Fortschritt der Promotionsarbeit: Erste deskriptive Statistik

Treffen: 26. Juni 2017

Letztes Treffen: 30.05.2017

Lea Oelkers

Thema der Promotion:

Pädiatrische LH- und FSH-Serumwerte und Bezüge zum BMI und Sozialstatus.

Kurze Inhaltsangabe:

1. 4 Gruppen von Datensätzen
2. Beschreibung der Bereinigung des Datensatzes der Gruppe 1
3. Erste deskriptive Statistik des Datensatzes der Gruppe 1
4. 4 Gruppen von Datensätzen mit unterschiedlichen Ausschlusskriterien

Ich habe aufgrund von meist sehr ungenauen Angaben zu Krankheiten und Medikamenten in der Anamnese vorerst 4 Gruppen von Datensätzen gebildet, in denen ich unterschiedlich streng bezüglich der Ausschlusskriterien vorgehe. Dementsprechend kleiner/größer sind die Datensätze der Gruppen. Die Krankheits- und Medikamentenausschlüsse sind als „FLAGs“ aufgeführt, also so, wie sie als Abkürzungen im Datensatz bezeichnet wurden. Der Vergleich der Analysen der Datensätze wird zeigen, inwiefern dies Sinn macht. Meine bisherigen ersten statistischen Analysen habe ich vorerst mit der Gruppe 1 durchgeführt. Auch wird genau aufgeführt werden, wie die Bereinigung des Datensatzes der Gruppe 1 vorgenommen wurde.

|  |  |  |
| --- | --- | --- |
| **Gruppe 1**: nur „gesunde“ Kinder  🡪n=963 | **Gruppe 2**: Ausgeschlossen werden u.a. Frühgeborene, endokrinologische/schwere chronische Erkrankungen  🡪 n= 1201 | **Gruppe 3**🡪 ausgeschlossen sind u.a. die Frühgeborenen, wenige schwere Erkrankungen und Kontrazeptiva🡪n= 1296  **Gruppe 4** : inklusive Kontrazeptiva 🡪n=1320 |
| Ausschlusskriterien |  |  |
| * Fehlender Pubertätsstatus * Mehrfachbesucher   Krankheiten:   * C\_DISEASE\_TX\_SD\_ALLG * C\_DISEASE\_TX\_SD\_HYPER * C\_DISEASE\_TX\_SD\_HYPO * C\_DISEASE\_TX\_FRUEHGEB * C\_DISEASE\_TX\_ENDOKR   (darunter z.B.: Café au lait Flecken ( da unklare Ätiologie), PCO, gestörte Glukosetoleranz, AGS, Diabetes typ 1, Hypoglykämien, SD-Dysfunktion, Homocystinurie,…)   * C\_DISEASE\_TX\_DM1 * C\_DISEASE\_TX\_DM2 * C\_DISEASE\_TX\_BLUT * C\_DISEASE\_TX\_GERIN C\_DISEASE\_TX\_PULMO * C\_DISEASE\_TX\_ASTHMA * C\_DISEASE\_TX\_GIT * C\_DISEASE\_TX\_INFEKT * C\_DISEASE\_TX\_HWI\_PN * C\_DISEASE\_TX\_NIERENFEHL * C\_DISEASE\_TX\_KARDIO\_RHYTH * C\_DISEASE\_TX\_KARDIO * C\_DISEASE\_TX\_ALLERG * C\_DISEASE\_TX\_EPIKRAMPF * C\_DISEASE\_TX\_PSY * C\_DISEASE\_TX\_ADHS * C\_DISEASE\_TX\_DEPRES * C\_DISEASE\_TX\_SUCHT * C\_DISEASE\_TX\_MUSKEL * **Medikamente:** Glukokortikoide, Wachstumshormon, Testosteron, Immunsuppressiva, Neuroleptika, Antipsychotika, „ADHS“, Hormone, „Stoffwechsel“, Metformin, Insulin, L-thyroxin, Kontrazeptiva, Sexualsteroide, Antibiotika, „Allergie“, Antihistamin, „Hyposensibiliseirung“, Mineralocorticoide * + individuell ausgeschlossene **Freitexte** | * Fehlender Pubertätsstatus * Mehrfachbesuche   Krankheiten:   * C\_DISEASE\_TX\_SD\_ALLG * C\_DISEASE\_TX\_SD\_HYPER (1) * C\_DISEASE\_TX\_SD\_HYPO (44) * C\_DISEASE\_TX\_FRUEHGEB (41 * C\_DISEASE\_TX\_ENDOKR (18) * C\_DISEASE\_TX\_**DM1** (4) * C\_DISEASE\_TX\_**DM2** (13) * C\_DISEASE\_TX\_BLUT (35) * C\_DISEASE\_TX\_GERIN (12) * C\_DISEASE\_TX\_EPIKRAMPF * Asthma: nur Glukokortikoid-Behandlungen * **Medikamente:** Glukokortikoide, Insulin, Metformin, L-Thyroxin, Stoffwechsel, Hormone, Wachstumshormon, Immunsuppression, Neuroleptika, Kontrazeptiva, Testosteron   **Freitexte**   * Frühgeburt (5)*,* SD-Dysfunktion (1),unklare Anamnese „ais“ (1), Morbus Crohn (1), Tumor (3), Long QT-Syndrom (1),   Blut/Gerinnung (2)   * Medis:   Insulin (1), L-Tyroxin (1) | Gruppe3:   * Fehlender Pubertätsstatus * Mehrfachbesuche * C\_DISEASE\_TX\_FRUEHGEB * Medikamente: Glukokortikoide, Wachstums-   Hormon, Testosteron, Kontrazeptiva,  Immunsuppressiva   * Freitext: Tumor, Frühgeburt   **Gruppe 4: wie Gruppe 3, außer dass Kontrazeptiva (31) und Testosteron (2) eingeschlossen bleiben.** |
| Kein Ausschlusskriterium |  |  |
| * C\_DISEASE\_TX\_ADIP * C\_DISEASE\_TX\_AUGE * C\_DISEASE\_TX\_HNO * C\_DISEASE\_TX\_MOE (Mittelohrentz.) * C\_DISEASE\_TX\_NEURODER * C\_DISEASE\_TX\_SOOR (43) * C\_DISEASE\_TX\_DERMA * C\_DISEASE\_TX\_KOPFSCH * C\_DISEASE\_TX\_SPRACH * C\_DISEASE\_TX\_ORTHOP * C\_DISEASE\_TX\_OP * C\_DISEASE\_TX\_HERNIE * C\_DISEASE\_TX\_KH * C\_DISEASE\_TX\_NEPHRO | Mineralokortikoide |  |

1. Beschreibung der Bereinigung des Datensatzes der GRUPPE 1

In dieser Gruppe wird am meisten ausgeschlossen. Die Anamnesen im Gesamtdatensatz sind teils sehr ungenau. Durch das „radikale“ Ausschließen ist die Sicherheit, dass Krankheiten (deren Schwere meist unbekannt ist) die folgenden statistischen Berechnungen beeinflussen, minimiert. Es werden vom Ausgangsdatensatz n=2907 Besuchen, bei denen LH und FSH gemessen wurde, insgesamt 1944 Besuche ausgeschlossen, sodass sich der Datensatz nach der Bereinigung auf n=963 reduziert. Im Folgenden wird dokumentiert, wie der Datensatz der Gruppe 1 bereinigt wurde. Die Krankheits- und Medikamentenausschlüsse sind teilweise als „FLAGs“ aufgeführt, also so, wie sie als Abkürzungen im Datensatz bezeichnet wurden.

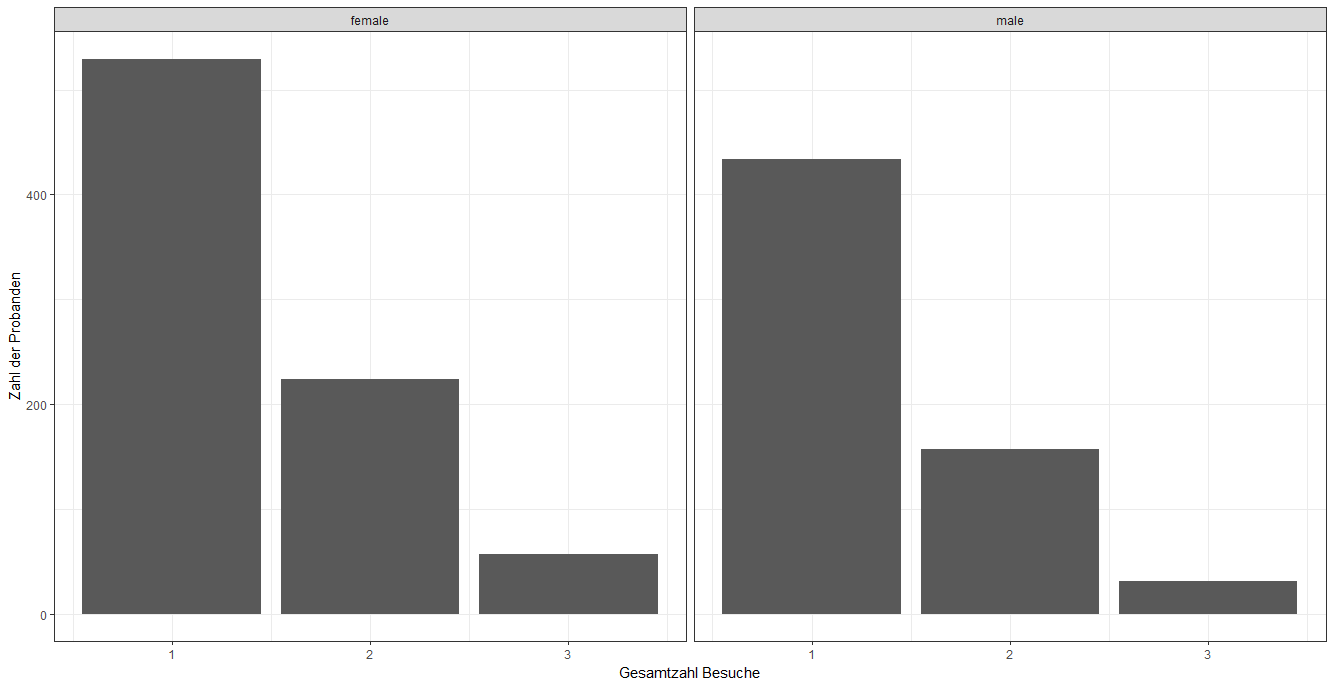
**Vor den Ausschlüssen: 2907 Besuche, davon:**

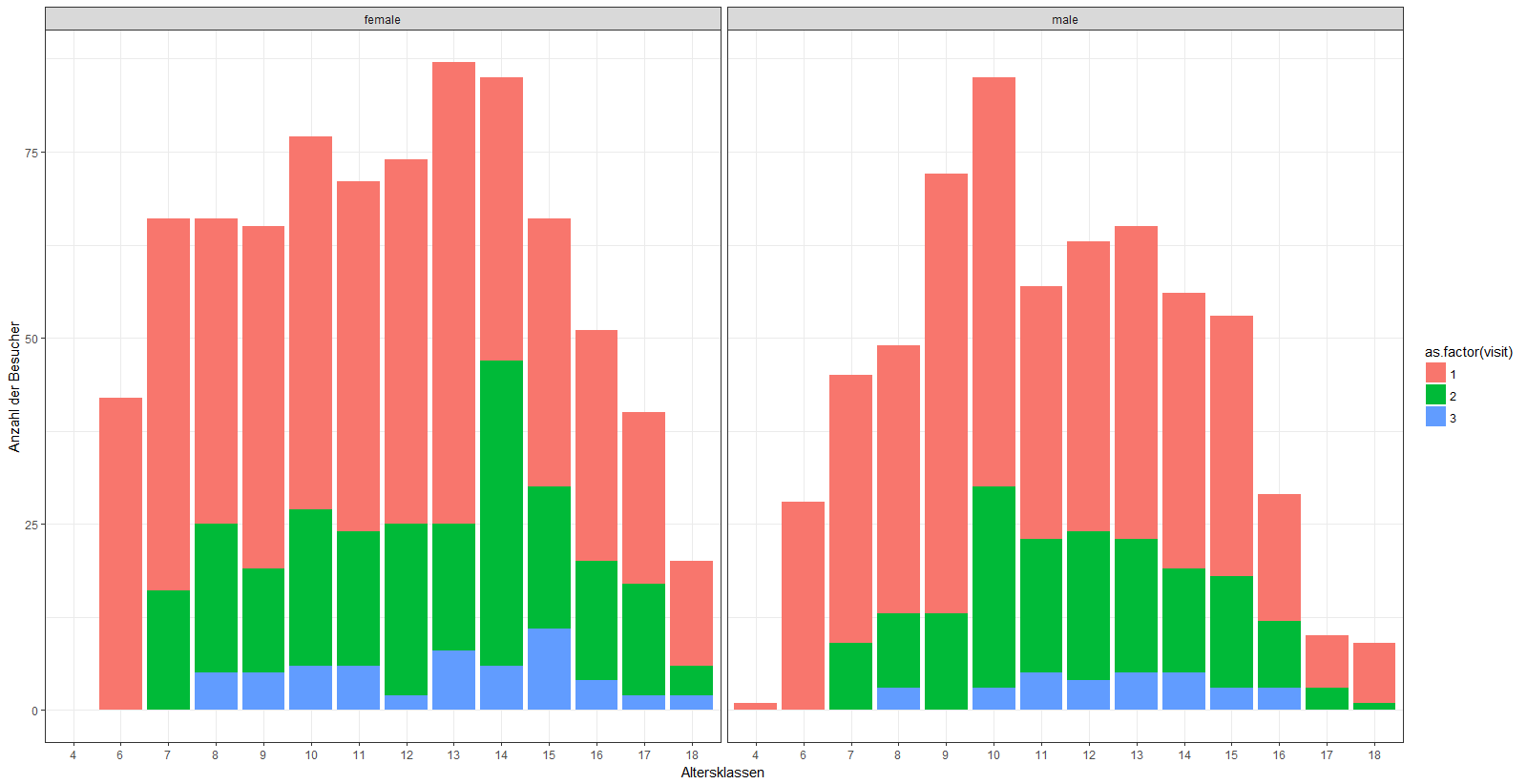
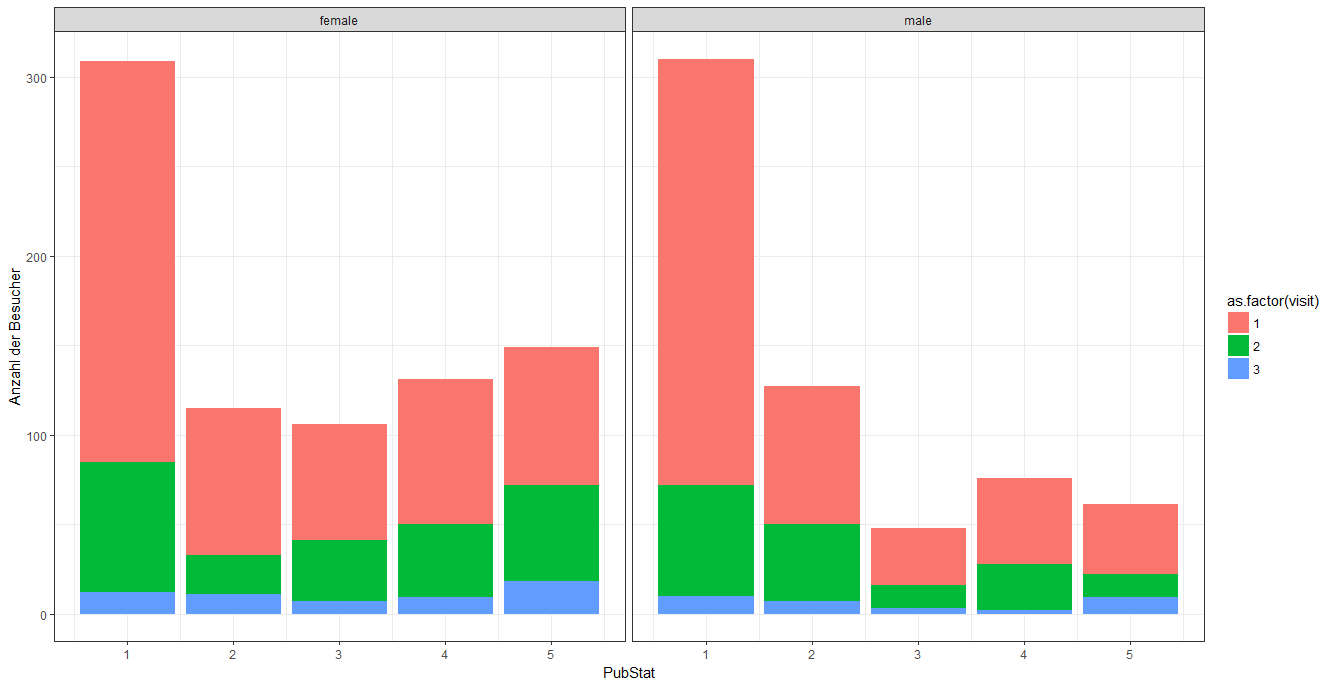
|  |
| --- |
| * Fehlenden Angabe zum Pubertätsstatus: 647 Besuche * Krankheiten   1 SD\_Allg 14  2 SD\_Hypo 23  3 Sd\_Hyper 3  4 Fruehgeb 1  5 Endokr 6  6 DM1 2  7 DM2 7  8 Blut 14  9 Gerinnung 9  10 Pulmo 71  11 Asthma 218  12 GIT 32  13 Infekt 24  14 HWI 9  15 Nierenfehl 17  16 Kardio\_Rhyth 28  17 Kardio 17  18 Allergie 281  19 Epikrampf 12  20 Psy 51  21 ADHS 38  22 Depres 14  23 Sucht 8  24 Muskel 6   * Medikamente |
| 1 Gluko 155  2 Asthma 173  3 Wachstumshormon 8  4 Immnunsuppr 4  5 Neuroleptika 59  6 Antipsychotika 3  7 ADHS 45  8 Hormone 135  9 Stoffwechsel 9  10 Metformin 10  11 Insulin 4  12 L-Tyroxin 12  13 Kontrazeptiva 80  14 Sexualsteroide 14  15 Testosteron 2  16 Antibiotika 40  17 Allergie 151  18 Antihis 108  19 Hyposensibilisierung 19  20 Mineralocorticoide 20  Dann werden noch ausgeschlossen:   * Zusätzliche auffällige Freitexte (die nicht in den „Krankheits-/Medikamenten-Flags“ markiert sind):   62 Besucher, 55 SICs, insgesamt also 55 Ausschlüsse  🡪*vor Ausschluss der Mehrfachbesucher habe ich deskriptive Statistik mit noch eingeschlossenen Mehrfachbesuchen*  *durchgeführt (siehe Dokument „Deskriptive Statistik Gruppe 1) .*  Ausgeschlossene Freitexte   |  |  | | --- | --- | | Ausschlusskriterium | Anzahl der Besucher | | Frühgeburt | 6 | | Infekt | 19 | | Endokrinologische Erkrankung   * AGS, Lupus, Hyperinsulinämie | 3 | | Blut, Gerinnung   * Petechien, von Willebrand-Syndrom | 2 | | Sonstiges   * Tumor, Cafe au lait Fleck, Nicht eindeutige Anamnese („ais“, „multiple lentigenes“), Zigaretten/Tag (1), offenes Bauchtrauma, Purpura Schönlein-Hennoch(1), Zyste im Koff des Fetus (1), neurolog. Grunderkrankung mit statomot. Entwicklungsverzögerung(1), Hemiparese (1) | 16 | | Kardio\_Rhyth | 5 | | Psychische Erkrankung | 1 | | Gastrointestinale Probleme   * Fructose-, Lactoseintoleranz, Gastritis | 8 | | Medikamente:   * Antibiotika (1), L-Thyroxin (1) | 2 |  * Mehrfachbesuche : 469   Graphische Darstellung der Ausschlusskriterien |
| |  | | --- | |  |  * Frage: Sollen Mehrfachbesuche ausgeschlossen werden?   Ausschlüsse   1. Fehlender Pubertätsstatus: 646 Ausschlüsse 🡪 verbleibende Besuche: 2260 2. Krankheiten: noch 532 Ausschlüsse 🡪 Verbleibende Besuche: (2260-532=)1728 3. Medikamente: noch 241 Ausschlüsse 🡪Verbleibende Besuche: 1487 4. Freitexte: 55 🡪 verbleibende Besuche: 1432 5. Mehrfachbesuche: 469 Ausschlüsse 🡪 Verbleibende Erstbesuche: **n=** **963.**   Graphische Darstellung der Bereinigung des Datensatzes der Gruppe 1 |

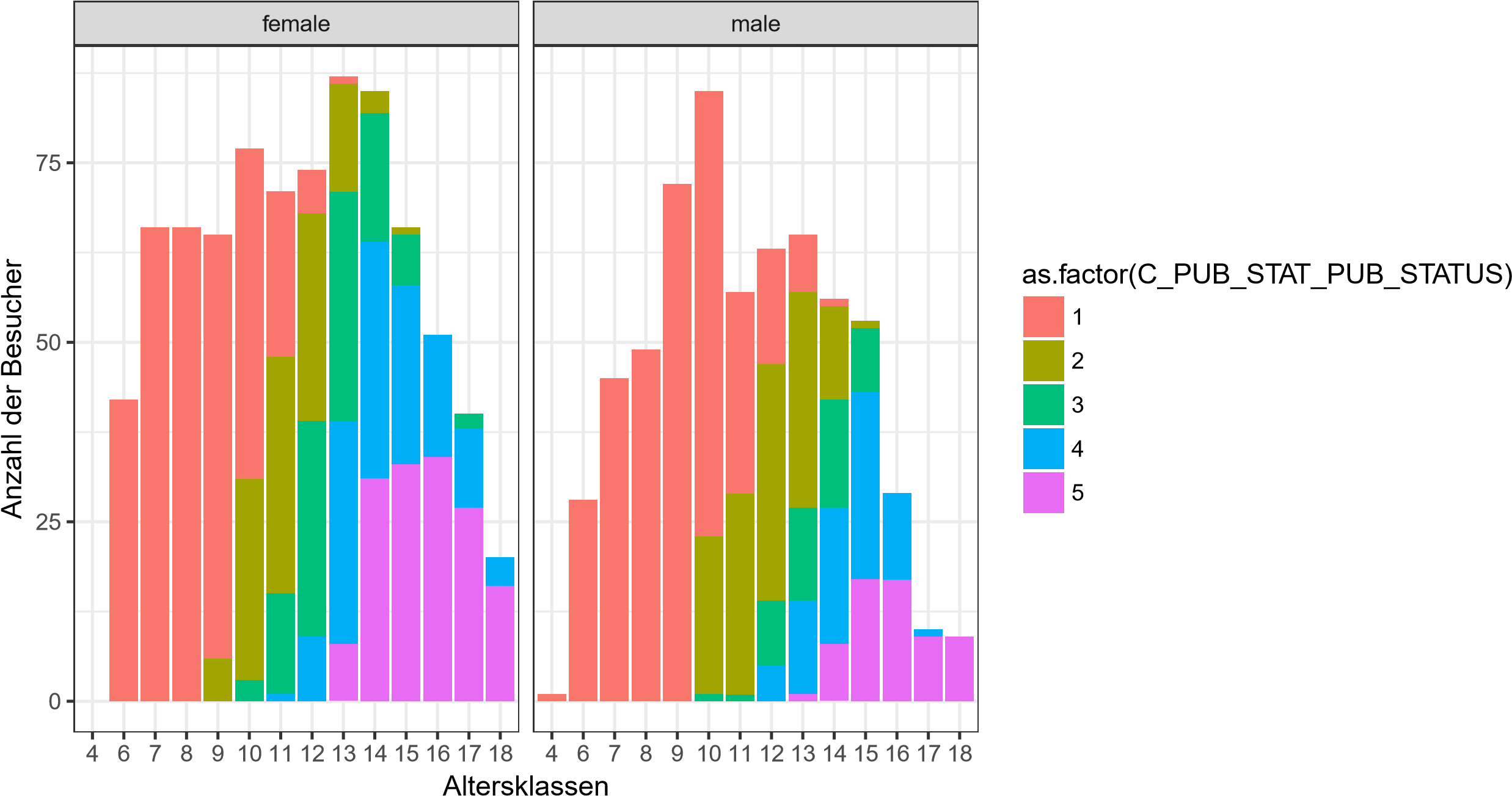
1. Erste deskriptive Statistik des Datensatzes der Gruppe 1
2. Deskriptive Statistik des Datensatzes **inklusive der Mehrfachbesucher** (n=1432)

* Geschlechterverteilung: 810 Mädchen, 522 Jungen

Wie oft kamen die Probanden? Das folgende Diagramm zeigt die Anzahl der Probanden nach der Gesamtzahl der Besuche.



* Wie viele Besucher waren je nach Geschlecht in den unterschiedlichen Altersklassen da? War es ihr Erst-, Zweit- oder Drittbesuch? Wie groß ist der Anteil an Erst-, Zweit- oder Drittbesuchern?
* Wie viele Besucher waren in den jeweiligen Pubertätsstadien da? War es ihr Erst-, Zweit- oder Drittbesuch im jeweiligen Pubertätsstadium?
* Wie viele Besucher waren in den unterschiedlichen Altersklassen da und ich welchem Pubertätsstadium befanden sie sich? Wie unterscheiden sich die Geschlechter diesbezüglich?



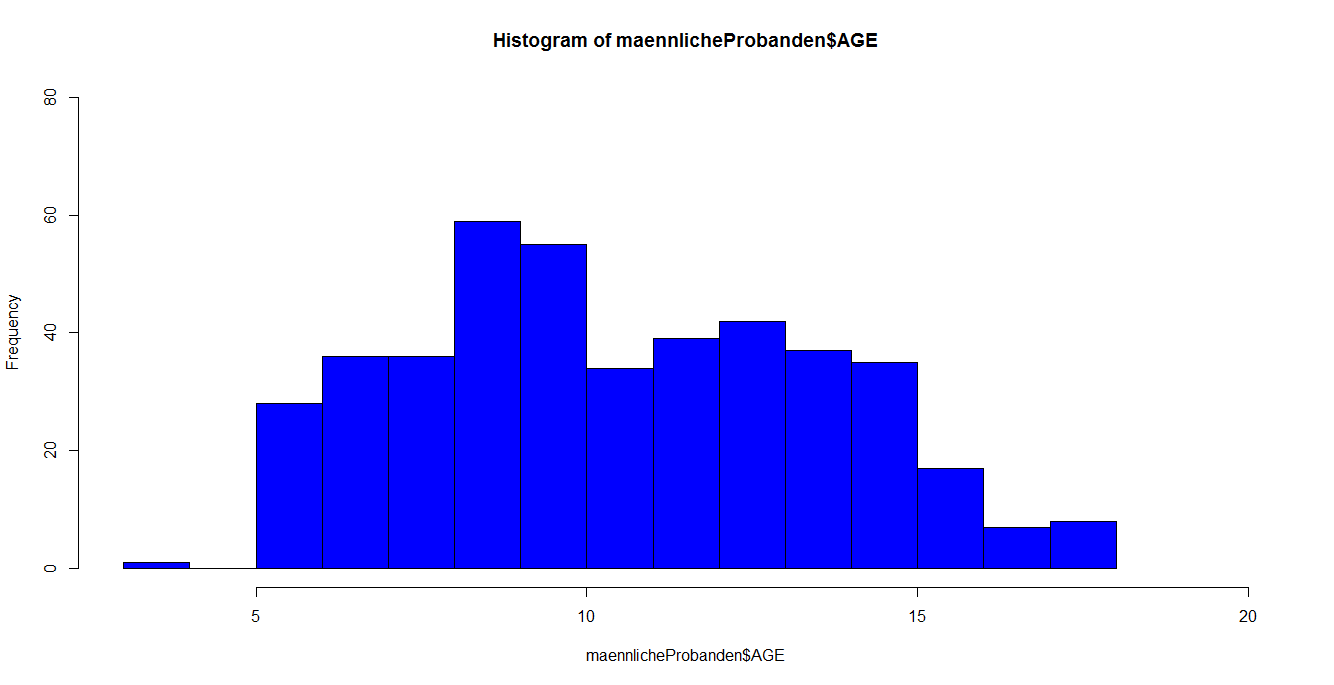
**2. Nun betrachte ich** **nur die Erstbesucher** (n=963).

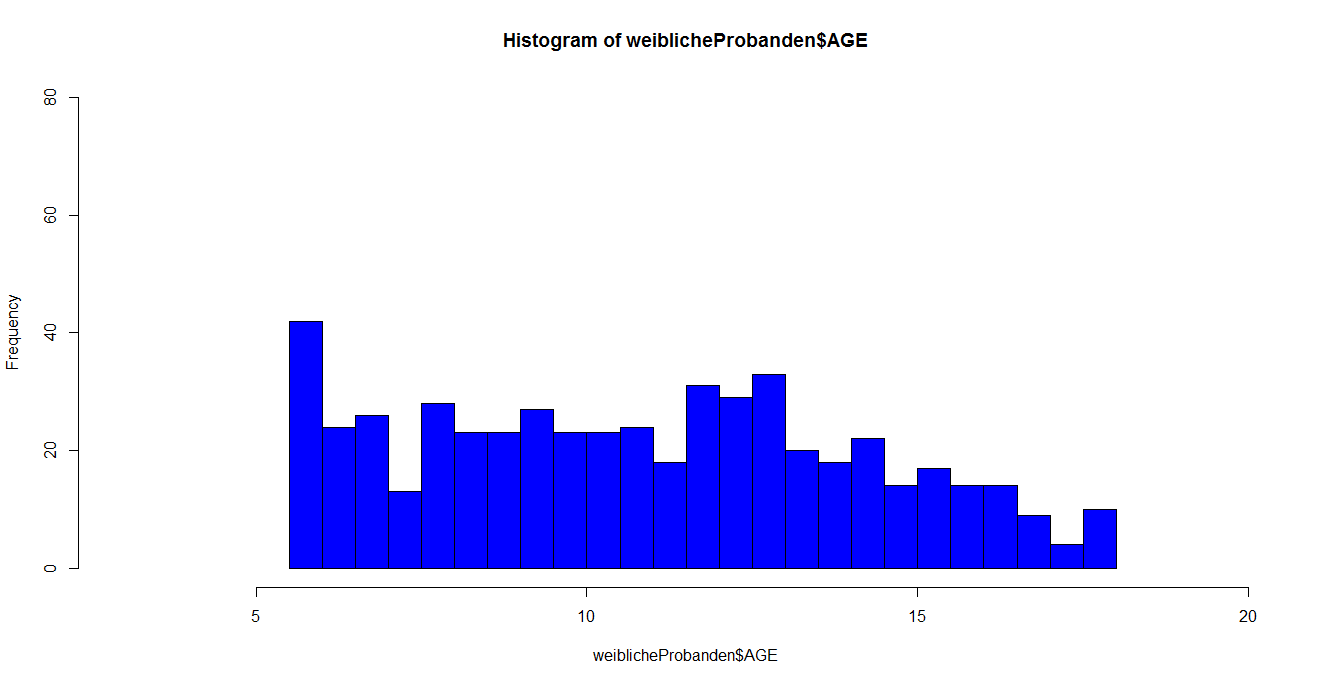
Geschlechterverteilung: 529 Mädchen, 434 Jungen

Wie stellt sich die Altersverteilung der Probanden dar?

* Druchschnittsalter Mädchen: 10,8 yr Durchschnittalter Jungen: 10,47yr
* Median Mädchen : 10,71yr Median Jungen : 10,71yr
* Standartabweichung : 3,3 Standartabweichung :3,03

Diagramme zur Veranschaulichung der Altersverteilung:





Wieviele Probanden befinden sich in den unterschiedlichen Pubertätsstadien?

|  |
| --- |
| Pubertätsstadium 1 2 3 4 5 |
| female 224 82 65 81 77 |
| male 238 77 32 48 39 |

* als Säulendiagramm: Anzahl der Probanden nach Pubertätsstatus („PubStat“), nach Geschlecht getrennt:

female

male

1

2

3

4

5

1

2

3

4

5

0

50

100

150

200

PubStat

Anzahl der Probanden

* als Säulendiagramm: Anzahl der Probanden nach Pubertätsstatus („PubStat“), nicht nach Geschlechtern getrennt:

0

100

200

300

400

1

2

3

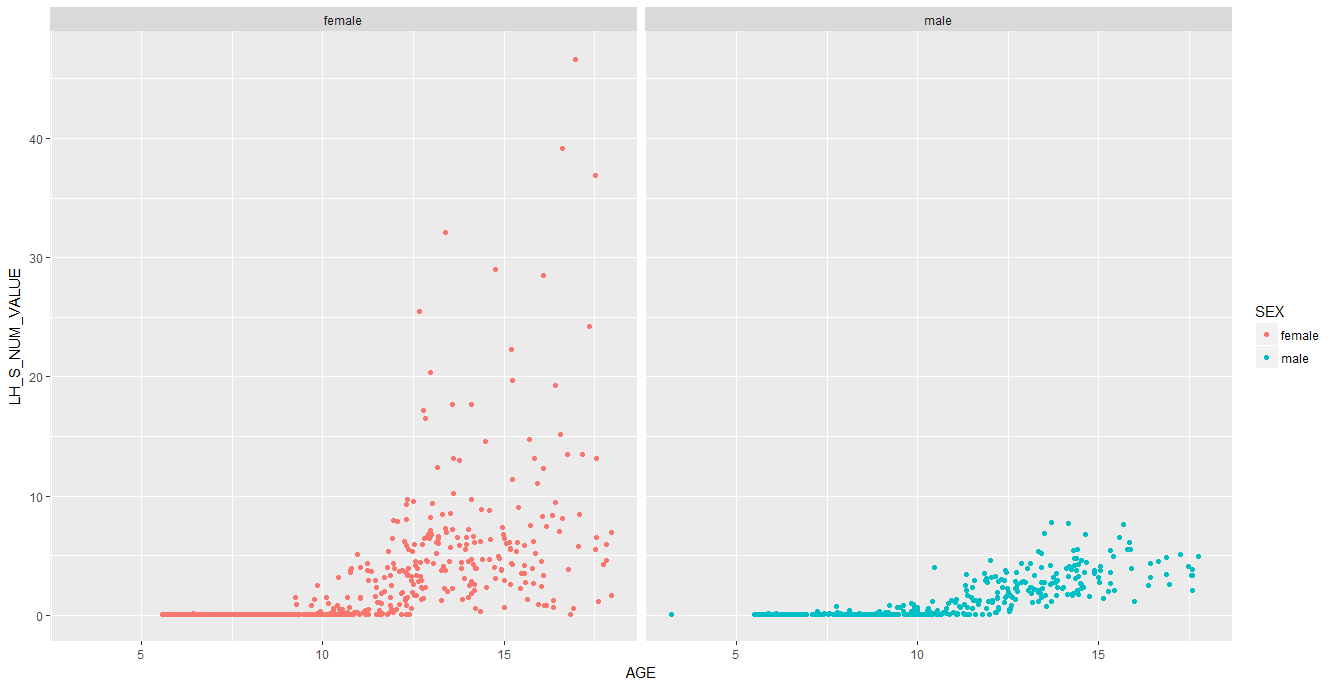
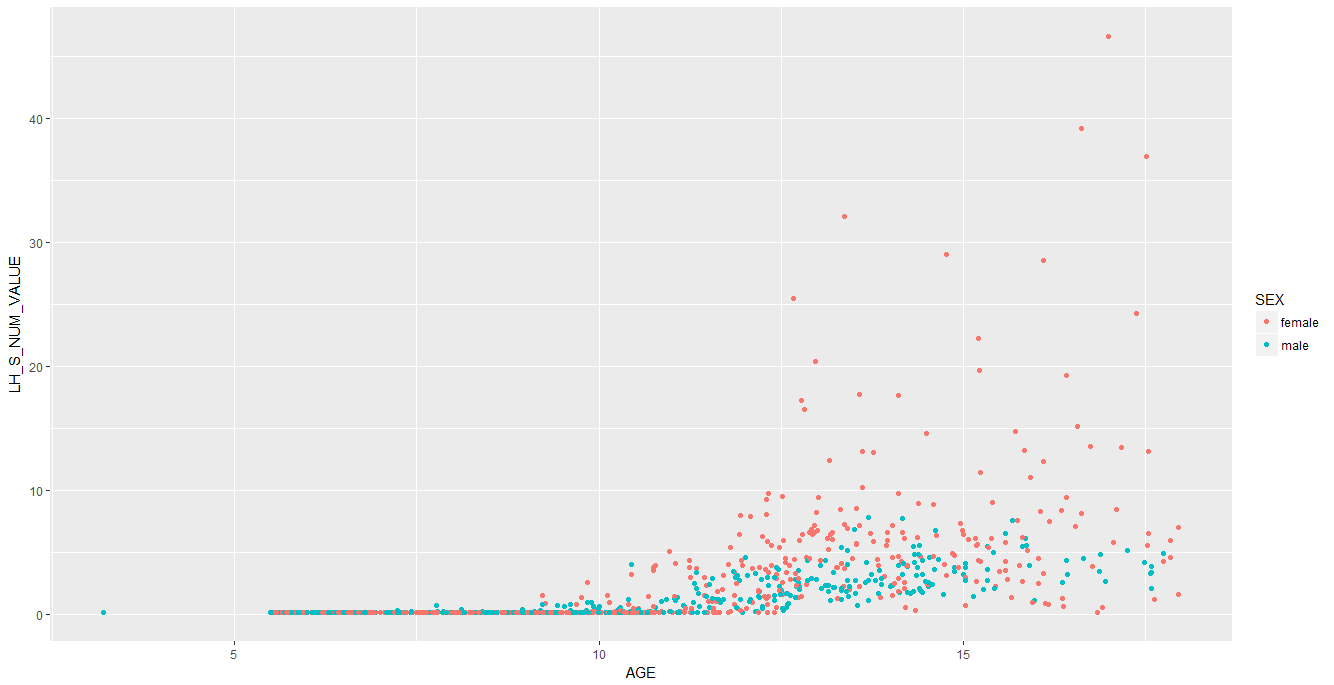
4

5

PubStat

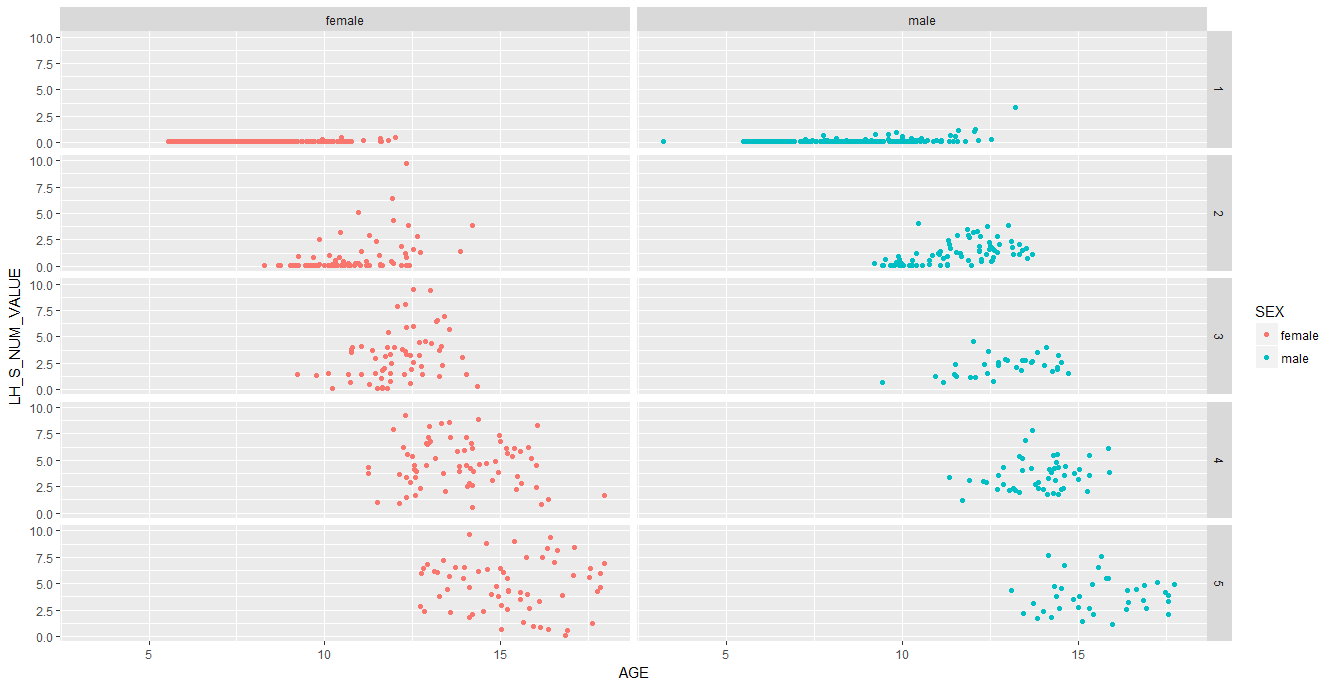
Anzahl der Probanden

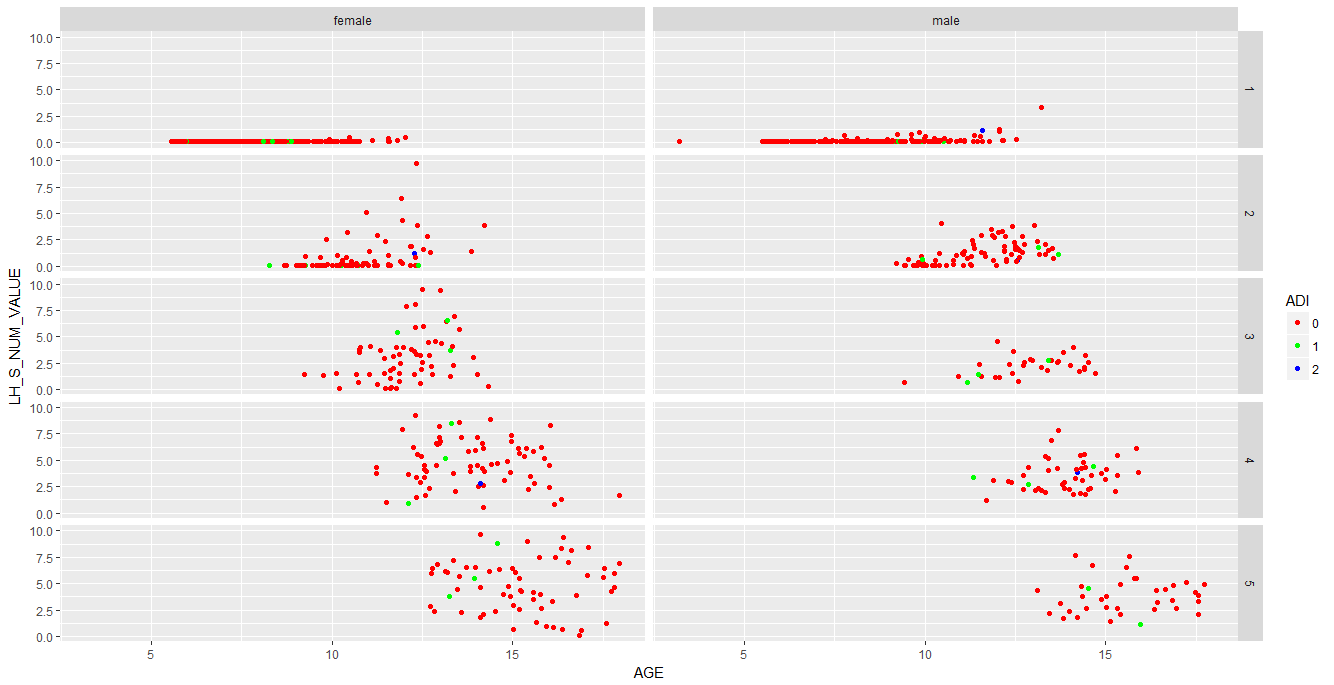
* Wie stellt sich geschlechtsspezifisch die LH-Verteilung bezüglich des Alters der gemessenen Probanden dar?



* Wie stellt sich die geschlechtsspezifische LH-Verteilung entlang des Alters bezüglich der Pubertätsstadien dar?

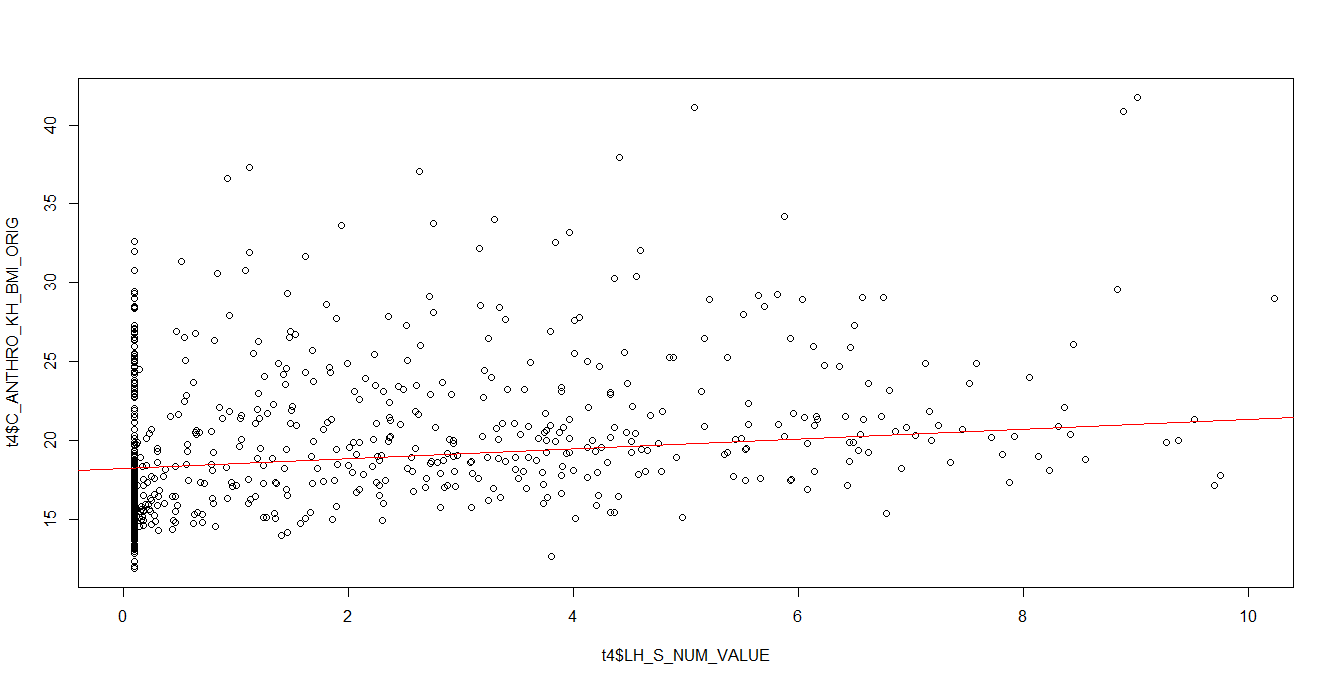


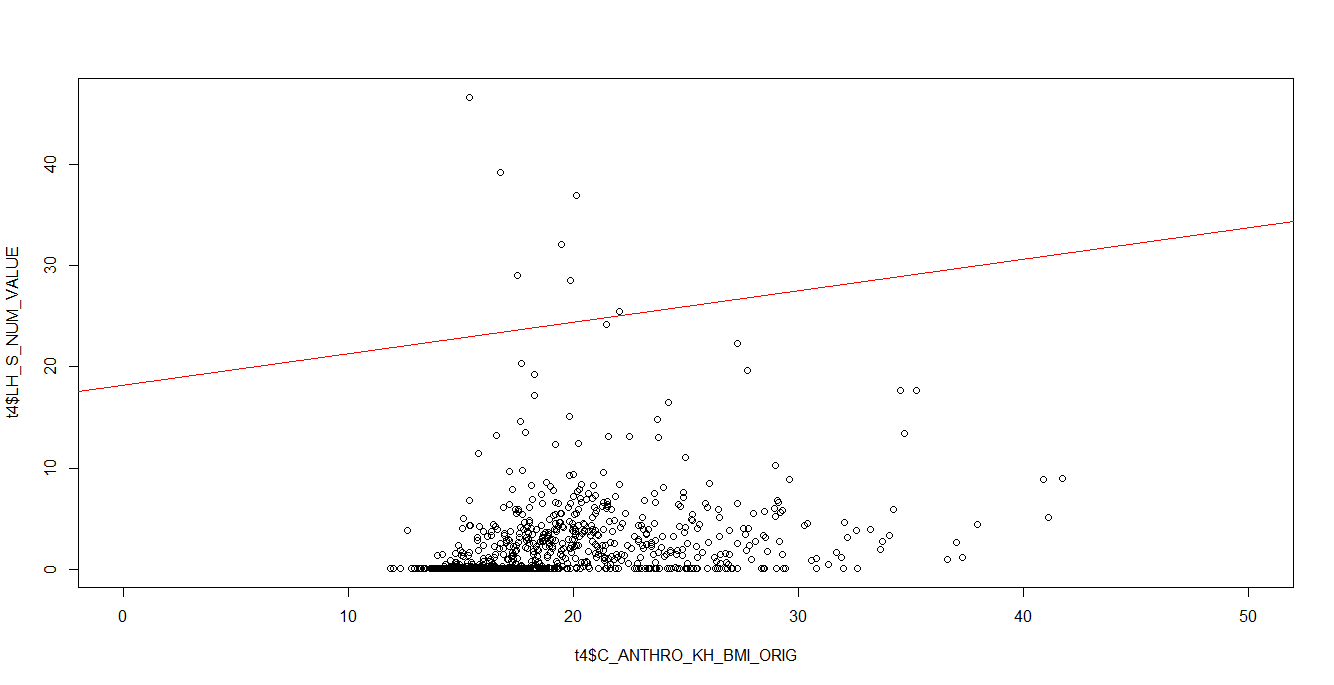
* Oder mit einer limitierteren Achse der LH-Werte :
* Die folgende Abbildung zeigt zusätzlich markierte Probanden, die als adipös eingestuft wurden *(in dieser Abbildung sind nur diejenigen markiert (46 von 963 Probanden), deren „Adipositas-Krankheitsflag“ markiert war. Es gibt sehr viel mehr übergewichtige und adipöse Probanden, deren Gewicht sich im BMI wiederspiegelt, die aber nicht als adipös „markiert“ sind. Sinnvoller wird also die Verwendung des BMIs sein. Auch die Einstufung „wahrscheinlich adipös“ wird überarbeitet werden.*).
* Erklärung zum Plot: ADI= Adipöse/Übergewichtige Probanden, 0=nicht adipös, 1= adipös/übergewichtig, 2=wahrscheinlich adipös/übergewichtig



* Die Probanden wurden bei LIFE in eine A-Kohorte (normalgewichtige) und in eine B-Kohorte (übergewichtige/adipöse) eingeteilt. In folgender Abbildung ist analog zu den vorigen Abbildungen die A- und die B-Kohorte bezüglich ihrer LH-Verteilung entlang des Alters und der Pubertätsstadien markiert. *(Auch diese Abbildung hält einer weiteren Überprüfung kaum stand, da auch in der A-Kohorte übergewichtige Kinder zu finden sind. Die Abbildung dient lediglich einem ersten Überblick.)*



Beispiel zweier Regressionsanalysen unter Verwendung des BMI- und LH-Werts:



Welchem sozialen Status lassen sich die Probanden zuordnen?

* Das Diagramm zeigt die Anzahl der Probanden nach den 3 Kategorien des Winkler-Index (niedriger Sozialstatus: 3–8 Punkte, mittlerer Sozialstatus: 9–14 und hoher Sozialstatus: 15–21 Punkte, NA= fehlende Angaben).

