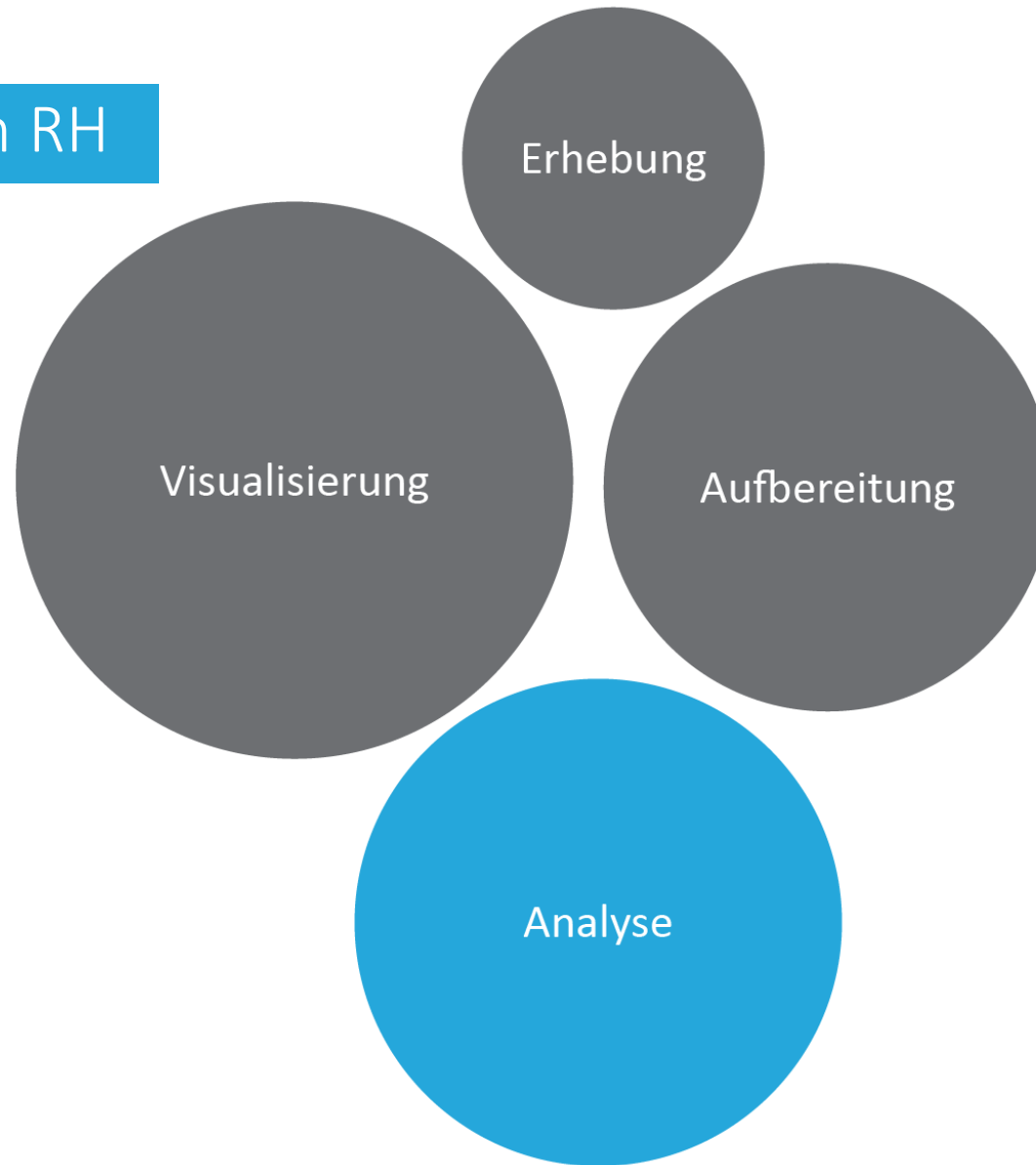
Datenanalyse im RH



Datenaufbereitung

- Aufbereitung großer Datenmengen
 - Excel \neq Rstudio
- Verknüpfung von Datensätzen
- Unstrukturierte Daten zu strukturiertem Datensatz verwandeln

Datenanalyse im RH



Datenanalyse

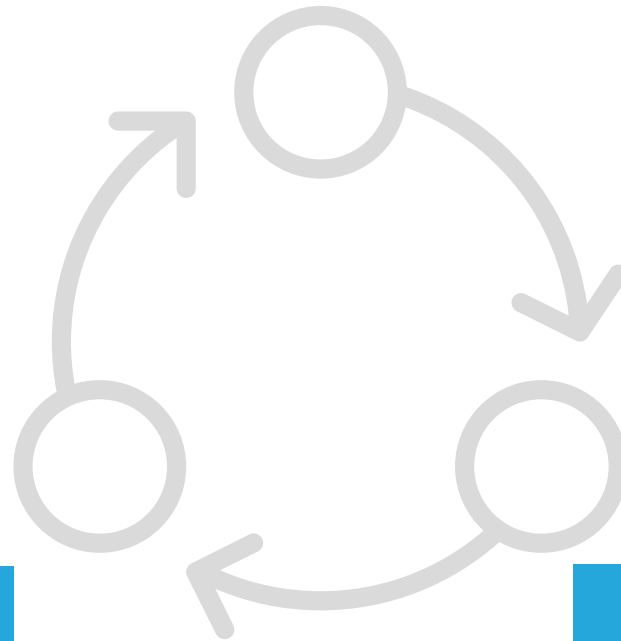
- Beantworten der Fragestellungen / Prüfen der Hypothesen
- Verteilung von Variablen
- Prüfen von statistischen Zusammenhängen

Prüfprozess

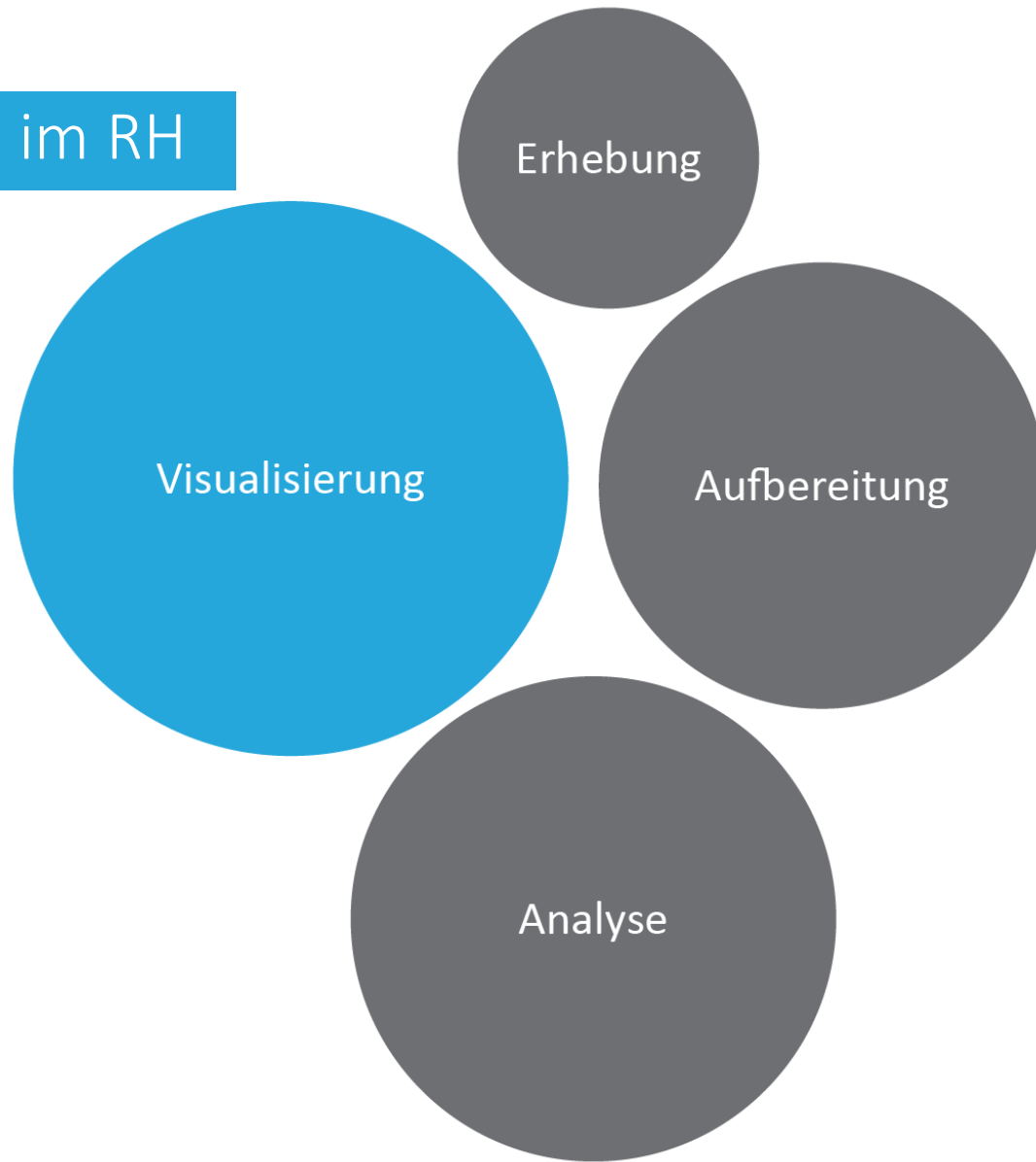
Datenerhebung und
-bereinigung

Datenanalyse

Datenvisualisierung



Datenanalyse im RH



Datenvisualisierung

ÜBERSICHT

BUND

BUNDESLÄNDER

TABELLE

Auf dieser Seite erhalten Sie eine interaktive Übersicht über die finanziellen Hilfsmaßnahmen, die sowohl der Bund als auch die Bundesländer im Zuge der COVID-19-Pandemie getätigt haben. Der Rechnungshof erfasste quartalsweise, welche Hilfsmaßnahmen seit 1. März 2020 implementiert wurden. Aktuell verfügt der Rechnungshof über Daten bis Ende Juni 2021.

Ersichtlich sind: das geplante Gesamtvolumen sowie die tatsächlich getätigten finanziellen Hilfsleistungen. Unter „finanziellen Hilfsmaßnahmen“ sind geldwerte Leistungen zu verstehen – wie Zuschüsse, Sachleistungen, Haftungen, Garantien und Darlehen. Ebenfalls erfasst wurden Mindereinnahmen der öffentlichen Hand – etwa Einnahmenverzichtse. Unter „keine Beschränkung“ finden Sie Leistungen, die keinem spezifischen Bereich zuordenbar sind.

Alle Hilfsmaßnahmen zusammengefasst, ergab sich bis 30. Juni 2021 folgendes Bild:



GEWÄHRTE HILFSLEISTUNGEN
34,481 Mrd. EUR
Juni 2021



GEPLANTES GESAMTVOLUMEN
75,342 Mrd. EUR
Juni 2021



PROZENT DER GEWÄHRTEN HILFSLEISTUNGEN
23,4%
Juni 2021



ANZAHL HILFSMASSNAHMEN
528
Juni 2021

[Link zur Grafik](#)

Prüfprozess

Vorbereitung

Einschau

PE-Erstellung

Veröffentlichung





R
—
H



**Rechnungshof
Österreich**

Unabhängig und objektiv für Sie.

Fallbeispiel: COVID-19-Kurzarbeit

COVID-19-Kurzarbeit

- Bericht des Rechnungshofs: COVID-19-Kurzarbeit
- Fragestellung der Prüfung: Darstellung und Beurteilung der Konzeption, Inanspruchnahme, Kosten und Abwicklung der COVID-19-Kurzarbeitsbeihilfe
- Prüfzeitraum: März 2020 bis März 2021

COVID-19-Kurzarbeit

- Kurzarbeitsbeihilfe: 7,849 Mrd. EUR (März 2020 – März 2021)
- Auszahlung in 3 Phasen:
 - Phase 1 (1. März 2020 – 31. Mai 2020)
 - Phase 2 (1. Juni 2020 – 30. September 2020)
 - Phase 3 (1. Oktober 2020 – 31. März 2021)
- Konzeption von Phase 1 erfolgte unter großem Zeitdruck
- Überföderung in Phase 1: rund 500 Mio. EUR

Unterstützung durch Datenanalyse-Team

- Wie stark wäre ausbezahlte Kurzarbeitsbeihilfe gesunken, wenn Phase 1 mit dem Modell aus Phase 2 berechnet worden wäre?



- Berechnung von Kurzarbeitsbeihilfe in Phase 1 mit Modell von Phase 2

Vorgehensweise

1. Daten von AMS zu allen Personen, die im Prüfzeitraum in Kurzarbeit waren, angefordert
2. Vorgaben für Phase 2-Modell in Statistiksoftware R nachbilden
3. Nachgebildetes Modell mit Daten aus Phase 2 validieren
4. Nachgebildetes Modell auf Phase 1 anwenden und Unterschied der Kurzarbeitsbeihilfe berechnen

Schritt 1: Datengrundlage

- Daten: AMS
- Jeder Kurzarbeitsbezieher für jedes Monat
- 5.957.386 Zeilen (Personen je Monat)
- 49 Spalten (Information zur Normalarbeitszeit, Ausfallstunden, Höhe der Kurzarbeitsbeihilfe...)
- 2.5 GB

Schritt 1: Datengrundlage

	svnr	kua_phase	bruttoentgelt	kurzarbeitshilfe	lehrling	abrechnungszeitraum_von	abrechnungszeitraum_bis	naz_abrechnungszeitraum	naz_kv	ausfallstunden
1	10000*****	1	2136.39	2485.08	N	2020-04-01	2020-04-30	176.0	40.0	156.00
2	10000*****	1	2179.99	1083.29	N	2020-05-01	2020-05-31	168.0	40.0	66.50
3	10000*****	1	2575.20	1490.40	N	2020-05-01	2020-05-31	133.0	35.0	67.50
4	10000*****	1	2575.20	2450.88	N	2020-05-01	2020-05-31	147.0	35.0	111.00
5	10000*****	1	2575.20	1700.16	N	2020-06-01	2020-06-30	141.0	35.0	77.00
6	10000*****	1	771.81	1020.10	N	2020-04-01	2020-04-30	33.6	8.0	33.60
7	10000*****	1	771.81	922.94	N	2020-05-01	2020-05-31	30.4	8.0	30.40
8	10000*****	1	771.81	631.49	N	2020-06-01	2020-06-30	32.0	8.0	20.80
9	10000*****	1	2000.68	2290.18	N	2020-04-01	2020-04-30	129.9	30.0	115.90
10	10000*****	1	6847.00	3194.17	N	2020-05-01	2020-05-31	146.3	38.5	78.50
11	10000*****	1	6847.00	5406.89	N	2020-04-01	2020-04-30	161.7	38.5	132.88
12	10000*****	1	3367.40	2046.24	N	2020-06-01	2020-06-30	159.0	40.0	84.00
13	10000*****	1	3367.40	2996.28	N	2020-05-01	2020-05-31	151.0	40.0	123.00
14	10000*****	1	3367.40	2825.76	N	2020-04-01	2020-04-30	168.0	40.0	116.00
15	10000*****	1	3465.30	2087.45	N	2020-03-16	2020-03-31	90.5	38.0	79.10

Wie hoch mit dem Modell
aus Phase 2?

Schritt 2: Model für Phase 2

- Datenmenge: nicht in MS Excel bearbeitbar
- Statistik-Software R: für größere Datenmengen optimiert
- Herausforderung:
 - Berechnung des Phase 2-Modells benötigt Mindestbruttoentgelt, das in den Daten des AMS nicht vorhanden sind
 - Mit eigener Tabelle zum Mindestbruttoentgelt verknüpfen

Schritt 2: Model für Phase 2

- AMS-Daten

	svnr	bruttoentgelt
1	10000*****	2136.39
2	10000*****	2179.99
3	10000*****	2575.20
4	10000*****	2575.20
5	10000*****	2575.20
6	10000*****	771.81
7	10000*****	771.81
8	10000*****	771.81
9	10000*****	2000.68
10	10000*****	6847.00

- Umrechnungstabelle

	ab	bis	nettoersatz	mindestbrutto
81	750.00	755.00	90	677.25
82	755.00	760.00	90	681.75
83	760.00	765.00	90	686.25
84	765.00	770.00	90	690.75
85	770.00	775.00	90	695.25
86	775.00	780.00	90	699.75
87	780.00	785.00	90	704.25
88	785.00	790.00	90	708.75
89	790.00	795.00	90	713.26
90	795.00	800.00	90	717.76

Für alle 5.957.386 Zeilen!

Schritt 2: Model für Phase 2

- Für jede Person im Datensatz bestimmen, in welches Intervall das Bruttoentgelt fällt und dementsprechend das Mindestbruttoentgelt zuordnen
- Verknüpfung von mehreren Datensätzen in R schnell umsetzbar
- Problemstellung, die auch für andere Gebahrungsprüfungen relevant ist

Schritt 2: Model für Phase 2

- Mit Information zum Mindestbruttoentgelt können nun die weiteren Berechnungsschritte des Phase 2-Modells berechnet werden
- Beispielfile des AMS als Grundlage

Spalte	vor Beginn der Kurzarbeit				im Monat		im Abrechnungszeitraum						Kostensatz für Lohnnebenkosten				
	1	2	3		4		5	6	7	8	9 *)	10	11	12 *)	13 *)	14	
		**)					Querprüfung		5 - 6	1 / 4 * 6	2 / 4 * 5 - 8	8 + 9	9 * 27%	(1/4*5-8)/6 + 30%	(1/4*5 - 10)* 39%	9 + 11 + 12 +13	
Personen- daten	Bruttoentgelt vor KUA	Mindest- Bruttoentgelt (im Monat)	wtl. Normal- arbeitszeit lt. KV/AV	Lehrling	NAZ-Stunden im Monat	Abrechnungs- zeitraum von	Abrechnungs- zeitraum bis	NAZ Stunden im Abrechnungs- zeitraum	in Abrechnung angegebene Stunden	KUA-bedingte Ausfallstunden	"Arbeits- entgelt"	Kurzarbeits- unterstützung	Brutto- Einkommen	DG SV-Beiträge + DG-Abgaben (27%)	anteilige Sonderzahlung (1/6 + 30%)	höhere SV-Beiträge (39%)	Kurzarbeits- beihilfe
N.N.	1.510,00	1.314,89	40,00	N	176,00	Mo, 01.06.2020	Di, 30.06.2020	176,00	105,00	71,00	900,85	414,04	1.314,89	111,79	131,98	76,09	733,91
N.N.	2.150,00	1.665,69	40,00	N	176,00	Mo, 01.06.2020	Di, 30.06.2020	168,00	75,00	93,00	916,19	673,79	1.589,98	181,92	246,15	180,29	1.282,16
N.N.	5.370,00	4.041,29	40,00	N	176,00	Mo, 01.06.2020	Di, 30.06.2020	176,00	135,00	41,00	4.119,03	0,00	4.119,03	0,00	271,04	487,88	758,92
N.N. **)	6.150,00	4.041,29	40,00	N	176,00	Mo, 01.06.2020	Di, 30.06.2020	168,00	100,00	68,00	3.494,32	363,28	3.857,60	98,09	353,51	494,64	1.309,52
Daten laut Abrechnungsliste												363,28	98,08	353,51	494,64	1.309,51	
														1.163,50	518,20		

Schritt 2: Model für Phase 2

```
80 data %>%
81   mutate(mindestbruttoentgelt = case_when(
82     bruttoentgelt < 350 ~ 0,
83     bruttoentgelt <= 460.66 & lehrling == "N" ~ 0,
84     bruttoentgelt >= hbmng ~ max_mindestbrutto,
85     bruttoentgelt >= 350 & bruttoentgelt < hbmng ~ round2(mindestbrutto,2)),
86   abrechnungszeitraum_bis_monat = ceiling_date(abrechnungszeitraum_bis, unit = "month") - 1,
87   abrechnungszeitraum_von_monat = floor_date(abrechnungszeitraum_von, unit = "month"),
88   naz_tage_monat = bizdays(from = abrechnungszeitraum_von_monat, to = abrechnungszeitraum_bis_monat, cal = "austria") + 1,
89   naz_tage_abrechnungszeitraum = bizdays(from = abrechnungszeitraum_von, to = abrechnungszeitraum_bis, cal = "austria") + 1,
90   naz_stunden_monat = round2(naz_tage_monat * (naz_kv / 5), 2),
91   naz_stunden_abrechnungszeitraum = naz_abrechnungszeitraum,
92   arbeitsentgelt = round2(bruttoentgelt / naz_stunden_monat * (naz_stunden_abrechnungszeitraum - ausfallstunden),2),
93   ku_unterstützung = mindestbruttoentgelt * (naz_stunden_abrechnungszeitraum / naz_stunden_monat) - arbeitsentgelt,
94   ku_unterstützung = case_when(
95     ku_unterstützung <= 0 ~ 0,
96     TRUE ~ round2(ku_unterstützung, 2)),
97   brutto_einkommen = ku_unterstützung + arbeitsentgelt,
98   dgsv = (ku_unterstützung * 1.27) - ku_unterstützung,
99   anteilige_sonderzahlung_help = case_when(
100     mindestbruttoentgelt == 0 ~ 0,
101     bruttoentgelt < hbmng & mindestbruttoentgelt > 0 ~ (bruttoentgelt / naz_stunden_monat * naz_stunden_abrechnungszeitraum) - arbeitsentgelt,
102     bruttoentgelt >= hbmng & mindestbruttoentgelt > 0 ~ (hbmng / naz_stunden_monat * naz_stunden_abrechnungszeitraum) - arbeitsentgelt),
103   anteilige_sonderzahlung = case_when(
104     anteilige_sonderzahlung_help <= 0 ~ 0,
105     anteilige_sonderzahlung_help >= hbmng ~ (hbmng / 6) * 1.3,
106     anteilige_sonderzahlung_help > 0 & anteilige_sonderzahlung_help < hbmng ~ (anteilige_sonderzahlung_help/6) * 1.3),
107   anteilige_sonderzahlung = case_when(
108     mindestbruttoentgelt == 0 ~ 0,
109     anteilige_sonderzahlung >= max_sonderzahlung ~ max_sonderzahlung,
110     TRUE ~ anteilige_sonderzahlung),
111   höhere_sv_beiträge_help = case_when(
112     mindestbruttoentgelt == 0 ~ 0,
113     (bruttoentgelt < hbmng) & (mindestbruttoentgelt > 0) ~ (bruttoentgelt / naz_stunden_monat * naz_stunden_abrechnungszeitraum - brutto_einkommen),
114     (bruttoentgelt >= hbmng) & (mindestbruttoentgelt > 0) ~ (hbmng / naz_stunden_monat * naz_stunden_abrechnungszeitraum - brutto_einkommen)),
115   höhere_sv_beiträge = case_when(
116     höhere_sv_beiträge_help < (hbmng - max_mindestbrutto) ~ höhere_sv_beiträge_help * .39,
117     höhere_sv_beiträge_help >= (hbmng - max_mindestbrutto) ~ (hbmng - max_mindestbrutto) * .39),
118   höhere_sv_beiträge = case_when(
119     höhere_sv_beiträge < 0 ~ 0,
120     TRUE ~ höhere_sv_beiträge),
121   kurzarbeitshilfe_berechnet = ku_unterstützung + dgsv + anteilige_sonderzahlung + höhere_sv_beiträge,
122   kurzarbeitshilfe_berechnet = round2(kurzarbeitshilfe_berechnet, 2) %>%
123   select(-c(mindestbrutto, abrechnungszeitraum_von_monat, abrechnungszeitraum_bis_monat, anteilige_sonderzahlung_help, höhere_sv_beiträge_help))
```


Schritt 2: Model für Phase 2

- In R ist jeder Berechnungsschritt vollständig dokumentiert und nachvollziehbar
- Berechnungsschritte können einfach angepasst werden und das Modell neu berechnet werden

Schritt 3: Modell validieren

- Kann mit nachgebautem Model Phase 2 korrekt reproduziert werden?
 - Warum gibt es Abweichungen? Welche Abweichungen lassen sich nicht erklären?
 - Sind die Vorgaben der geprüften Stellen nicht nachvollziehbar?
 - Enger Austausch mit dem Prüfteam

Schritt 4: Phase 1 neu berechnen

- Für jede Person aus Phase 1 wird das (validierte) Modell aus Phase 2 angewandt
 - In R sehr einfach anwendbar
- Für Phase 1: Differenz zwischen tatsächlicher Kurzarbeitsbeihilfe und mit hypothetischer Kurzarbeitsbeihilfe (Modell von Phase 2)
- Ergebnis: 500 Mio. EUR Unterschied

Abschluss

- Visualisierung von zentralen Prüfungsergebnissen auf der Webpage des Rechnungshofes