Wissenschaftliches Rechnen 405.100
R: Grundlegendes, Appetizer, learning by doing

Univ.-Prof. Dr. Wolfgang Trutschnig

Arbeitsgruppe Stochastik/Statistik
Fachbereich AIHI
Universität Salzburg
www.trutschnig.net

Salzburg, WS 2023/24

Beurteilungsmodus:

- $\frac{2}{3}$ Kommanote Teil 1 (Schröder) + $\frac{1}{3}$ Kommanote Trutschnig (R-Teil).
- Max 100 Punkte im R-Teil.
- Max. 40 Punkte vom inkludierten Übungsteil: Jede Woche 5-6 R-Aufgaben, Ankreuzen mittels doodle, Vorrechnen bzw. Vorzeigen am Laptop in der UV (via WebEx).
- Max. 40 Punkte bei den Kleinprojekten - Ausarbeitung eines R-packages ODER Analyse echter Datensätze mit R.
- Prüfungsimmanente LV - Anwesenheitspflicht.
- Max. 20 Punkte Mitarbeit.
- Der R-Teil ist positiv, wenn Sie mind. 60 Punkte erreichen.

Warum R?

- Beschreibung auf der R-homepage: 'R is a free software environment for statistical computing and graphics'
- **R (und RStudio) ist freeware!**
- Läuft auf allen üblichen Plattformen.
- Weltweit verwendet, state of the art, standard in vielen Data Science Abteilungen.
- Sehr gute online Kurse (youtube, codeschool etc.) und tutorials.
- Laufend Aktualisierungen.
- Breites Anwendungsspektrum, Vielzahl an verwendbaren Paketen (derzeit > 20k) für diverseste Problemstellungen.
- Zusammen mit Python **die** Sprache für Data Science.
- Interaktive html-Grafiken und Dashboards (shiny).
- Schnittstellen mit anderer Software, insbesondere mit Latex und html (knitR/Markdown).
- Sehr einfache Simulation von Daten - erhöht die Verständlichkeit von Statistik & Wahrscheinlichkeitsrechnung.

Grobe Gliederung der kommenden Termine:

- Dezember:
  - How not to do statistics.
  - R Appetizer
  - Erste Schritte in R - learning by doing: Lösung konkreter (spannender stochastischer) Problemstellungen.
- Jänner - learning by doing.
  - Approximative Berechnung von Wahrscheinlichkeiten mit Hilfe von Simulationen.
  - Grundlagen Datenanalyse und deskriptive Statistik via ggplot2 und dplyr.

Zielsetzung:

Ende Jänner sollten Sie

1. Grundbegriffe der deskriptiven Statistik verstanden haben.
2. Einfache reale Datensätze mit Hilfe von R bearbeiten, analysieren und grafisch darstellen können.
3. Daten simulieren und Wahrscheinlichkeiten approximativ berechnen können.
4. Erkannt haben, dass es sich lohnt, R zu lernen.

- Der Ruf der Statistik könnte besser sein...
- *'Ich stehe Statistiken etwas skeptisch gegenüber. Denn laut Statistik haben ein Millionär und ein armer Kerl jeder eine halbe Million.' (Theodore Roosevelt)*
- *Traue keiner Statistik, die du nicht selbst gefälscht hast.*
- Wahrscheinlichkeitstheorie und Statistik sind Teilgebiete der Mathematik $\rightarrow$ Eigenschaften mathematisch beweisbar.
- **Problem ist die falsche Anwendung der Methodik!**
- Statistiken sind in print und online Medien sowie im Fernsehen allgegenwärtig.
- Die folgenden Beispiele sind grösstenteils [1], sowie der homepage *IFAS - Unsinn in den Medien* entnommen.

[1] A. Quatember: *Statistischer Unsinn - Wenn Medien an der Prozenthürde scheitern*, Springer Spektrum, 2015

MENSA
RICHTIG GUT ESSEN

ANTWORTEN und GEWINNEN Sie*:

1. Welche Suppe haben Sie heute gegessen?

_____

2. und wie hat Sie Ihnen geschmeckt:

☐ sehr gut ☐ gut ☐ befriedigend

3. sagt Ihnen unser Suppenangebot zu?

☐ sehr gut ☐ gut ☐ befriedigend

4. welche ist Ihre Lieblingssuppe?

_____

- Umfrage der Mensa der Johannes Kepler Universität Linz.
- Überraschenderweise hat sich absolut niemand über die Suppen in irgendeiner Weise beschwert!
- Lässt sich die Umfrage noch toppen?

Onlineumfrage: Nutzen Sie das Internet?

Dienstag, 16. Januar 2007 | 11:54 Uhr

Die Frage nach der Internetnutzung ist so alt wie das Internet selbst. Man wurde telefonisch befragt, in Fussgängerzonen darauf angeprochen, bei Gewinnspielen in Zeitungen um seine Meinung gebeten. Alles schön und gut, doch warum sollte man für so eine leichte Frage diesen Aufwand betreiben. Man könnte sie doch einfach einmal online stellen. Unsinn? Nein, hier im Land der Gegensätze ist alles möglich.

Denn der brasilianische Mediengigant *Rede Globo* wollte es auf seiner Webseite "*Globo Online*" tatsächlich ganz genau wissen:

(Nutzen Sie das Internet? Ja / Nein)

- An der Umfrage nahmen 2000 Personen teil.
- 68% gaben an, das Internet zu nutzen.
- 32% nutzen das Internet also nicht, füllen aber ONLINE Fragebögen aus?!

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Ein Viertel der Studenten is...

www.krone.at/Oesterreich/Ein_Viertel_der_Studenten_ist_alkoholabhaengig-Alarmierende_Studie-Story-22

Willkommen | krone.at-Forum | Leser-Aktionen | Login

Gebrauchtwagen Immobilien Jobs Kleinanzeigen

Kronen Zeitung

Krone-Abo
Titelseite
Ticketshop
Mobile Welt

Wien
5°C
Montag
wetter.tv

Das neue BonusCard-Programm

Nachrichten | Sport | Stars & Society | Digital | Freizeit | Auto | Videos | TV-Programm | Horoskop | epaper

Startseite > Österreich > Ein Viertel der Studenten ist alkoholabhängig

07.03.2016 - 17:37

Alarmierende Studie

29.09.2010, 13:23

Ein Viertel der Studenten ist alkoholabhängig

Foto: APA/HELMUT FOHRINGER

Ein Viertel der heimischen Studenten konsumiert Alkohol "in einem sehr hohen, gesundheitsgefährdenden Ausmaß", wie eine Studie der Österreichischen JungArbeiterBewegung ergeben hat. Außerdem zeigte die Untersuchung unter rund 1.400 Bewohnern von Studentenheimen, dass ein Viertel aller Befragten regelmäßig raucht. Das ist allerdings signifikant weniger als bei einer Studie aus dem Jahr 2002 unter Stellungspflichtigen beim Bundesheer.

289 Kommentare

Jetzt kommentieren

Insgesamt gaben rund 90 Prozent der Heimbewohner an, Alkohol zu konsumieren. Bei der Hälfte davon ist das Ausmaß unbedenklich, bei jeweils einem Viertel allerdings bedenklich bzw. liegt bereits eine Abhängigkeit vor - vor allem junge Männer sind davon betroffen.

Hauptmotiv für den Alkoholkonsum ist "weil es mir schmeckt" (48,5

Infobox

- Die Videos zum Tagesgeschehen
- zur krone.at-Startseite

Verwandte

Mehr Österreich-Nachrichten

7 Verletzte in Wien
Nach Bandenkrieg: Tschetschenen wollen "Frieden"

Musterprozess in OÖ
Psychisch bedingte Krankmeldung ist kein Freibrief

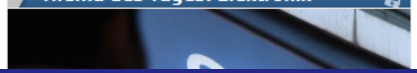
Nicht verpassen!
Außenminister Sebastian Kurz im krone.at-Talk

"Besorgniserregend"
24 Sexübergriffe auf Frauen zu Silvester!

Mittlerlehner im TV:
"EU-Grenzen notfalls mit Gewalt schützen"

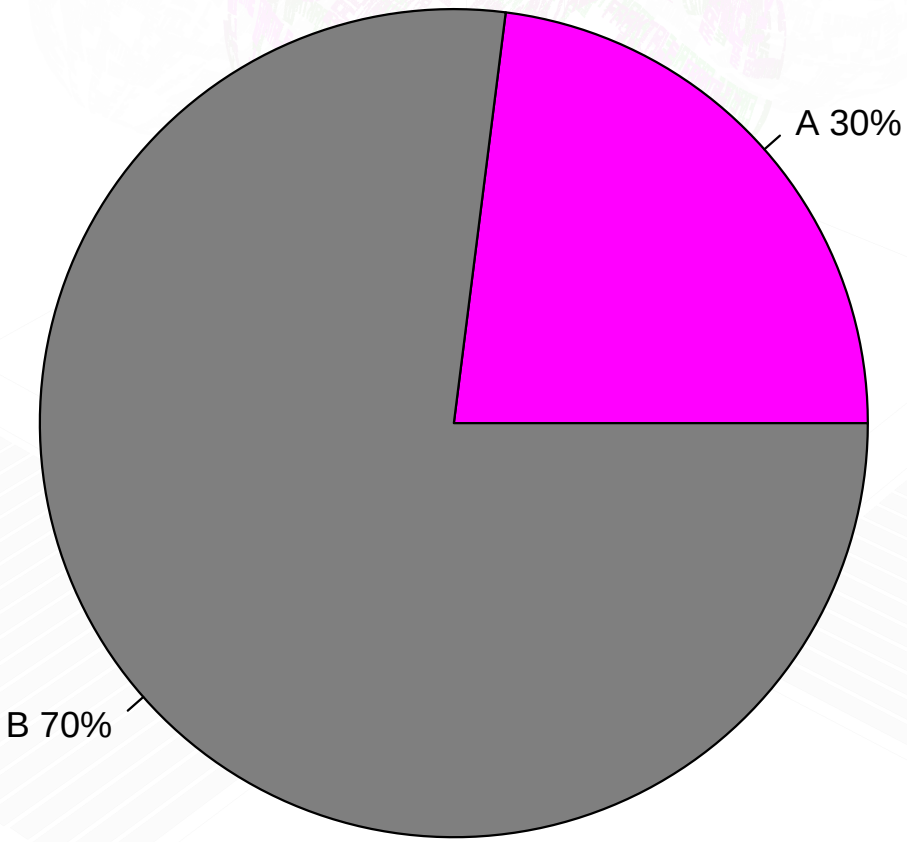
Werbung

Thema des Tages: Elektronik



VERBREITETE AUFLAGE DER FÜHRENDEN TAGESZEITUNGEN
Auflage Montag-Samstag
883.580
317.043
291.996
204.530
107.810
104.619
Krone
ÖSTERREICH
Kleine Zeitung
Kurier
Presse
Standard
a

Der Anteil der Kinder mit Leseschwäche beträgt nur 30 %

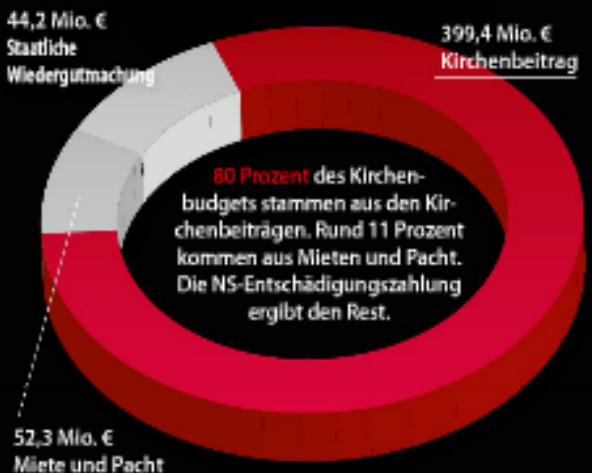


finanzielles Problem hat die Kirche dadurch nicht. Sie sitzt auf einem Schatz
Grundstücken, Gebäuden, Steuerprivilegien und staatlichen Zuschüssen.
das überhaupt noch zeitgemäß? *Von Oliver Schönsleben*

Die Gläubigen laufen der katholischen Kirche in Scharen
n. Alleine im Vorjahr
dten 52.425 Personen
Kirche den Rücken zu.
it fehlen im Klingelbeu-
nd fünf Millionen Euro
Kirchenbeiträgen. Trotz-
müssen sich die Gottes-
ner keine Gedanken
ihre Finanzen machen.
insgesamt 3,8 Milliarden
pro Jahr subventio-
en die österreichischen
erzahler die Einrichtun-
der Kirchen und Religi-
gemeinschaften, rechnet
istoph Baumgarten in
em Buch „Gottes Werk
unser Beitrag – Kir-
finanzierung in Öster-
" vor. Statistisch finan-
jeder von uns mit rund
Euro die religiösen Ein-
tungen. Aus den ver-
edensten Quellen ergie-
sich die Geldströme und
sten auch das Bundes-
get. Weekend Magazin
die Privilegien der ka-
tholischen Kirche auf und zeigt, auf welchem Milliardenschatz der Klerus sitzt.

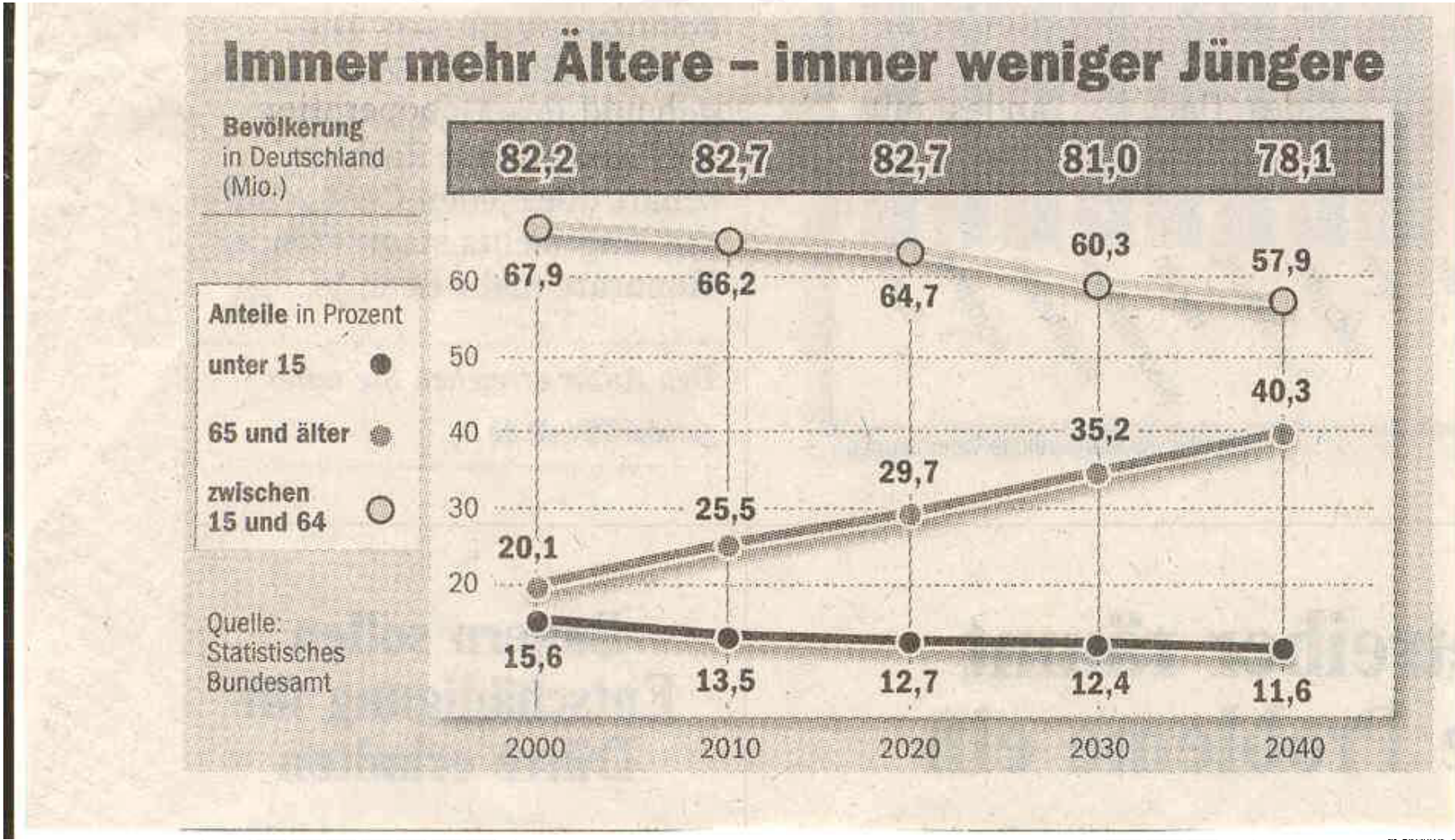
Faktor 1: Großgrundbesitz. Insgesamt verfügt die katholische Kirche über einen Grundbesitz von 207.000 Hektar – das entspricht der fünffachen Größe der Bundeshauptstadt Wien. Damit ist die Kirche der größte Grundbesitzer nach den Bundesforsten. Dafür gibt es auch satte EU-Agrarförderungen. Laut Transparenzdatenbank fließen jährlich vier Millionen Euro aus Brüssel. Nicht nur das Benediktinerstift Admont profitiert davon. Die Steirer gelten mit ihren 25.000 Hektar Grundbesitz als das reichste Stift des Landes und sind nicht nur wegen des himmlischen Weins aus ihrem Weingut „Friedenstür" bekannt.

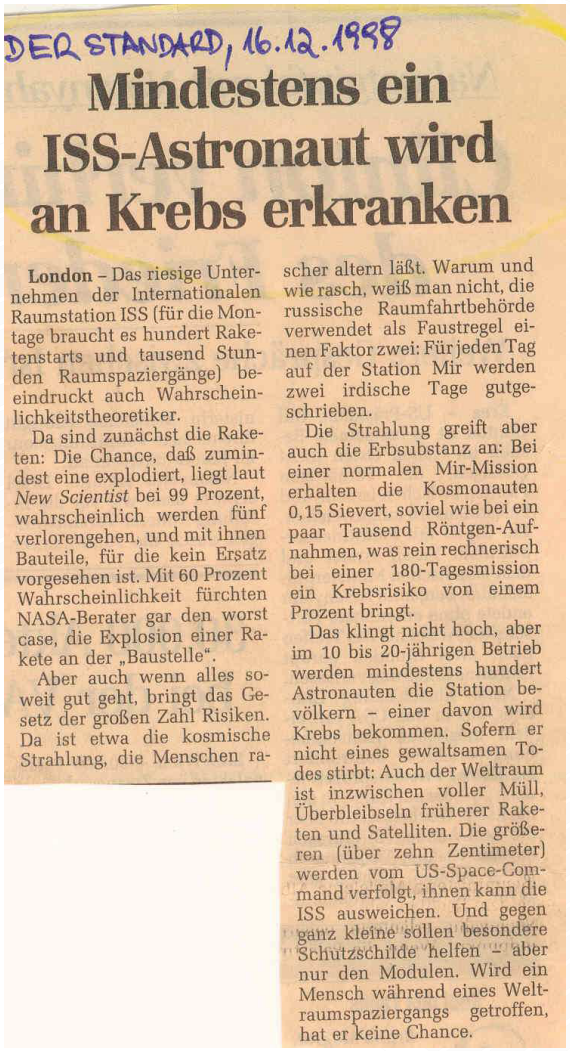
BUDGET DER KATHOLISCHEN KIRCHE



Auch das Erzbistum Wien kann fast eine halbe Million Euro an EU-Förderung verbuchen. Grundstückskäufe der Erzdiözese Wien in Niederösterreich sorgten vor einigen Jahren für Aufregung. „Mit überhöhten Preisen wurden Ortsansässige ausgebootet", kritisiert Niko Alm. Der Mit-Initiator des Volksbegehrens gegen Kirchenprivilegien geht einen Schritt weiter: „Der kirchliche Grundbesitz stieg in den vergangenen Jahren um rund 30 Prozent. Das erinnert mich an ‚land grabbing', die Aneignung von Land mit zweifelhaften Mitteln."

Faktor 2: Steuerprivilegien. Die zwölf staatlich anerkannten Religionsgemeinschaften in Österreich genießen bei Finanzministerin Maria Fekter zahlreiche Privilegien, von denen Privatpersonen nur träumen können. Kirchen sind vom Stiftungs- und Fondsgesetz ausgenommen und müssen für ihren Grundbesitz keine Grundsteuer zahlen, »


Immer mehr Ältere – immer weniger Jüngere
Bevölkerung in Deutschland (Mio.)
82,2
82,7
82,7
81,0
78,1
Anteile in Prozent
unter 15
65 und älter
zwischen 15 und 64
Quelle: Statistisches Bundesamt
60
50
40
30
20
67,9
66,2
64,7
60,3
57,9
40,3
35,2
29,7
25,5
20,1
15,6
13,5
12,7
12,4
11,6
2000
2010
2020
2030
2040



DER STANDARD, 16.12.1998

Mindestens ein ISS-Astronaut wird an Krebs erkranken

London – Das riesige Unternehmen der Internationalen Raumstation ISS (für die Montage braucht es hundert Raketenstarts und tausend Stunden Raumspaziergänge) beeindruckt auch Wahrscheinlichkeitstheoretiker.

Da sind zunächst die Raketen: Die Chance, daß zumindest eine explodiert, liegt laut *New Scientist* bei 99 Prozent, wahrscheinlich werden fünf verlorengehen, und mit ihnen Bauteile, für die kein Ersatz vorgesehen ist. Mit 60 Prozent Wahrscheinlichkeit fürchten NASA-Berater gar den worst case, die Explosion einer Rakete an der „Baustelle".

Aber auch wenn alles soweit gut geht, bringt das Gesetz der großen Zahl Risiken. Da ist etwa die kosmische Strahlung, die Menschen rascher altern läßt. Warum und wie rasch, weiß man nicht, die russische Raumfahrtbehörde verwendet als Faustregel einen Faktor zwei: Für jeden Tag auf der Station Mir werden zwei irdische Tage gutgeschrieben.

Die Strahlung greift aber auch die Erbsubstanz an: Bei einer normalen Mir-Mission erhalten die Kosmonauten 0,15 Sievert, soviel wie bei ein paar Tausend Röntgen-Aufnahmen, was rein rechnerisch bei einer 180-Tagesmission ein Krebsrisiko von einem Prozent bringt.

Das klingt nicht hoch, aber im 10 bis 20-jährigen Betrieb werden mindestens hundert Astronauten die Station bevölkern – einer davon wird Krebs bekommen. Sofern er nicht eines gewaltsamen Todes stirbt: Auch der Weltraum ist inzwischen voller Müll, Überbleibseln früherer Raketen und Satelliten. Die größeren (über zehn Zentimeter) werden vom US-Space-Command verfolgt, ihnen kann die ISS ausweichen. Und gegen ganz kleine sollen besondere Schutzschilde helfen – aber nur den Modulen. Wird ein Mensch während eines Weltraumspaziergangs getroffen, hat er keine Chance.

- Wo liegt hier der Denkfehler ?
- Angenommen, das tatsächliche Krebsrisiko eines Astronauten ist 1%
- In einer Urne liegen 100 Kugeln, von denen eine weiss und 99 schwarz sind.
- Sie ziehen (mit geschlossenen Augen) 100 Mal eine Kugel mit Zurücklegen.
- Kann es auch passieren, dass Sie nie die weisse Kugel gezogen haben?
- Natürlich!
- Wie gross ist die Wahrscheinlichkeit dafür, dass keine weisse Kugel gezogen wird?

Falsche Schlussfolgerungen

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

In Linz gibt es die meisten ...

www.nachrichten.at/oberoesterreic

Archiv | ePaper | Digital | AAA

Montag, 7. März 2016, 17:11 Uhr

Linz: 6°C
Ort wählen »

nachrichten

STARTSEITE | OBERÖSTERREICH | SPORT | FREIZEIT | REISEN | BESTE STELLEN | IMMOBILIEN | MOTOR

Linz | Wels | Steyr | Mühlviertel | Innviertel | Salzkammergut | Nachrufe | Wetter | Gutscheine

Oberösterreich

In Linz gibt es die meisten Single-Haushalte Österreichs

LINZ. In den vergangenen 40 Jahren hat die Zahl der Single-Haushalte in Österreich enorm zugenommen.

Ein Prosit auf das „Hotel Mama" in Oberösterreich. Bild:

Das geht aus einer Analyse der Registerzählung der Statistik Austria 2011 hervor. Während 1971 noch jeder vierte Privathaushalt von nur einer Person bewohnt wurde, war es 2011 bereits jeder dritte.

Single-Hochburg Linz: Oberösterreichs Landeshauptstadt ist im Österreich-Vergleich die Stadt mit den allermeisten Einpersonenhaushalten. 52,5 Prozent der Linzer leben allein in einer Wohnung. „Das liegt daran, dass in Linz viele Menschen leben, die älter als 65 Jahre sind. Viele ältere Menschen, besonders Frauen, leben allein. Ab dem 80. Lebensjahr sind das fast zwei Drittel", sagt Manuela Lenk, Bereichsleiterin Registerzählung von der Statistik Austria. Der Trend zum Alleine-Wohnen ist übrigens vor allem ein Phänomen der großen Städte. Der Anteil der Einpersonen-Haushalte beträgt in Steyr 44 und in Wels 39,7 Prozent, während er beispielsweise in den Bezirken Perg bei 25,2 Prozent und Freistadt bei 25,3 liegt.

Cookies erleichtern die Bereitstellung unserer Dienste. Mit der Nutzung unserer Dienste erklären

OK

- Singlestadt Linz?
- Hintergrundinfo: Linz hat ca. 200.000 Einwohner und 90.000 Haushalte.
- Es gibt daher ca. $0.52 \cdot 90000 = 46800$ Single Haushalte und 43200 Haushalte mit mehr als einer Person.
- In den 43200 Nicht-Single-Haushalten wohnen daher $200000 - 46800 = 153200$ Personen.
- Preisfrage: Wie viel Prozent der Linzer wohnen alleine?
- Antwort: $p = \frac{46800}{200000} = 0.234$.
- 'Nur' 23.4 % wohnen alleine.

Hohe Anteile von Geimpften in der Altersgruppe 75+ auf der Covid
t die Impfung nicht besser als ein
Münzwurf?
e Medien voll mit Aussagen von doppelt Geimpften
auf den Covid-19+Intensivstationen. Gehen wir noch einen Schritt
weiter und nehmen fiktiv an, in der **Altersgruppe 75+ wären sogar 50%**
der Covid-19-Intensivpatient*innen doppelt geimpft: Bedeutet ein so
hoher Prozentsatz tatsächlich, dass Impfen nichts bringt?

Um diese Frage beantworten zu können, benötigen wir neben dem Anteil von doppelt geimpften 75+ Covid-19-Intensivpatient*innen auch den Prozentsatz an doppelt geimpften Österreicher*innen in dieser Altersgruppe: Mitte November 2021 lag dieser Anteil bei knapp über 90% – im Schnitt sind also **9 von 10 Österreicher*innen in der Altersgruppe 75+ sind doppelt geimpft** (Quelle: https://info.gesundheitsministerium.gv.at/impflage). Zusätzlich nehmen wir vereinfachend noch an, dass die Komorbiditäten (Vorerkrankungen) in beiden Gruppen – **doppelt** Geimpfte und **nicht doppelt** Geimpfte – in etwa gleich verteilt sind.

Unter den 75+ Österreicher*innen ist das Verhältnis der **doppelt** Geimpften zu den **nicht doppelt** Geimpften also **9:1**, das Verhältnis der Covid-19-Intensivpatient*innen in der selben Altersgruppe beträgt aber 1:1.

Das statistische Risiko nach einer Ansteckung mit Covid-19 als nicht doppelt geimpfte 75+ Person auf der Intensivstation behandelt werden zu müssen, ist demnach um den Faktor 9 höher als für eine doppelt geimpfte 75+ Person.

IDA Lab Salzburg

Home Teams About Us Collaborations Projects Publications Events Covid-19

News PR Contact

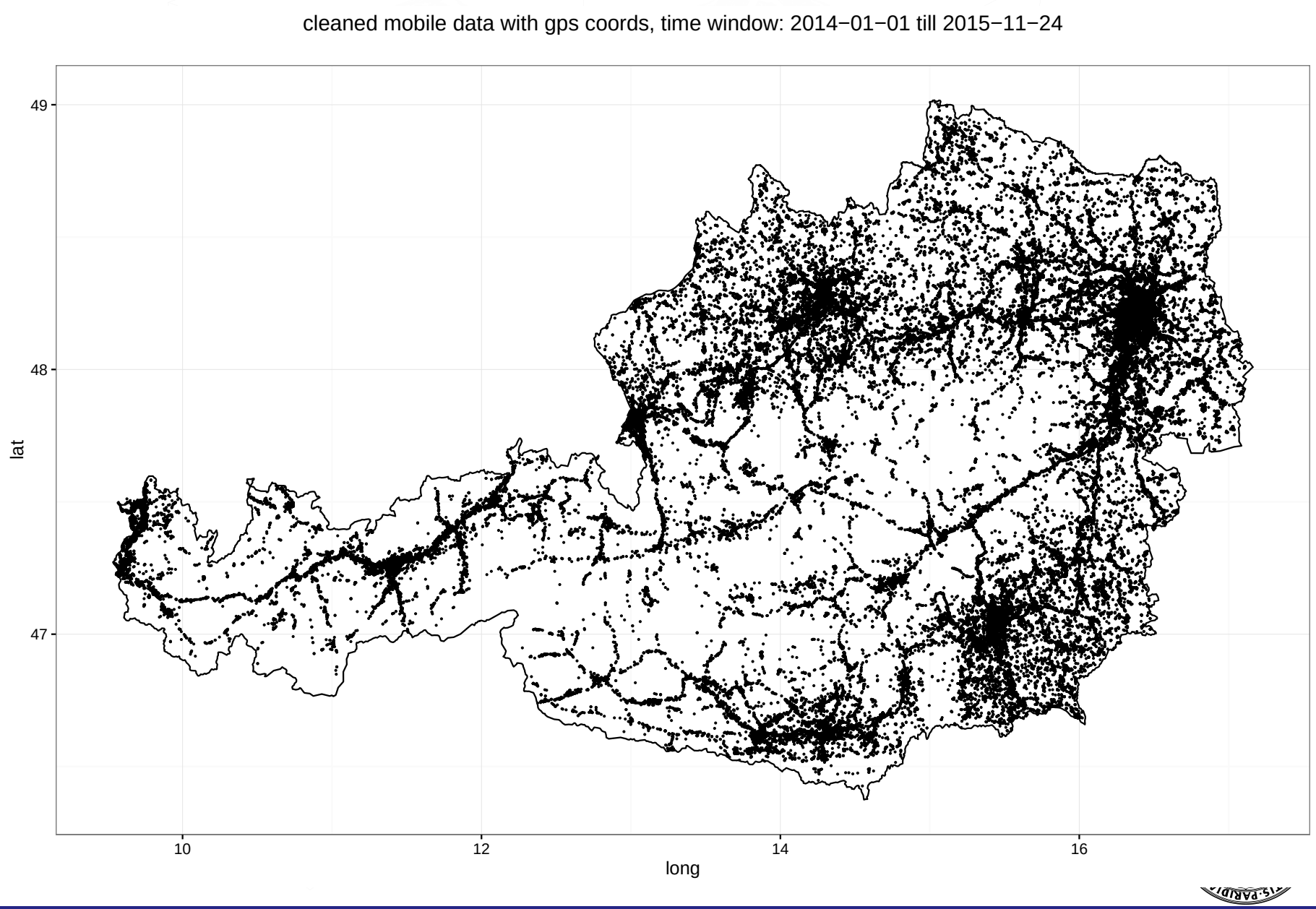
About IDA Lab

The Lab for Intelligent Data Analytics Salzburg (IDA Lab) funded by Land Salzburg (WISS 2025) serves as central facility for basic and applied research, as well as for knowledge and technology transfer in Data Science, Machine Learning and Statistics. It is a cooperation of the Paris Lodron Universität Salzburg (PLUS, Lead), the Paracelsus Medical University (PMU), the Salzburg Research (SRFG) and the University of Applied Sciences (FHS).

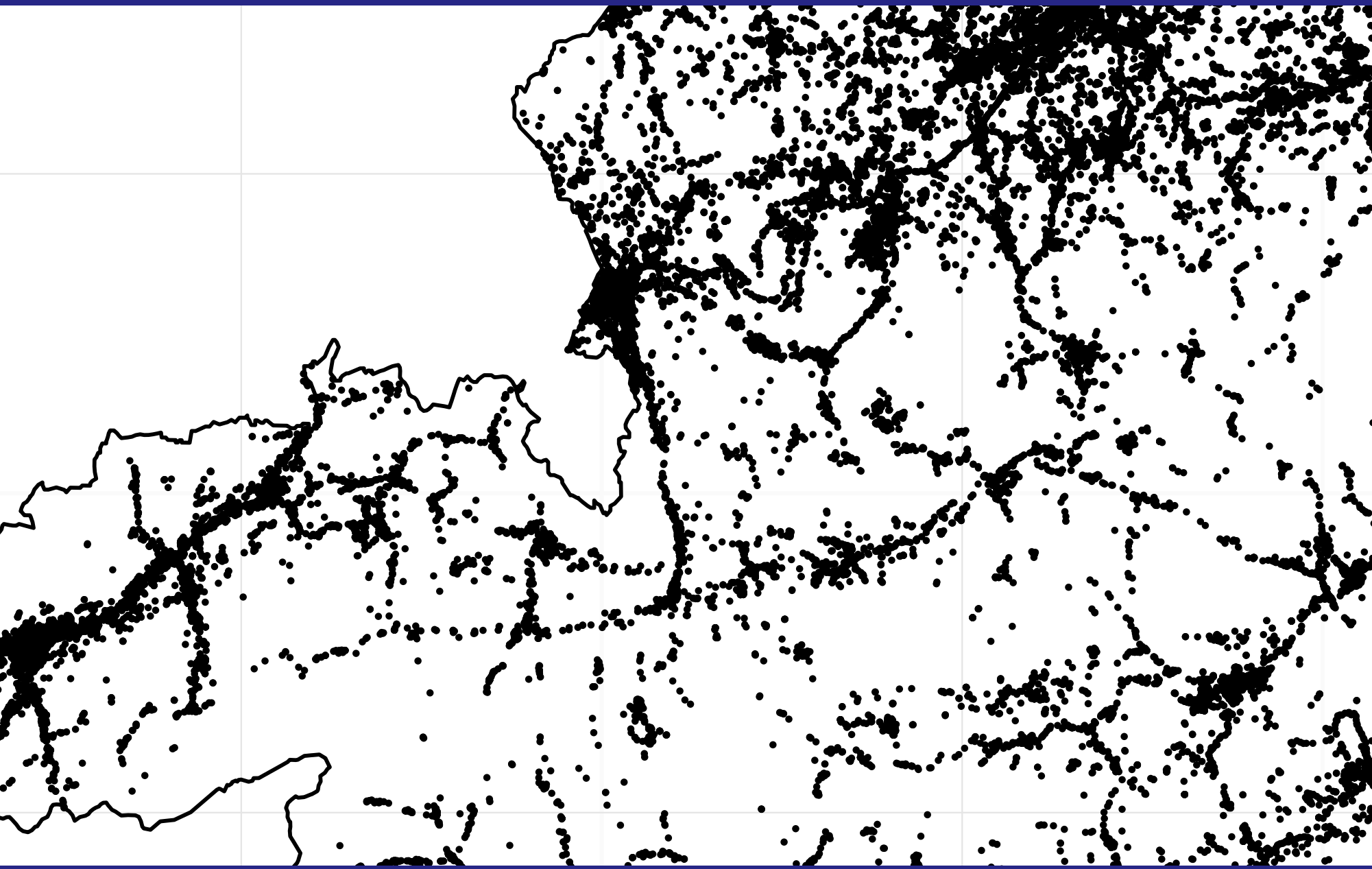
Example (RTR-Netztest)

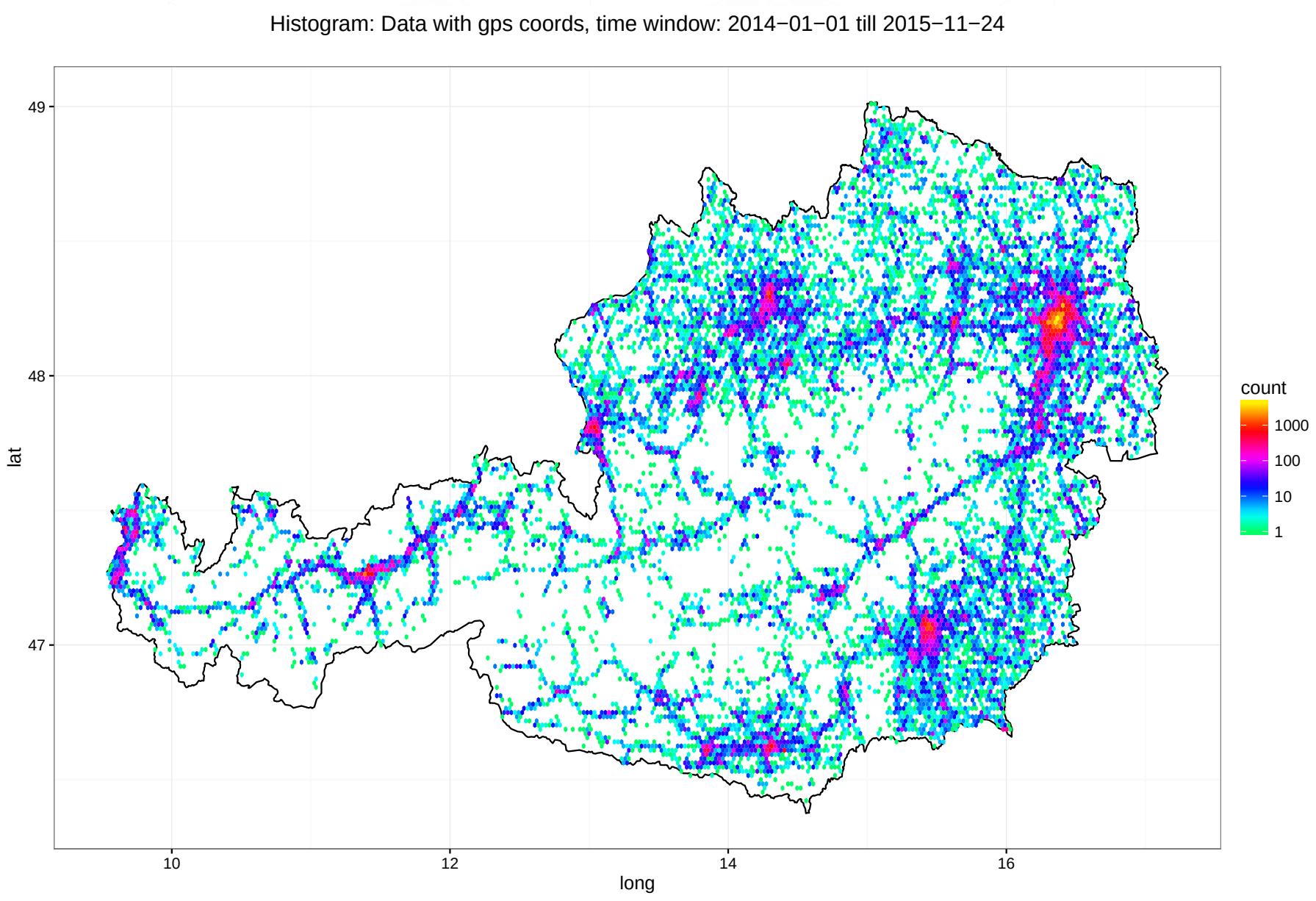
- Grosser, offener Datensatz (ca. 1.5 Mio. Einträge 2014-2015).
- Geschwindigkeitsmessungen mobile und fixed (App downloadbar für iOS und Adroid).
- Siehe https://www.netztest.at/de/ für Hintergrundinfo
- In der Vorlesung betrachten wir im folgenden nur mobile Messungen, die nicht von 'heavy-usern' stammen (239.060 Messungen).

id	mtime	longitude	latitude	device	op_name	rtr_speed_dl
0adff95f3-880a	2014-01-01 00:10:00	13.2076	47.3430	Galaxy Note 2	Pear	3226.00
0c2852599-37bd	2014-01-01 00:16:00	13.2286	47.3459	Galaxy Note 2	Pear	2068.00
048a47c74-c7a6	2014-01-01 00:21:00	15.7516	48.2046	Galaxy Note 10.1 LTE	Kiwi	11604.00
069f25c23-5089	2014-01-01 00:21:00	16.3971	48.2307	Galaxy S3	Kiwi	1948.00
0ede0db57-1d56	2014-01-01 00:27:00	15.6321	48.2080	Galaxy Note 10.1 LTE	Kiwi	6277.00
0383e9306-9612	2014-01-01 00:41:00	15.5196	47.1249	iPhone 5s	Apple	3577.00
04b28bf72-a489	2014-01-01 00:42:00	15.5194	47.1250	iPhone 5s	Apple	6526.00
0ef71e021-3414	2014-01-01 00:48:00	13.1981	47.3490	iPhone 4s	Pear	3155.00
0ef957387-3204	2014-01-01 01:05:00	13.7706	46.8187	Galaxy S4 LTE	Pear	11354.00
06c813ae1-e247	2014-01-01 01:07:00	15.1820	48.7588	iPhone 5s	Kiwi	13600.00

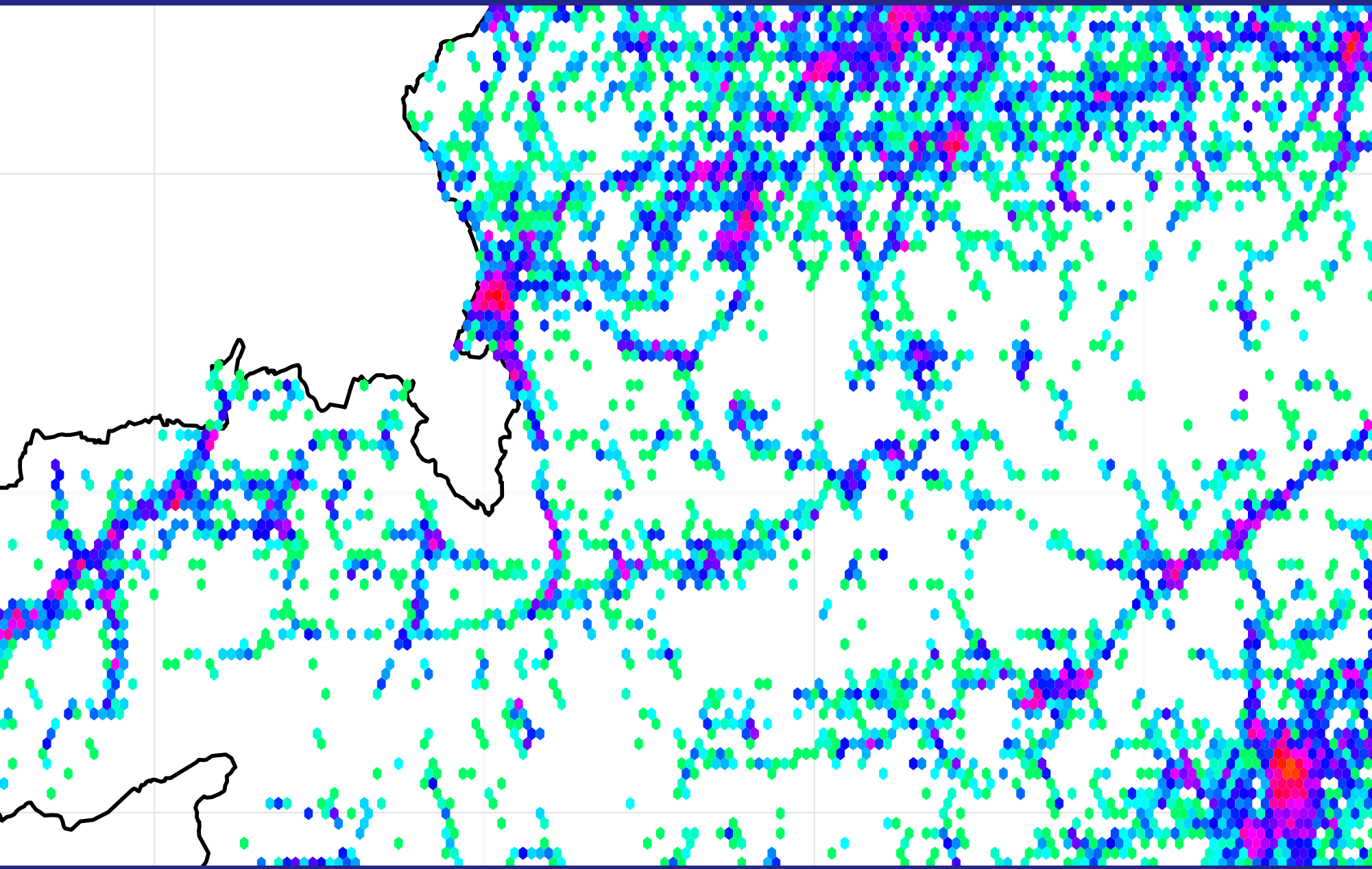

cleaned mobile data with gps coords, time window: 2014−01−01 till 2015−11−24
49
48
47
lat
10
12
14
16
long

RTR-Netztest Data

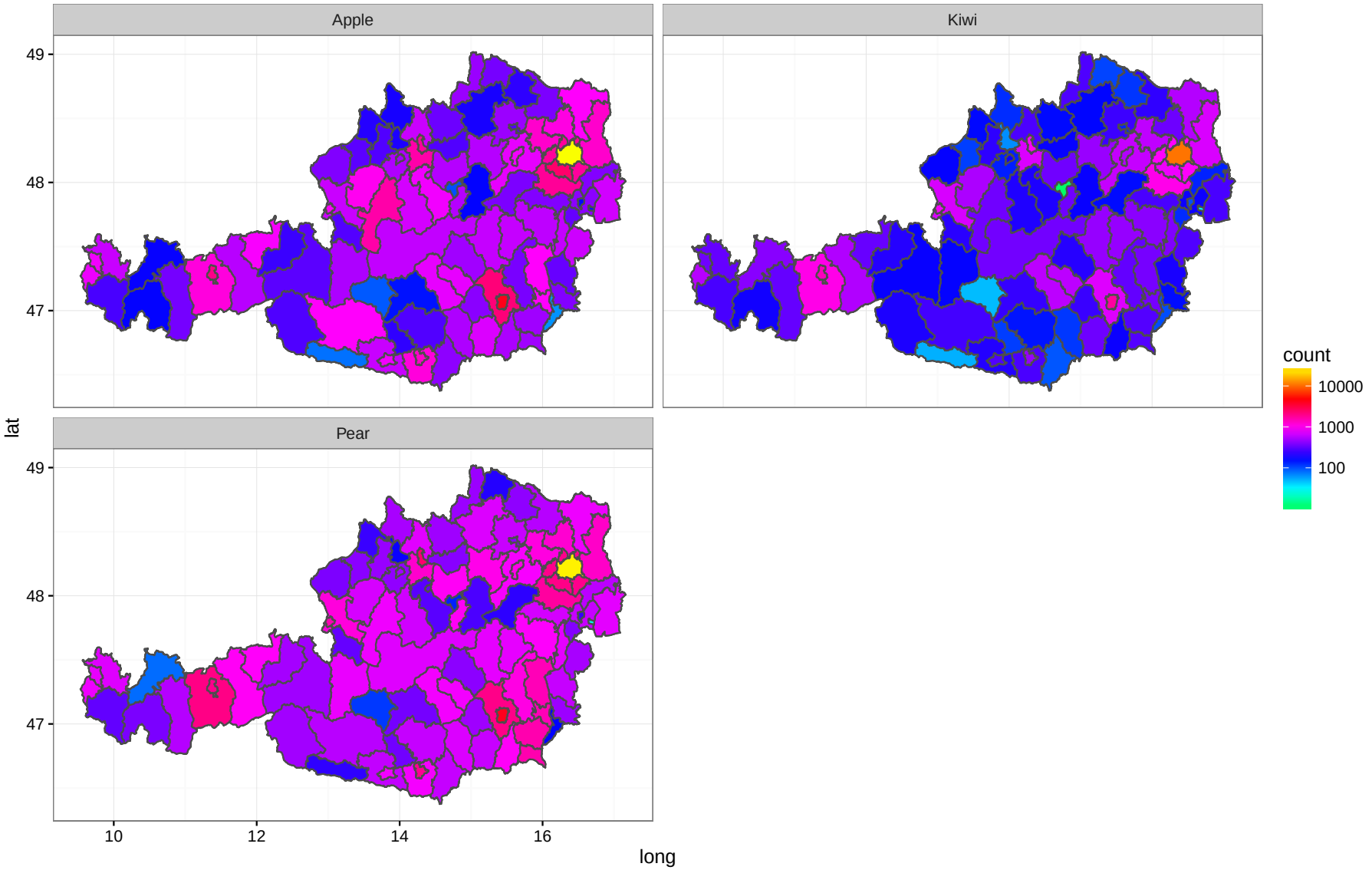


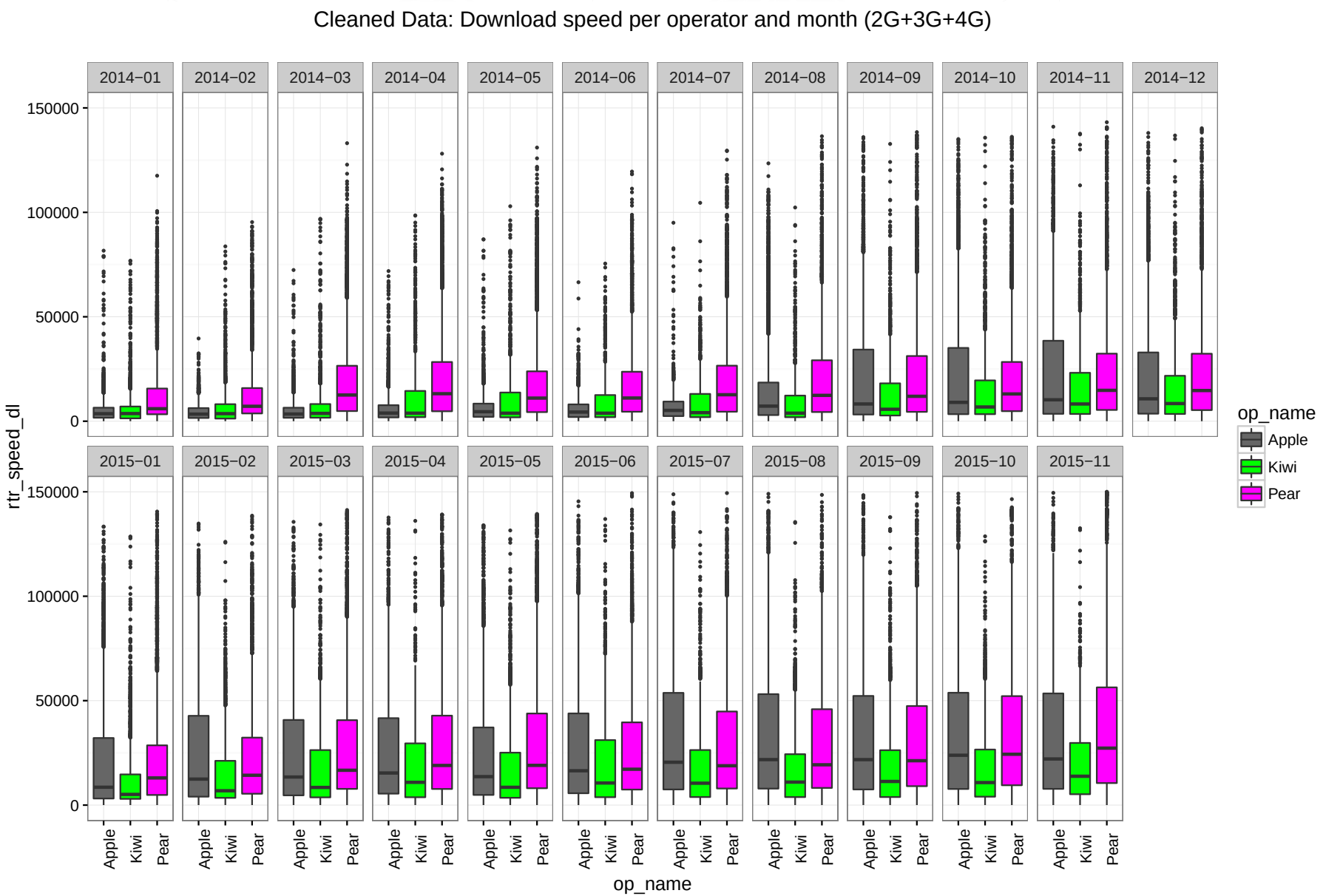

Histogram: Data with gps coords, time window: 2014−01−01 till 2015−11−24
49
48
47
lat
10
12
14
16
long
count
1000
100
10
1

RTR-Netztest Data

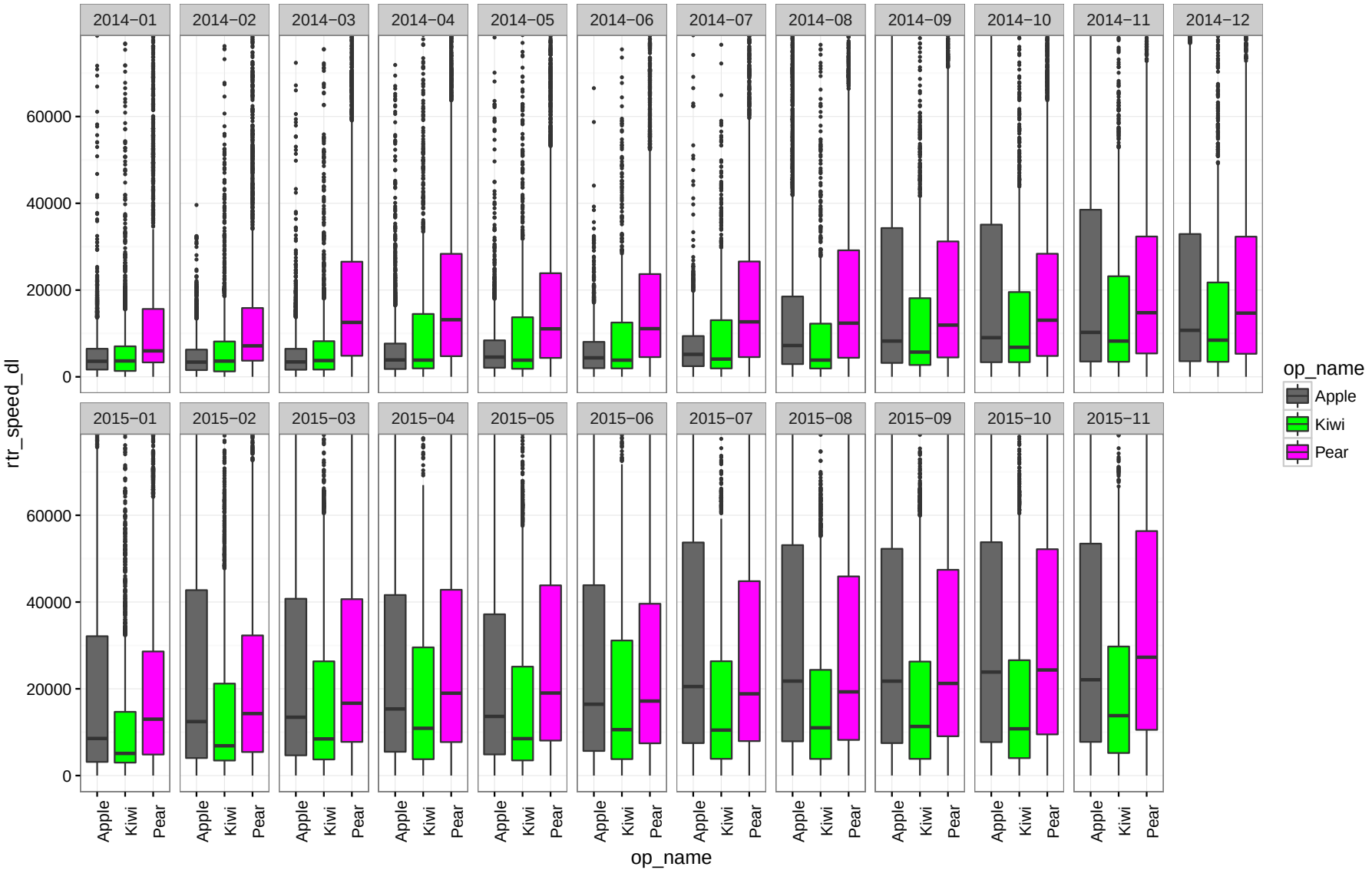


Overall Sample size per district: Data from 2014−01−01 till 2015−11−24

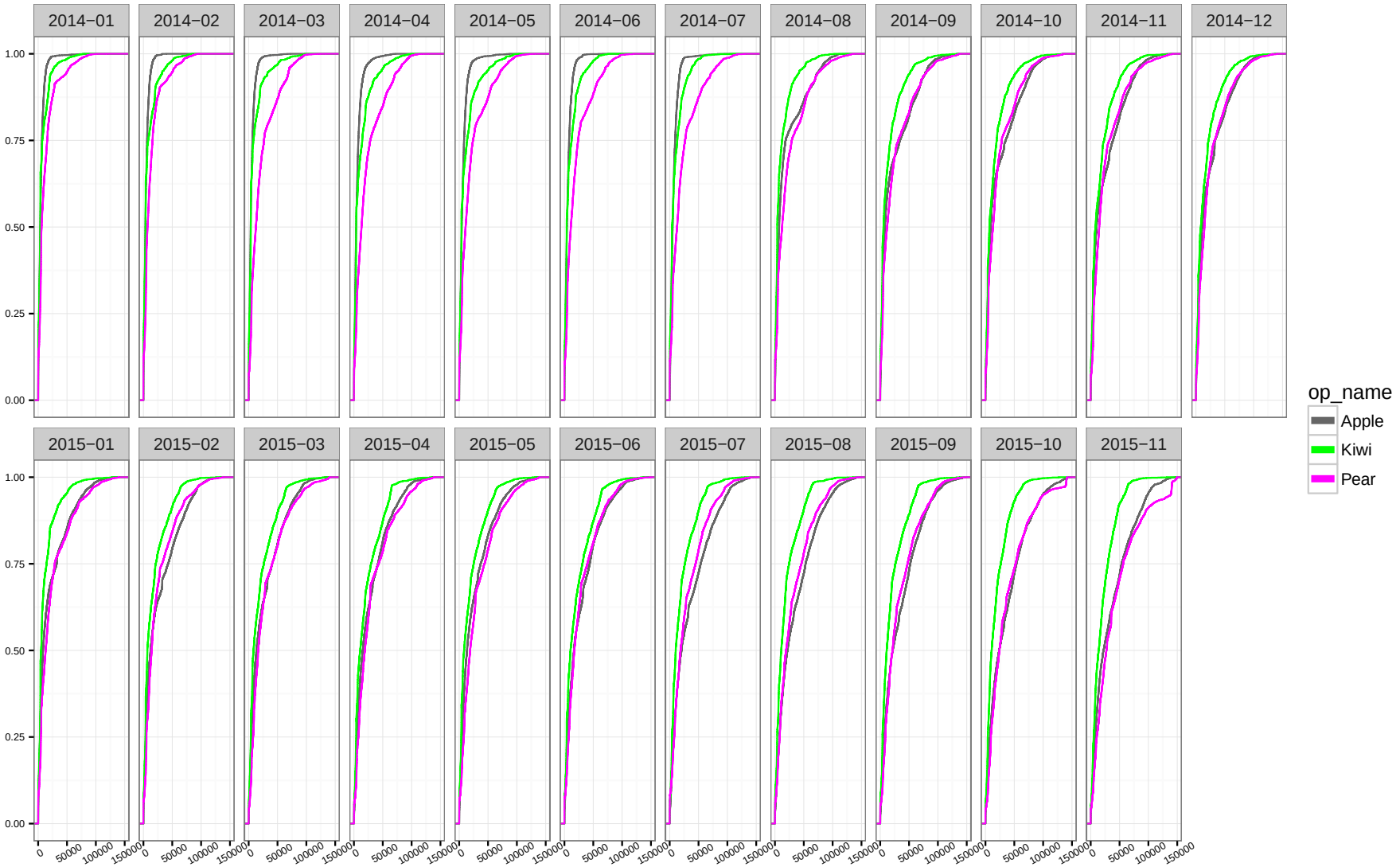



Cleaned Data: Download speed per operator and month (2G+3G+4G)
2014−01
2014−02
2014−03
2014−04
2014−05
2014−06
2014−07
2014−08
2014−09
2014−10
2014−11
2014−12
2015−01
2015−02
2015−03
2015−04
2015−05
2015−06
2015−07
2015−08
2015−09
2015−10
2015−11
rtr_speed_dl
150000
100000
50000
0
Apple
Kiwi
Pear
op_name
op_name
Apple
Kiwi
Pear

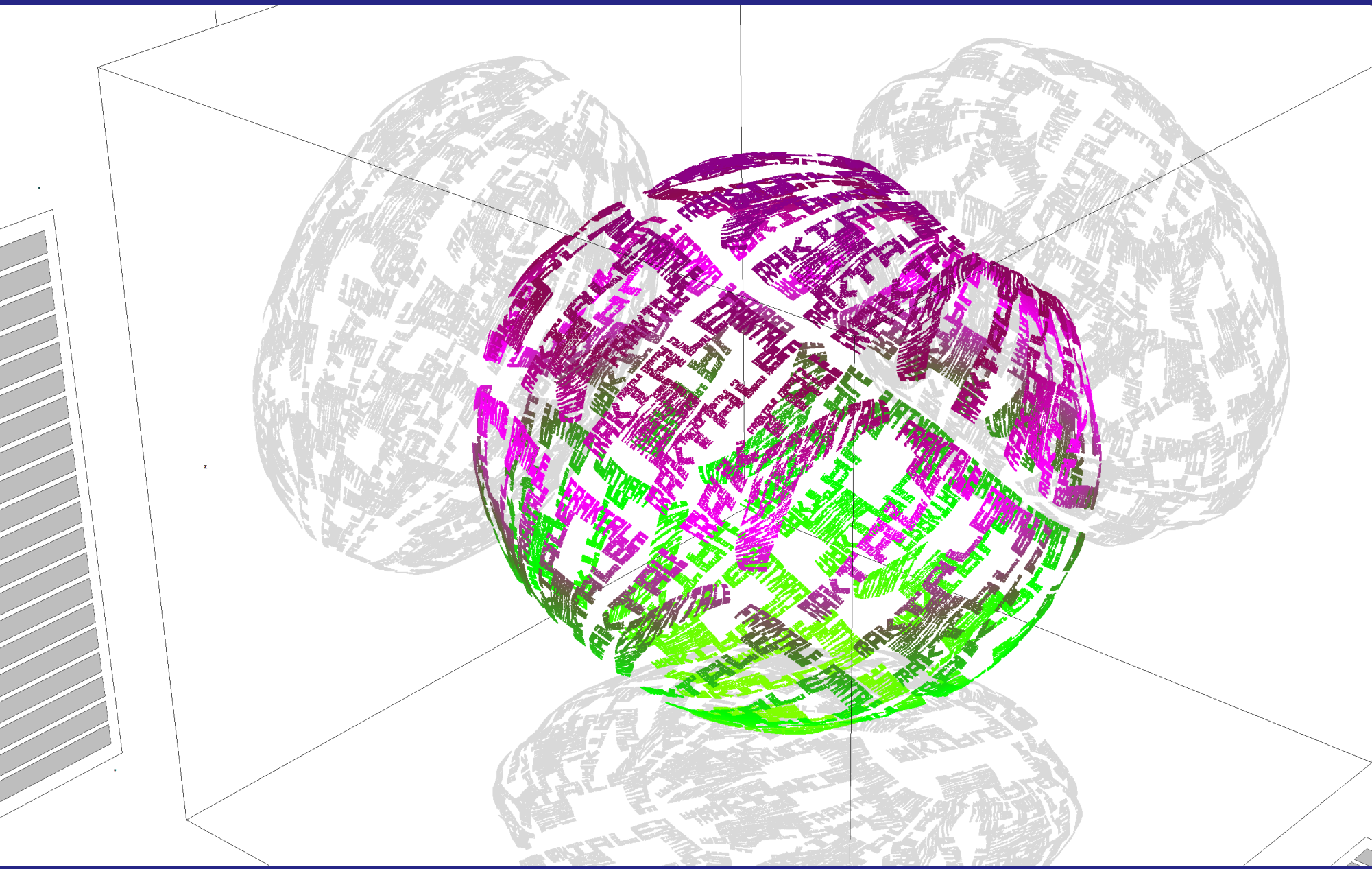
Cleaned Data: Download speed per operator and month (2G+3G+4G)



ecdf: Download speed per operator



Fraktale



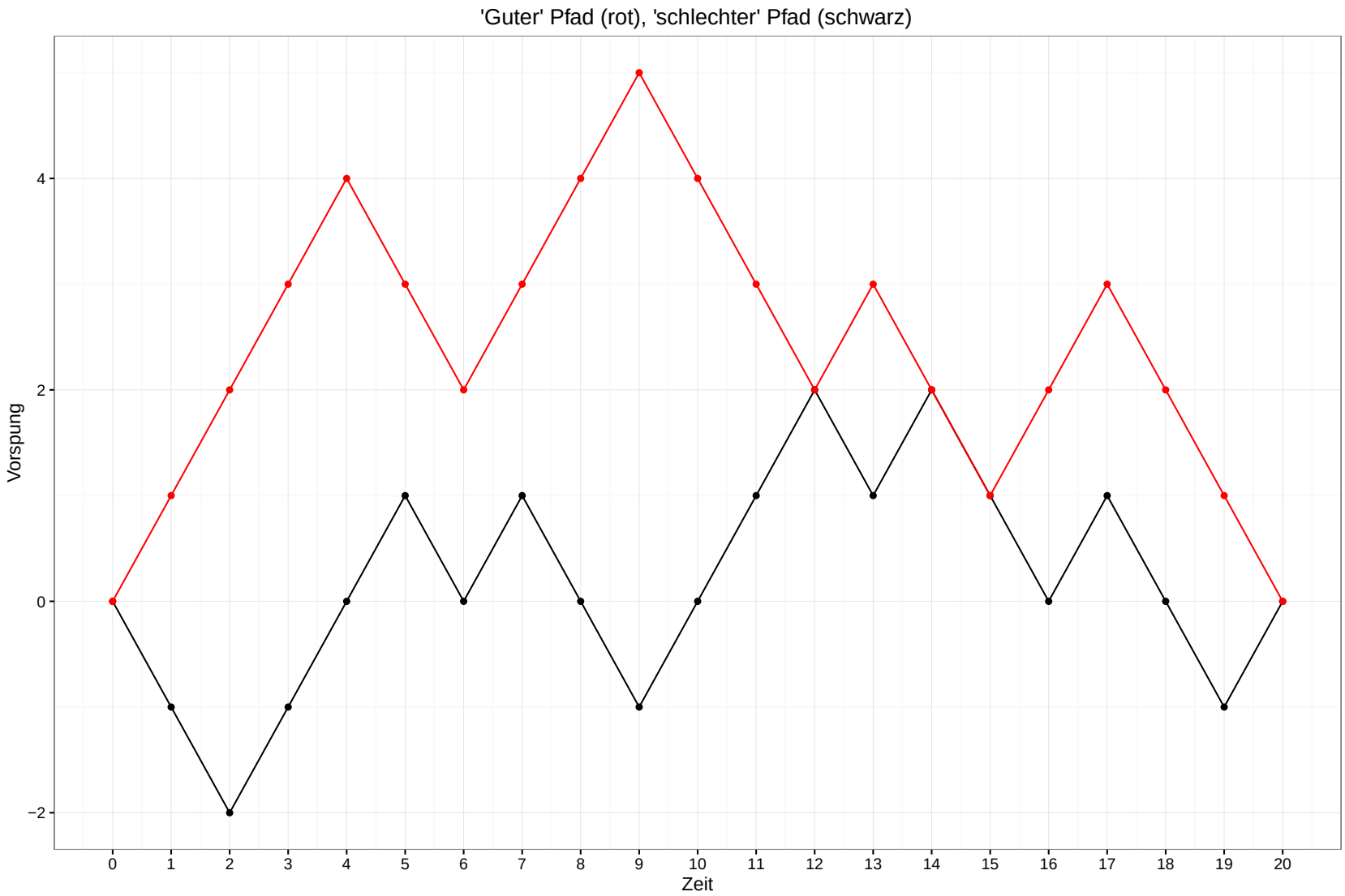
Problem 1: 50:50 Wahlergebnis

- Ein Dorf wählt seinen Bürgermeister.
- Es gibt zwei Kandidaten, V und H.
- $2n$ Personen gehen zur Wahl, beide Kandidaten erhalten genau n Stimmen.
- Es wird Stimme für Stimme nacheinander ausgezählt.
- Gesucht: Wahrscheinlichkeit dafür, dass V zu jedem Zeitpunkt mindestens so viele Stimmen hat wie H?
- Lernziel: Befehle rep, (sample), cumsum, if.

Vereinfachung des Problems:

- Schreiben 1 falls V eine Stimme bekommt und -1 falls H eine Stimme bekommt.
- Für $n = 5$ erhalten wir z.B. $1, -1, 1, -1, 1, -1, -1, 1, -1, 1$.
- Hat im Bsp. V in jedem Schritt mind. so viele Stimmen wie H?

Problem 1: 50:50 Wahlergebnis



Problem 1: 50:50 Wahlergebnis, cont.

- Sind $x_1, x_2, \ldots, x_{2n} \in \{-1, 1\}$ die so erhaltenen Werte dann gilt (nach Konstruktion) $\sum_{i=1}^{2n} x_i = 0$.
- V hat in jedem Schritt mindestens so viele Stimmen wie H genau dann wenn

$$\sum_{i=1}^{t} x_i \geq 0$$

  für jeden Zeitpunkt $t \in \{1, 2, \ldots, 2n-1, 2n\}$.
- Damit ist klar, wie die Situation in R simuliert werden kann

```
1 n <- 10
2 x <- sample(c(rep(1,n),rep(-1,n)),size=2*n)      #Simulation des
      Auszaehlens
3 x
4 cumsum(x)                                        #alle
      Partialsummen
5 min(cumsum(x))
```

Problem 1: 50:50 Wahlergebnis, cont.

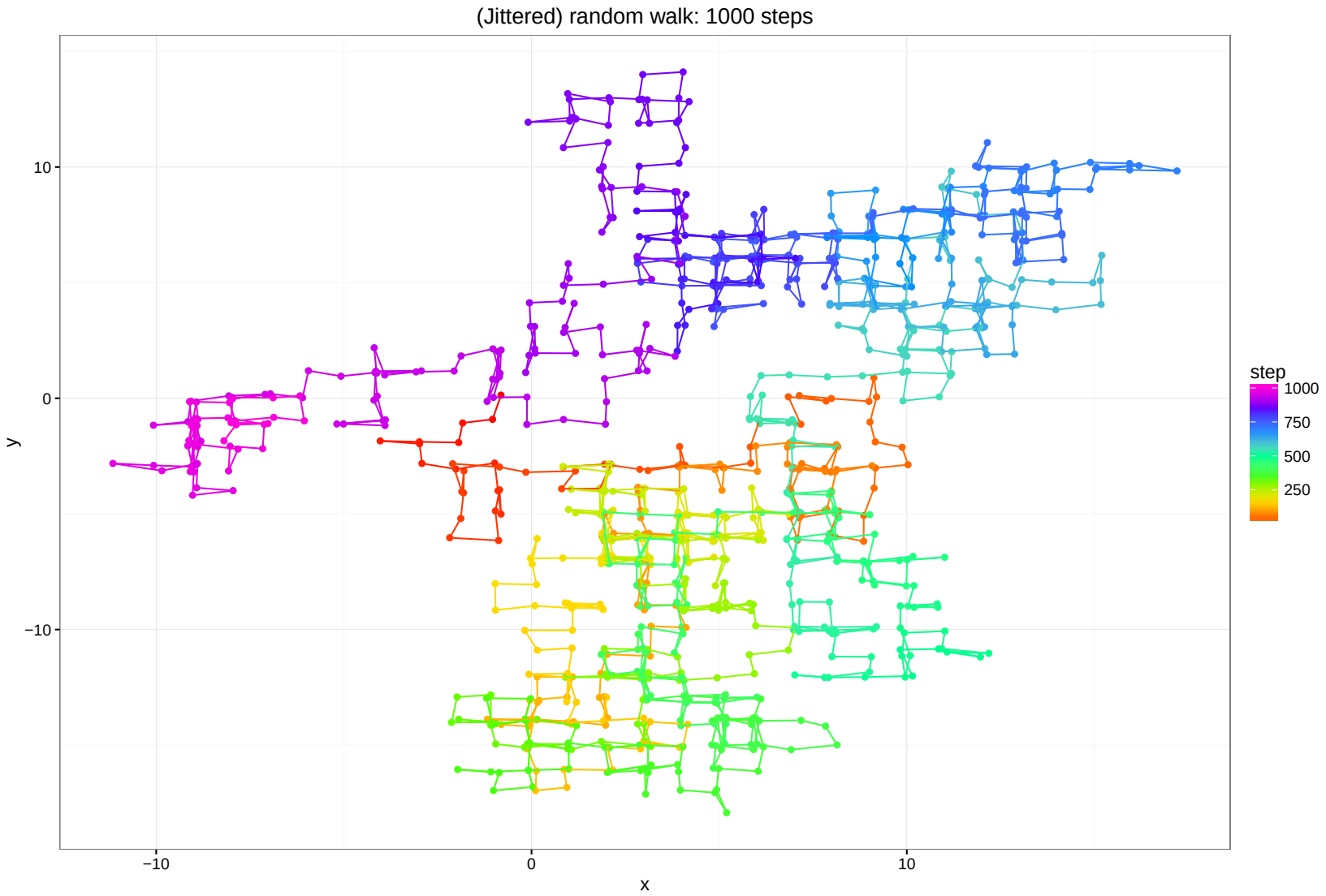
```
R <- 10000                                  #R Wiederholungen
erg <- rep(0,R)                             #Abspeichern ob Bedingung erfuellt oder nicht
n <- 1000
for(i in 1:R){
  x <- sample(c(rep(1,n),rep(-1,n)),size=2*n)
  if(min(cumsum(x))>=0){erg[i] <- 1}        #falls Bedingung erfuellt kommt 1 in Vektor
}
mean(erg)                                   #Prozentsatz (= Schaetzung fuer Wahrscheinlichkeit)
```

- Liefert ca. 0.09
- Für $n = 50$ erhalten wir analog (wie genau?) ca. den Wert 0.0196
- Lässt sich eine Formel erraten?

Problem 2: Random walk in $\mathbb{Z}^2$.

- Ein Partikel befindet sich im Punkt $X_0 = (0, 0)$.
- Mit Wahrscheinlichkeit von jeweils $p = 1/4$ springt es entweder um einen Schritt nach rechts, nach oben, nach links oder nach unten.
- Bezeichnet X_1 die Position nach dem ersten Sprung, dann gilt also $X_1 \in \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$.
- Ausgehend von X_1 springt das Partikel wieder mit Wahrscheinlichkeit von jeweils $p = 1/4$ entweder um einen Schritt nach rechts, um einen Schritt nach oben, um einen Schritt nach links oder um einen Schritt nach unten.
- X_2 bezeichne die Position nach dem zweiten Sprung.
- Allgemein bezeichne X_n die Position nach dem n-ten Sprung.
- Gesucht: Wahrscheinlichkeit dafür, nach $n = 10$ Schritten wieder im Punkt $(0, 0)$ zu landen, i.e. $\mathbb{P}(X_n = (0, 0))$.

Problem 2: Random walk in $\mathbb{Z}^2$



- Lernziel:
  - list & data.frame
  - if, & und logische Abfrage ==
  - Ergänzende Erklärungen in den Übungen

Schritt 1: Simulation eines zufälligen Sprünges

- ```
  jumps<-list(c(1,0),c(0,1),c(-1,0),c(0,-1))      #4 Sprungrichtungen
  richtung<-sample(1:4,1)
  jumps[[richtung]]                                #Sprung 1
  ```

Schritt 2: Simulation eines random walks (i.e.mehrerer zufälliger Sprünge), Speichern der Positionen

- Wiederholen Schritt 1 und merken uns (Zwischenspeichern!) die Position des Partikels nach jedem Sprung.

  ```

## Schritt 2: Simulation eines random walks (i.e.mehrerer zufälliger Sprünge), Speichern der Positionen, cont.

- Zuätzliche Kommentare direkt im R-Code im R-Code 'R-Code_WR01.R' auf http://www.trutschnig.net/courses.
- Generieren random walk der Länge $n = 100$.
- 
```
n<-100
A<-data.frame(x=rep(0,n),y=rep(0,n))
position<-c(0,0)
for(i in 1:n){
  richtung <- sample(1:4,1)
  position <- position + jumps[[richtung]]
  A[i,] <- position
}
plot(A,type="l")
lines(A,type="p")
```

Schritt 3: Simulation mehrerer random walks und approximative Berechnung der Wahrscheinlichkeit

- Simulieren $R = 1000$ random walks und berechnen den Prozentsatz jener, die nach $n = 10$ Schritten in $(0, 0)$ enden.

```
1  R<-1000                                  #Anzahl Wiederholungen
2  retour<-rep(0,R)                         #Abspeichern ob Bedingung
       erfuellt oder nicht
3  for(i in 1:R){
4   n<-10
5   A<-data.frame(x=rep(0,n),y=rep(0,n))
6   position<-c(0,0)                        #Anfangsposition
7   for(k in 1:n){
8     richtung <- sample(1:4,1)
9     position <- position + jumps[[richtung]]      #neue Position
          = alte Position + Sprung
10    A[k,1:2] <- position                          #
          Zwischenspeichern der neuen Position
11  }
12  if(A$x[n]==0 & A$y[n]==0){retour[i] <- 1}       #Falls wieder
       in (0,0) wird 1 in Vektor 'retour' geschrieben
13 }
14 mean(retour)      #Prozentsatz (=Schaetzung fuer
       Wahrscheinlichkeit)
```

Ein bisschen Theorie: Wahrscheinlichkeitsdichten

- Sie kennen die Normalverteilung (Stichwort 'Glockenkurve') schon aus der Schule.
- Das zugrunde liegende Konzept ist jenes der Wahrscheinlichkeitsdichte (kurz: W-Dichte):

Definition (W-Dichte)

Eine Funktion $f : \mathbb{R} \to [0, \infty)$ heisst *W-Dichte* falls $\int_{\mathbb{R}} f(x)dx = 1$.

Definition (Zufallsvariable mit W-Dichte)

Eine Zufallsvariable X hat W-Dichte f (wir schreiben $X \sim f$) wenn für jede 'brave'[1] Menge $A \subseteq \mathbb{R}$ gilt

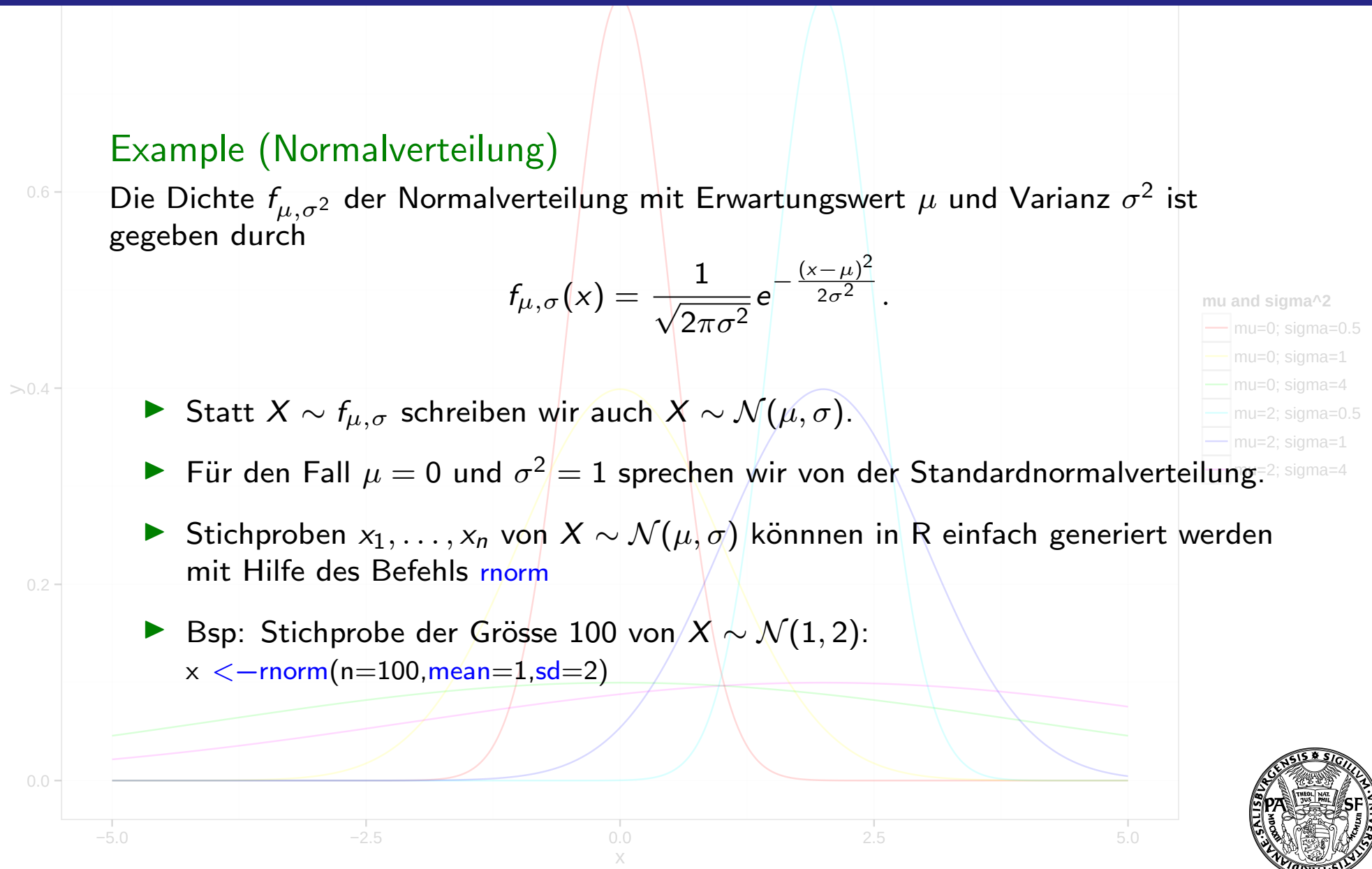
$$\mathbb{P}(X \in A) = \int_A f(x)dx.$$

[1] Mit 'brav' sind hier Borel-Mengen gemeint; jedes Interval ist eine 'brave' Menge

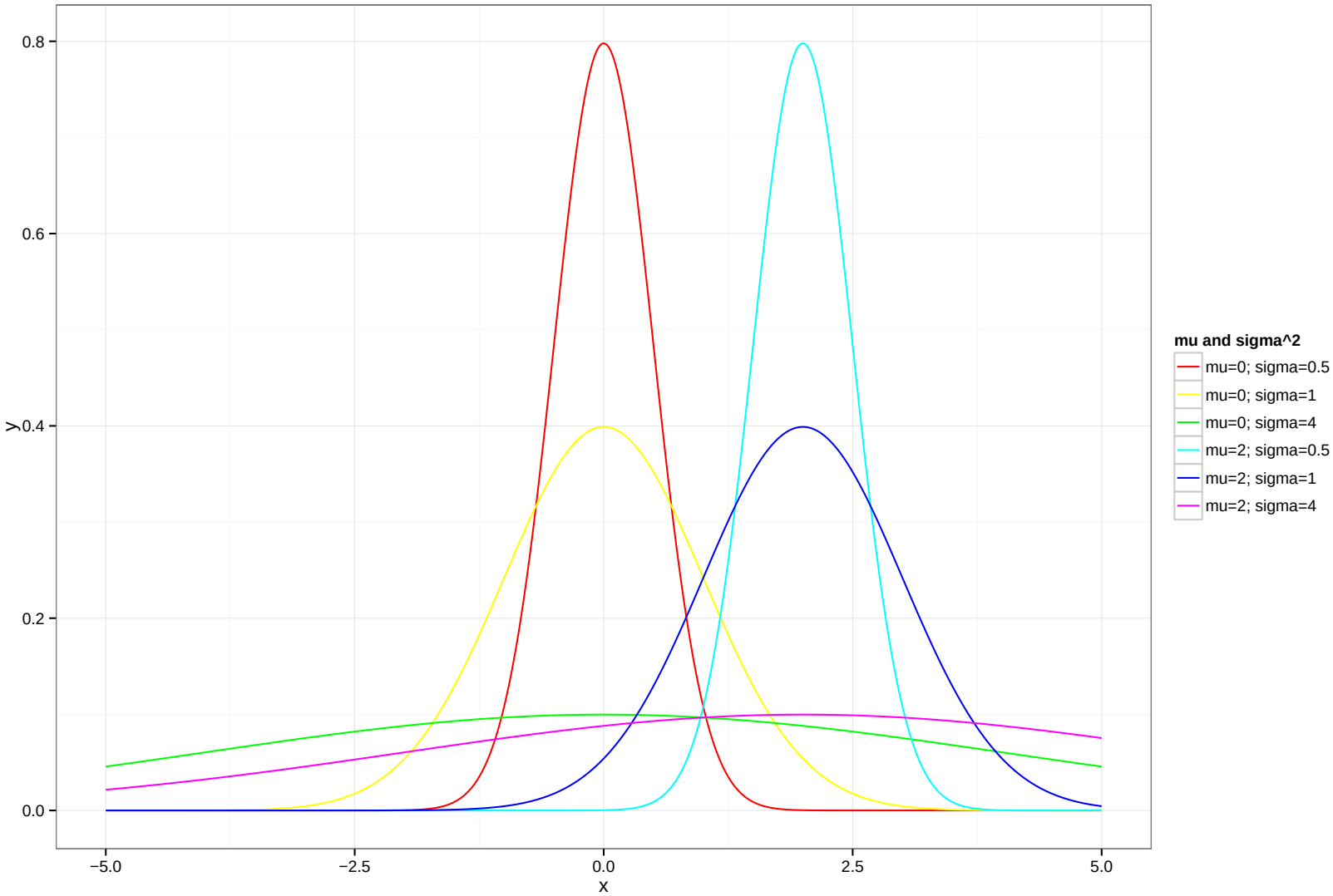
Example (Normalverteilung)

Die Dichte f_{μ,σ^2} der Normalverteilung mit Erwartungswert μ und Varianz σ^2 ist gegeben durch

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$



- Statt $X \sim f_{\mu,\sigma}$ schreiben wir auch $X \sim \mathcal{N}(\mu, \sigma)$.
- Für den Fall $\mu = 0$ und $\sigma^2 = 1$ sprechen wir von der Standardnormalverteilung.
- Stichproben $x_1, \ldots, x_n$ von $X \sim \mathcal{N}(\mu, \sigma)$ könnnen in R einfach generiert werden mit Hilfe des Befehls rnorm
- Bsp: Stichprobe der Grösse 100 von $X \sim \mathcal{N}(1, 2)$:
  x <−rnorm(n=100,mean=1,sd=2)


0.8
0.6
0.4
0.2
0.0
y
−5.0
−2.5
0.0
2.5
5.0
x
mu and sigma^2
mu=0; sigma=0.5
mu=0; sigma=1
mu=0; sigma=4
mu=2; sigma=0.5
mu=2; sigma=1
mu=2; sigma=4

Example (Stetige Gleichverteilung)

Sei $a < b$. Dann heisst $g_{a,b}$, definiert durch

$$g_{a,b}(x) = \mathbf{1}_{[a,b]}(x) = \begin{cases} 1 & \text{if } x \in [a,b], \\ 0 & \text{if } x \notin [a,b]. \end{cases}$$

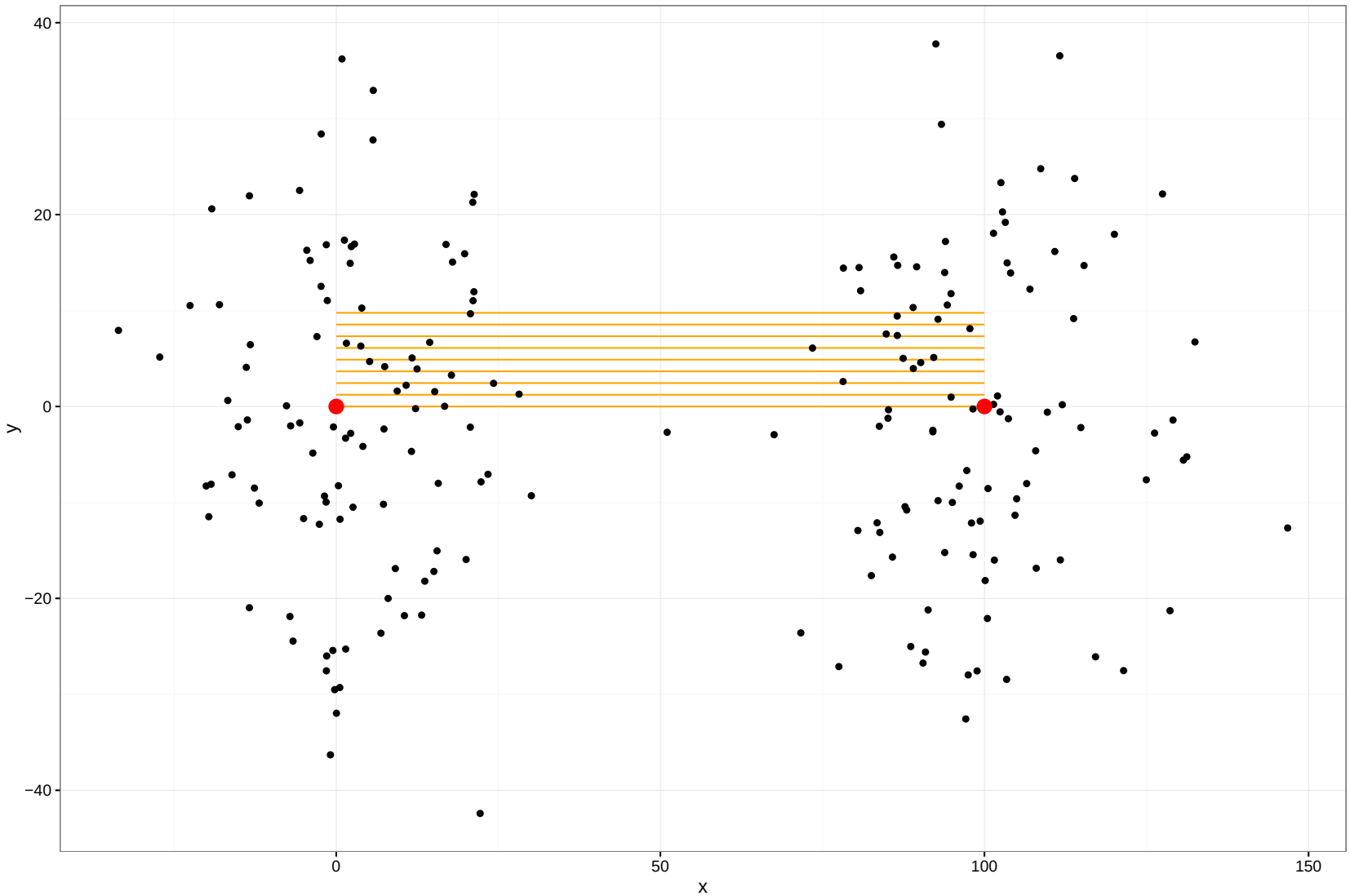
W-Dichte der stetigen Gleichverteilung auf dem Intervall $[a, b]$.

- Statt $X \sim g_{a,b}$ schreiben wir auch $X \sim \mathcal{U}(a,b)$.
- Stichproben $x_1, \ldots, x_n$ von $X \sim \mathcal{U}(a,b)$ könnnen in R einfach generiert werden mit Hilfe des Befehls runif.
- Bsp: Stichprobe der Grösse 100 von $X \sim \mathcal{U}(1,2)$:
  x <−runif(n=100,min=1,max=2).

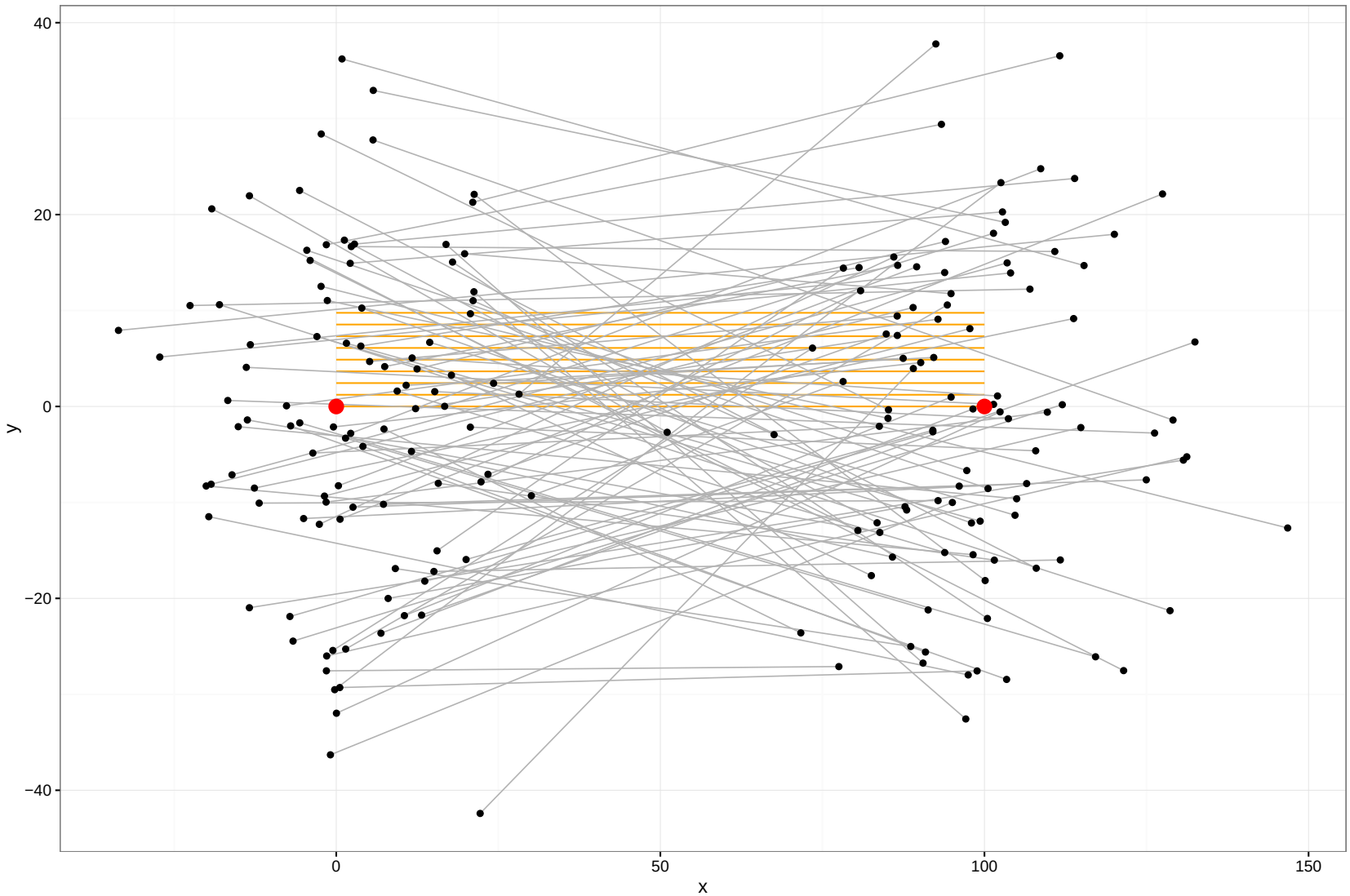
Problem 3: GPS-Verzerrung?

- Zehn Geoinformatik-Studenten möchten die Genauigkeit GPS-basierter Abstandsmessungen testen.
- Sie erfassen (hintereinander) die GPS-Koordinaten der Startlinie des 100m-Sprints auf der äussersten Bahn im benachbarten Leichtathletikstadium, gehen dann (hintereinander) die Bahn entlang bis zur Ziellinie, und erfassen dort wiederum die GPS Koordinaten.
- Jeder von ihnen wiederholt das Experiment 50 Mal.
- Für die so erhaltenen 500 Koordinatenpaare berechnen sie jeweils den Abstand der Punkte.
- Sie erwarten, dass der Mittelwert der Abstände ziemlich genau 100m ist (Fehler heben sich auf).
- Umso grösser ist die Überraschung, als der berechnete Mittelwert etwas über 102m ist.
- Wie kann das sein - just bad luck?

Problem 3: GPS-Verzerrung?



Problem 3: GPS-Verzerrung?



- Wie kann das sein - just bad luck (cont.)?
- Beantworten die Frage mit Hilfe von Simulationen.
- O.B.d.A. nehmen wir an, dass der Startpunkt S und der Zielpunkt Z folgende Koordinaten haben: $S = (0, 0), Z = (100, 0)$
- Die fehlerbehafteten Messungen in S und Z bezeichnen wir mit S', Z', den Fehler betreffend S mit $F = (X_1, Y_1)$ und jenen betreffend Z mit $G = (X_2, Y_2)$.
- Es gilt also

$$\begin{aligned} S' &= S + (X_1, Y_1) = (X_1, Y_1) \\ Z' &= Z + (X_2, Y_2) = (100 + X_2, Y_2). \end{aligned}$$

- Der gemessene Abstand ist damit

$$\|S' - Z'\|_2 = \sqrt{(100 + X_2 - X_1)^2 + (Y_2 - Y_1)^2}.$$

- Nehmen an, dass alle Fehler normalverteilt sind, genauer $X_1, X_2, Y_2, Y_2 \sim \mathcal{N}(0, 15)$.
- Simulieren nicht nur 500, sondern gleich 10.000 (oder noch mehr) Abstandsmessungen.
- Direkte Umsetzung inkl. Kommentare siehe 'R-Code_WR01.R' auf http://www.trutschnig.net/courses.
- Wir erhalten tatsächlich ca. 102,2 als mittleren Abstand.
- Was passiert, wenn die Varianz der Fehler vergrössert oder verkleinert wird?
- Wie kann diese Überschätzung erklärt werden?
- Bekanntes Phänomen?

[1] P. Ranacher, R. Brunauer, W. Trutschnig, S. Van der Spek, S. Reich: *Why GPS makes distances bigger than they are*, International Journal of Geographical Information Science **30**, 316-333 (2016)

Problem 4: Wartezeit - Aufgabe am Übungsblatt

- Die U1 in Wien fährt innerhalb der Stosszeiten von 17-20h lt. Plan (in die von Ihnen gewünschte Richtung) alle zwei Minuten, i.e. um 17:00, 17:02, 17:04,..., 19:58, 20:00.
- Sie kommen werktags gemäss einer stetigen Gleichverteilung im Intervall 18:00-19:00 am Bahnsteig an und würden erwarten, im Mittel eine Minute warten zu müssen.
- Von der Statistik begeistert notieren Sie über einen langen Zeitraum hinweg Ihre Wartezeiten und stellen fest, dass Sie im Mittel etwas länger als eine Minute warten müssen.
- Wie kann das sein - just bad luck?
- Parallele zur GPS-Verzerrung?