



REPORT SERIES WITH DLOOKR

---

# Data Quality Diagnosis Report

---

*Author:*  
dlookr package

*Version:*  
0.4.0

February 3, 2024

# Contents

<b>1</b>	<b>Diagnose Data</b>	<b>3</b>
1.1	Overview of Diagnosis	3
1.1.1	List of all variables quality	3
1.1.2	Diagnosis of missing data	4
1.1.3	Diagnosis of unique data(Text and Category)	6
1.1.4	Diagnosis of unique data(Numerical)	6
1.2	Detailed data diagnosis	6
1.2.1	Diagnosis of categorical variables	6
1.2.2	Diagnosis of numerical variables	13
1.2.3	List of numerical diagnosis (zero)	15
1.2.4	List of numerical diagnosis (minus)	15
<b>2</b>	<b>Diagnose Outliers</b>	<b>17</b>
2.1	Overview of Diagnosis	17
2.1.1	Diagnosis of numerical variable outliers	17
2.2	Detailed outliers diagnosis	18

# Chapter 1

## Diagnose Data

### 1.1 Overview of Diagnosis

#### 1.1.1 List of all variables quality

Table 1.1: Data quality overview table

variables	type	missing (n)	missing (%)	unique
ano	numeric	0	0.000	
clase	character	0	0.000	
pertenencia_etnica	character	0	0.000	
pertenencia_etnica_agregada	character	0	0.000	
estrato	character	0	0.000	
colegio_caracter	character	486,881	96.477	
region	character	26,500	5.251	
region_dane	character	26,500	5.251	
departamento	character	26,500	5.251	
cod_dpto	numeric	26,500	5.251	
sexo	character	0	0.000	
grupo_decenal_edad	character	0	0.000	
regimen_afiliacion_p6100	character	31,875	6.316	
departamento_co	character	0	0.000	
quintil_ingreso_percapita	character	0	0.000	
quintil_ingreso_percapita_clase	character	0	0.000	
fex_c	numeric	0	0.000	87,
accion_tratar_enfermedad_p8563	character	467,811	92.698	
accion_tratar_enfermedad_agregada_p8563	character	467,811	92.698	
no_asistencia_razon_p6153	character	491,980	97.487	
no_asistencia_agregada_p6153	character	491,569	97.406	
no_afiliacion_razon_p768	character	480,195	95.152	
dias_incapacidad_p6134	numeric	459,922	91.135	
ocupacion_p6240	character	102,042	20.220	
ocupacion_posicion_p6435	character	287,670	57.003	
estado_salud_general_p6127	character	0	0.000	
calidad_atencion_prevision_p8595	character	301,123	59.668	
enfermedad_ultimo_mes_p5665	character	0	0.000	
cubrir_costos_atencion_salud_eps_p8556s1	character	459,922	91.135	
cubrir_costos_atencion_salud_seguro_voluntario_p8556s2	character	459,922	91.135	

Table 1.1: Data quality overview table (*continued*)

variables	type	missing (n)	missing (%)	unique
cubrir_costos_atencion_salud_seguro_obligatorio_p8556s4	character	459,922	91.135	
cubrir_costos_atencion_salud_alcaldia_p8556s5	character	459,922	91.135	
cubrir_costos_atencion_salud_propios_familiares_p8556s6	character	459,922	91.135	
cubrir_costos_atencion_salud_otras_personas_p8556s9	character	459,922	91.135	
cubrir_costos_atencion_salud_no_requiere_p8556s10	character	459,922	91.135	
percepcion_calidad_entidad_p6181	character	32,718	6.483	
percepcion_calidad_entidad_factor_p798	character	444,192	88.018	
percepcion_calidad_prestacion_p8554	logical	504,660	100.000	
percepcion_calidad_prestacion_factor_p801	character	502,075	99.488	
percepcion_pobreza_p5230	character	73	0.014	
percepcion_suficiencia_ingresos_p9090	character	73	0.014	
quejas_ultimo_ano	character	320,834	63.574	
tutelas_ultimo_ano	character	73	0.014	
tutelas_ultimo_ano_razon	character	496,664	98.416	
cronica_diagnostico_p1930	character	391,750	77.627	
cronica_tratamiento_p1930s1	character	488,762	96.850	
discapacidad_oir_p1906s1	character	365,250	72.375	
consulta_prevision_medico	character	290,901	57.643	
consulta_prevision_odontologo	character	290,901	57.643	
consulta_prevision_medico_odontologo	character	213,759	42.357	
discapacidad_hablar_p1906s2	character	365,250	72.375	
discapacidad_ver_p1906s3	character	365,250	72.375	
discapacidad_moverse_caminar_p1906s4	character	365,250	72.375	
discapacidad_usar_manos_p1906s5	character	365,250	72.375	
discapacidad_entender_p1906s6	character	365,250	72.375	
discapacidad_comer_vestirse_banarse_solo_p1906s7	character	365,250	72.375	
discapacidad_relacionarse_p1906s8	character	365,250	72.375	
dias_cita_consulta_acudio_especialista_p6199	logical	504,660	100.000	
dias_cita_consulta_medico_p6199	numeric	490,020	97.099	
dias_cita_consulta_odontologo_p6199	numeric	503,313	99.733	
algun_miembro_hogar_no_consumio_ninguna_comida_un_dia_o_mas_p8706	character	74,240	14.711	
controles_crecimiento_desarrollo_p6161	character	491,947	97.481	
controles_crecimiento_desarrollo_numero_p6161	numeric	493,243	97.738	
alfabeta_p6160	character	38,476	7.624	
asiste_educacion_p8586	character	38,476	7.624	
razon_no_estudio_p6218	character	390,923	77.463	
razon_no_estudio_agregada_p6218	character	390,923	77.463	
nivel_educativo_actual_p8587	character	176,005	34.876	
nivel_educativo_alcanzado_p1088	character	367,131	72.748	
anos_de_educacion	numeric	0	0.000	
anos_de_educacion_jefe	numeric	0	0.000	
analfabetismo_ipm	character	0	0.000	
rezago_escolar_ipm	character	399,906	79.243	
inasistencia_escolar_ipm	character	400,430	79.346	

### 1.1.2 Diagnosis of missing data

Table 1.2: Variables that include missing values

variables	type	missing (n)	missing (%)	unique
percepcion_calidad_prestacion_p8554	logical	504,660	100.000	
dias_cita_consulta_acudio_especialista_p6199	logical	504,660	100.000	
dias_cita_consulta_odontologo_p6199	numeric	503,313	99.733	
percepcion_calidad_prestacion_factor_p801	character	502,075	99.488	
tutelas_ultimo_ano_razon	character	496,664	98.416	
controles_crecimiento_desarrollo_numero_p6161	numeric	493,243	97.738	
no_asistencia_razon_p6153	character	491,980	97.487	
controles_crecimiento_desarrollo_p6161	character	491,947	97.481	
no_asistencia_agregada_p6153	character	491,569	97.406	
dias_cita_consulta_medico_p6199	numeric	490,020	97.099	
cronica_tratamiento_p1930s1	character	488,762	96.850	
colegio_caracter	character	486,881	96.477	
no_afiliacion_razon_p768	character	480,195	95.152	
accion_tratar_enfermedad_p8563	character	467,811	92.698	
accion_tratar_enfermedad_agregada_p8563	character	467,811	92.698	
dias_incapacidad_p6134	numeric	459,922	91.135	
cubrir_costos_atencion_salud_eps_p8556s1	character	459,922	91.135	
cubrir_costos_atencion_salud_seguro_voluntario_p8556s2	character	459,922	91.135	
cubrir_costos_atencion_salud_seguro_obligatorio_p8556s4	character	459,922	91.135	
cubrir_costos_atencion_salud_alcaldia_p8556s5	character	459,922	91.135	
cubrir_costos_atencion_salud_propios_familiares_p8556s6	character	459,922	91.135	
cubrir_costos_atencion_salud_otras_personas_p8556s9	character	459,922	91.135	
cubrir_costos_atencion_salud_no_requiere_p8556s10	character	459,922	91.135	
percepcion_calidad_entidad_factor_p798	character	444,192	88.018	
inasistencia_escolar_ipm	character	400,430	79.346	
rezago_escolar_ipm	character	399,906	79.243	
cronica_diagnostico_p1930	character	391,750	77.627	
razon_no_estudio_p6218	character	390,923	77.463	
razon_no_estudio_agregada_p6218	character	390,923	77.463	
nivel_educativo_alcanzado_p1088	character	367,131	72.748	
discapacidad_oir_p1906s1	character	365,250	72.375	
discapacidad_hablar_p1906s2	character	365,250	72.375	
discapacidad_ver_p1906s3	character	365,250	72.375	
discapacidad_moverse_caminar_p1906s4	character	365,250	72.375	
discapacidad_usar_manos_p1906s5	character	365,250	72.375	
discapacidad_entender_p1906s6	character	365,250	72.375	
discapacidad_comer_vestirse_banarse_solo_p1906s7	character	365,250	72.375	
discapacidad_relacionarse_p1906s8	character	365,250	72.375	
quejas_ultimo_ano	character	320,834	63.574	
calidad_atencion_preencion_p8595	character	301,123	59.668	
consulta_preencion_medico	character	290,901	57.643	
consulta_preencion_odontologo	character	290,901	57.643	
ocupacion_posicion_p6435	character	287,670	57.003	
consulta_preencion_medico_odontologo	character	213,759	42.357	
nivel_educativo_actual_p8587	character	176,005	34.876	
ocupacion_p6240	character	102,042	20.220	
algun_miembro_hogar_no_consumio_ninguna_comida_un_dia_o_mas_p8706	character	74,240	14.711	
alfabeta_p6160	character	38,476	7.624	

Table 1.2: Variables that include missing values (*continued*)

variables	type	missing (n)	missing (%)	unique
asiste_educacion_p8586	character	38,476	7.624	
percepcion_calidad_entidad_p6181	character	32,718	6.483	
regimen_afiliacion_p6100	character	31,875	6.316	
region	character	26,500	5.251	
region_dane	character	26,500	5.251	
departamento	character	26,500	5.251	
cod_dpto	numeric	26,500	5.251	
percepcion_pobreza_p5230	character	73	0.014	
percepcion_suficiencia_ingresos_p9090	character	73	0.014	
tutelas_ultimo_ano	character	73	0.014	

### 1.1.3 Diagnosis of unique data(Text and Category)

No variable with a high proportion greater than 0.5

### 1.1.4 Diagnosis of unique data(Numerical)

Table 1.3: Variables where the proportion of unique data is less than 0.1

variables	type	missing (n)	missing (%)	unique (n)	unique (n/N)
dias_incapacidad_p6134	numeric	459,922	91.135	71	0
cod_dpto	numeric	26,500	5.251	34	0
anos_de_educacion	numeric	0	0.000	32	0
anos_de_educacion_jefe	numeric	0	0.000	32	0
dias_cita_consulta_medico_p6199	numeric	490,020	97.099	29	0
dias_cita_consulta_odontologo_p6199	numeric	503,313	99.733	22	0
controles_crecimiento_desarrollo_numero_p6161	numeric	493,243	97.738	19	0
ano	numeric	0	0.000	7	0

## 1.2 Detailed data diagnosis

### 1.2.1 Diagnosis of categorical variables

Table 1.4: Top 10 levels of categories

variables	levels
clase	Cabecera
clase	Centros poblados, inspección de policía o corregimiento
pertenencia_etnica	Ninguno de los anteriores
pertenencia_etnica	Negro, mulato (afrodescendiente)
pertenencia_etnica	Indígena
pertenencia_etnica	Raizal del archipiélago
pertenencia_etnica	Palenquero
pertenencia_etnica	Gitano (rom)
pertenencia_etnica_agregada	Sin Pertenencia étnica
pertenencia_etnica_agregada	Con Pertenencia étnica

Table 1.4: Top 10 levels of categorical variables

variables	levels
estrato	1. Bajo - Bajo
estrato	2. Bajo
estrato	3. Medio - Bajo
estrato	0. Sin estrato
estrato	4. Medio
estrato	5. Medio - Alto
estrato	6. Alto
colegio_caracter	NA
colegio_caracter	Oficial
region	Atlántica
region	Central
region	Oriental
region	Pacífica (sin Valle)
region	Valle del Cauca
region	Antioquia
region	Bogotá
region	Orinoquía - Amazonia
region	NA
region	San Andrés, Providencia y Santa Catalina
region_dane	Pacífica
region_dane	Eje Cafetero - Antioquia
region_dane	Central
region_dane	Caribe
region_dane	NA
region_dane	Llanos - Orinoquía
region_dane	Amazonía
departamento	Valle del Cauca
departamento	Antioquía
departamento	Bogotá
departamento	Cauca
departamento	NA
departamento	Nariño
departamento	Cundinamarca
departamento	Atlántico
departamento	Tolima
departamento	Boyacá
sexo	Mujer
sexo	Hombre
grupo_decenal_edad	[10,20)
grupo_decenal_edad	[0,10)
grupo_decenal_edad	[20,30)
grupo_decenal_edad	[30,40)
grupo_decenal_edad	[40,50)
grupo_decenal_edad	[50,60)
grupo_decenal_edad	[60,70)
grupo_decenal_edad	[70,80)
grupo_decenal_edad	[80,90)
grupo_decenal_edad	[90 y más)

Table 1.4: Top 10 levels of categorical variables

variables	levels
regimen_afiliacion_p6100	2. Subsidiado (eps-s)
regimen_afiliacion_p6100	1. Contributivo (EPS)
regimen_afiliacion_p6100	NA
regimen_afiliacion_p6100	3. Especial (fuerzas armadas, ecopetrol, univers
regimen_afiliacion_p6100	4. No sabe, no informa
departamento_co	Valle del Cauca Colombia
departamento_co	Antioquía Colombia
departamento_co	Bogotá Colombia
departamento_co	Cauca Colombia
departamento_co	NA Colombia
departamento_co	Nariño Colombia
departamento_co	Cundinamarca Colombia
departamento_co	Atlántico Colombia
departamento_co	Tolima Colombia
departamento_co	Boyacá Colombia
quintil_ingreso_percapita	Quintil 4
quintil_ingreso_percapita	Quintil 3
quintil_ingreso_percapita	Quintil 5
quintil_ingreso_percapita	Quintil 2
quintil_ingreso_percapita	Quintil 1
quintil_ingreso_percapita_clase	Quintil 2
quintil_ingreso_percapita_clase	Quintil 4
quintil_ingreso_percapita_clase	Quintil 3
quintil_ingreso_percapita_clase	Quintil 5
quintil_ingreso_percapita_clase	Quintil 1
accion_tratar_enfermedad_p8563	NA
accion_tratar_enfermedad_p8563	Acudió a Institución prestadora de servicios de
accion_tratar_enfermedad_p8563	Autorecetado
accion_tratar_enfermedad_p8563	Usó remedios caseros
accion_tratar_enfermedad_p8563	Acudió de forma particular a un médico genera
accion_tratar_enfermedad_p8563	Acudió a un boticario, farmaceuta, droguista
accion_tratar_enfermedad_p8563	No hizo nada
accion_tratar_enfermedad_p8563	Consultó a un empírico, curandero, yerbatero, c
accion_tratar_enfermedad_p8563	Asistió a terapias alternativas (acupuntura, ese
accion_tratar_enfermedad_agregada_p8563	NA
accion_tratar_enfermedad_agregada_p8563	Médica Profesional
accion_tratar_enfermedad_agregada_p8563	Boticario, Medicina alternativa, remedios casero
accion_tratar_enfermedad_agregada_p8563	No hizo nada
no_asistencia_razon_p6153	NA
no_asistencia_razon_p6153	Caso leve
no_asistencia_razon_p6153	Mal servicio o cita distanciada en el tiempo
no_asistencia_razon_p6153	Falta de dinero
no_asistencia_razon_p6153	Centro de atención lejano
no_asistencia_razon_p6153	Muchos trámites para la cita
no_asistencia_razon_p6153	Espera demasiado prolongada
no_asistencia_razon_p6153	No confianza en los médicos
no_asistencia_razon_p6153	Otro
no_asistencia_razon_p6153	Problema no resuelto previamente



Table 1.4: Top 10 levels of categorical va

variables	levels
no_asistencia_agregada_p6153	NA
no_asistencia_agregada_p6153	Demanda
no_asistencia_agregada_p6153	Oferta
no_afiliacion_razon_p768	NA
no_afiliacion_razon_p768	Afiliación en trámite
no_afiliacion_razon_p768	No interesa
no_afiliacion_razon_p768	Trámites complejos
no_afiliacion_razon_p768	Otras razones
no_afiliacion_razon_p768	Falta de dinero
no_afiliacion_razon_p768	No vinculado laboralmente
no_afiliacion_razon_p768	Desconocimiento
no_afiliacion_razon_p768	Entidad cercana no disponible
ocupacion_p6240	Trabajando
ocupacion_p6240	Oficios del hogar
ocupacion_p6240	NA
ocupacion_p6240	Estudiando
ocupacion_p6240	Otra actividad
ocupacion_p6240	Buscando trabajo
ocupacion_p6240	Incapacitado permanentemente para trabajar
ocupacion_posicion_p6435	NA
ocupacion_posicion_p6435	Trabajador independiente o por cuenta propia
ocupacion_posicion_p6435	Obrero o empleado de empresa particular
ocupacion_posicion_p6435	Trabajador de finca, tierra o parcela propia en
ocupacion_posicion_p6435	Jornalero de peón
ocupacion_posicion_p6435	Obrero o empleado del gobierno
ocupacion_posicion_p6435	Patrón o empleado
ocupacion_posicion_p6435	Trabajador sin remuneración
ocupacion_posicion_p6435	Empleado doméstico
ocupacion_posicion_p6435	Profesional independiente
estado_salud_general_p6127	2. Bueno
estado_salud_general_p6127	3. Regular
estado_salud_general_p6127	1. Muy bueno
estado_salud_general_p6127	4. Malo
calidad_atencion_prevencion_p8595	NA
calidad_atencion_prevencion_p8595	Buena
calidad_atencion_prevencion_p8595	Mala
calidad_atencion_prevencion_p8595	Muy buena
calidad_atencion_prevencion_p8595	Muy Mala
enfermedad_ultimo_mes_p5665	No
enfermedad_ultimo_mes_p5665	Sí
cubrir_costos_atencion_salud_eps_p8556s1	NA
cubrir_costos_atencion_salud_eps_p8556s1	Entidad de seguridad social de la cual es afiliad
cubrir_costos_atencion_salud_eps_p8556s1	Otros
cubrir_costos_atencion_salud_seguro_voluntario_p8556s2	NA
cubrir_costos_atencion_salud_seguro_voluntario_p8556s2	Otros
cubrir_costos_atencion_salud_seguro_voluntario_p8556s2	Seguro médico, plan complementario o medicin
cubrir_costos_atencion_salud_seguro_obligatorio_p8556s4	NA
cubrir_costos_atencion_salud_seguro_obligatorio_p8556s4	Otros
cubrir_costos_atencion_salud_seguro_obligatorio_p8556s4	Seguro obligatorio de accidentes de tránsito (SC

Table 1.4: Top 10 levels of categorical va

variables	levels
cubrir_costos_atencion_salud_alcaldia_p8556s5	NA
cubrir_costos_atencion_salud_alcaldia_p8556s5	Otros
cubrir_costos_atencion_salud_alcaldia_p8556s5	Secretaría de salud o la alcaldía
cubrir_costos_atencion_salud_propios_familiares_p8556s6	NA
cubrir_costos_atencion_salud_propios_familiares_p8556s6	Otros
cubrir_costos_atencion_salud_propios_familiares_p8556s6	Recursos propios y/o familiares
cubrir_costos_atencion_salud_otras_personas_p8556s9	NA
cubrir_costos_atencion_salud_otras_personas_p8556s9	Otros
cubrir_costos_atencion_salud_otras_personas_p8556s9	Recursos de otras personas
cubrir_costos_atencion_salud_no_requiere_p8556s10	NA
cubrir_costos_atencion_salud_no_requiere_p8556s10	Otros
cubrir_costos_atencion_salud_no_requiere_p8556s10	No se requirió pago
percepcion_calidad_entidad_p6181	2. Bueno
percepcion_calidad_entidad_p6181	3. Mala
percepcion_calidad_entidad_p6181	NA
percepcion_calidad_entidad_p6181	1. Muy bueno
percepcion_calidad_entidad_p6181	4. Muy Mala
percepcion_calidad_entidad_p6181	5. No Sabe
percepcion_calidad_entidad_factor_p798	NA
percepcion_calidad_entidad_factor_p798	Demora en citas
percepcion_calidad_entidad_factor_p798	Mala atención
percepcion_calidad_entidad_factor_p798	Trámites excesivos
percepcion_calidad_entidad_factor_p798	Demora en atención médica
percepcion_calidad_entidad_factor_p798	Infraestructura
percepcion_calidad_entidad_factor_p798	Falta capacidad médica
percepcion_calidad_entidad_factor_p798	Otro
percepcion_calidad_prestacion_factor_p801	NA
percepcion_calidad_prestacion_factor_p801	Mala atención del personal administrativo o asi
percepcion_calidad_prestacion_factor_p801	Demora en la atención por parte del personal m
percepcion_calidad_prestacion_factor_p801	Demora en la asignación de citas
percepcion_calidad_prestacion_factor_p801	Trámites excesivos o dispendiosos
percepcion_calidad_prestacion_factor_p801	Falta de capacidad, conocimientos o habilidad o
percepcion_calidad_prestacion_factor_p801	Problemas relacionados con los medicamentos
percepcion_calidad_prestacion_factor_p801	Otro
percepcion_calidad_prestacion_factor_p801	Condiciones deficientes de infraestructura, dota
percepcion_pobreza_p5230	No
percepcion_pobreza_p5230	Sí
percepcion_pobreza_p5230	NA
percepcion_suficiencia_ingresos_p9090	Alcanza para cubrir gastos mínimos
percepcion_suficiencia_ingresos_p9090	No alcanza para cubrir gastos mínimos
percepcion_suficiencia_ingresos_p9090	Cubre más que gastos mínimos
percepcion_suficiencia_ingresos_p9090	NA
quejas_ultimo_ano	NA
quejas_ultimo_ano	No
quejas_ultimo_ano	Sí
tutelas_ultimo_ano	No
tutelas_ultimo_ano	Sí
tutelas_ultimo_ano	NA

Table 1.4: Top 10 levels of categorical va

variables	levels
tutelas_ultimo_ano_razon	NA
tutelas_ultimo_ano_razon	No lo atendieron en una institución de salud
tutelas_ultimo_ano_razon	No le dieron los medicamentos
tutelas_ultimo_ano_razon	Otra
tutelas_ultimo_ano_razon	No tiene dinero para pagar un tratamiento
tutelas_ultimo_ano_razon	No está afiliado al sistema general de seguridad
tutelas_ultimo_ano_razon	Le negaron el cambio de régimen de afiliación a
tutelas_ultimo_ano_razon	No está de acuerdo con el nivel de sisben asigna
cronica_diagnostico_p1930	NA
cronica_diagnostico_p1930	No
cronica_diagnostico_p1930	Sí
cronica_tratamiento_p1930s1	NA
cronica_tratamiento_p1930s1	Sí
cronica_tratamiento_p1930s1	No
discapacidad_oir_p1906s1	NA
discapacidad_oir_p1906s1	Sin dificultad
discapacidad_oir_p1906s1	Sí, con alguna dificultad
discapacidad_oir_p1906s1	Sí, con mucha dificultad
discapacidad_oir_p1906s1	No puede hacerlo
consulta_prevision_medico	NA
consulta_prevision_medico	Si
consulta_prevision_medico	No
consulta_prevision_odontologo	NA
consulta_prevision_odontologo	Si
consulta_prevision_odontologo	No
consulta_prevision_medico_odontologo	NA
consulta_prevision_medico_odontologo	Al médico y al odontólogo
consulta_prevision_medico_odontologo	A ninguno
consulta_prevision_medico_odontologo	Sólo al médico
consulta_prevision_medico_odontologo	Sólo al odontólogo
discapacidad_hablar_p1906s2	NA
discapacidad_hablar_p1906s2	Sin dificultad
discapacidad_hablar_p1906s2	Sí, con alguna dificultad
discapacidad_hablar_p1906s2	No puede hacerlo
discapacidad_hablar_p1906s2	Sí, con mucha dificultad
discapacidad_ver_p1906s3	NA
discapacidad_ver_p1906s3	Sin dificultad
discapacidad_ver_p1906s3	Sí, con alguna dificultad
discapacidad_ver_p1906s3	Sí, con mucha dificultad
discapacidad_ver_p1906s3	No puede hacerlo
discapacidad_moverse_caminar_p1906s4	NA
discapacidad_moverse_caminar_p1906s4	Sin dificultad
discapacidad_moverse_caminar_p1906s4	Sí, con alguna dificultad
discapacidad_moverse_caminar_p1906s4	Sí, con mucha dificultad
discapacidad_moverse_caminar_p1906s4	No puede hacerlo
discapacidad_usar_manos_p1906s5	NA
discapacidad_usar_manos_p1906s5	Sin dificultad
discapacidad_usar_manos_p1906s5	Sí, con alguna dificultad

Table 1.4: Top 10 levels of categorical va

variables	levels
discapacidad_usar_manos.p1906s5	Sí, con mucha dificultad
discapacidad_usar_manos.p1906s5	No puede hacerlo
discapacidad_entender.p1906s6	NA
discapacidad_entender.p1906s6	Sin dificultad
discapacidad_entender.p1906s6	Sí, con alguna dificultad
discapacidad_entender.p1906s6	No puede hacerlo
discapacidad_entender.p1906s6	Sí, con mucha dificultad
discapacidad_comer_vestirse_banarse_solo.p1906s7	NA
discapacidad_comer_vestirse_banarse_solo.p1906s7	Sin dificultad
discapacidad_comer_vestirse_banarse_solo.p1906s7	Sí, con alguna dificultad
discapacidad_comer_vestirse_banarse_solo.p1906s7	No puede hacerlo
discapacidad_comer_vestirse_banarse_solo.p1906s7	Sí, con mucha dificultad
discapacidad_relacionarse.p1906s8	NA
discapacidad_relacionarse.p1906s8	Sin dificultad
discapacidad_relacionarse.p1906s8	Sí, con alguna dificultad
discapacidad_relacionarse.p1906s8	No puede hacerlo
discapacidad_relacionarse.p1906s8	Sí, con mucha dificultad
algun_miembro_hogar_no_consumio_ninguna_comida_un_dia_o_mas.p8706	No
algun_miembro_hogar_no_consumio_ninguna_comida_un_dia_o_mas.p8706	NA
algun_miembro_hogar_no_consumio_ninguna_comida_un_dia_o_mas.p8706	Sí
controles_crecimiento_desarrollo.p6161	NA
controles_crecimiento_desarrollo.p6161	Sí
controles_crecimiento_desarrollo.p6161	No
alfabeta.p6160	Sí
alfabeta.p6160	No
alfabeta.p6160	NA
asiste_educacion.p8586	No
asiste_educacion.p8586	Sí
asiste_educacion.p8586	NA
razon_no_estudio.p6218	NA
razon_no_estudio.p6218	Necesita trabajar
razon_no_estudio.p6218	Costo elevado o falta de dinero
razon_no_estudio.p6218	Debe encargarse del hogar
razon_no_estudio.p6218	Considera que ya terminó
razon_no_estudio.p6218	No le gusta o no interesa el estudio
razon_no_estudio.p6218	Considera que no está en edad escolar
razon_no_estudio.p6218	Otros
razon_no_estudio.p6218	Por embarazo
razon_no_estudio.p6218	No existe centro educativo cercano
razon_no_estudio_agregada.p6218	NA
razon_no_estudio_agregada.p6218	Barreras_Acceso
razon_no_estudio_agregada.p6218	Razones_Personales
nivel_educativo_actual.p8587	NA
nivel_educativo_actual.p8587	Primaria
nivel_educativo_actual.p8587	Media
nivel_educativo_actual.p8587	Secundaria
nivel_educativo_actual.p8587	Ninguno
nivel_educativo_actual.p8587	Universitaria con título
nivel_educativo_actual.p8587	Técnico con título

Table 1.4: Top 10 levels of categorical variables

variables	levels
nivel_educativo_actual_p8587	Tecnológico con título
nivel_educativo_actual_p8587	Posgrado con título
nivel_educativo_actual_p8587	Universitaria sin título
nivel_educativo_alcanzado_p1088	NA
nivel_educativo_alcanzado_p1088	Primaria
nivel_educativo_alcanzado_p1088	Secundaria
nivel_educativo_alcanzado_p1088	Media
nivel_educativo_alcanzado_p1088	Universitario
nivel_educativo_alcanzado_p1088	Preescolar
nivel_educativo_alcanzado_p1088	Técnico
nivel_educativo_alcanzado_p1088	Tecnológico
nivel_educativo_alcanzado_p1088	Postgrado
analfabetismo_ipm	No
analfabetismo_ipm	Si
rezago_escolar_ipm	NA
rezago_escolar_ipm	Si
rezago_escolar_ipm	No
inasistencia_escolar_ipm	NA
inasistencia_escolar_ipm	No
inasistencia_escolar_ipm	Si

### 1.2.2 Diagnosis of numerical variables

Table 1.5: General list of numerical diagnosis

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
ano	2,012.000	2,013.00	2,015.059	2,015.000	2,017.00	2,018.000	0	0	0
cod_dpto	5.000	15.00	41.054	41.000	70.00	99.000	0	0	0
fex_c	2.429	151.07	584.590	343.032	747.89	9,401.786	0	0	39,776
dias_incapacidad_p6134	0.000	0.00	6.649	2.000	7.00	120.000	12,478	0	3,300
dias_cita_consulta_medico_p6199	0.000	0.00	3.038	1.000	3.00	30.000	5,046	0	1,954
dias_cita_consulta_odontologo_p6199	0.000	0.00	3.513	1.000	3.00	30.000	458	0	235
controles_crecimiento_desarrollo_numero_p6161	1.000	2.00	2.571	2.000	3.00	24.000	0	0	853
anos_de_educacion	0.000	2.00	6.530	5.000	11.00	31.000	87,375	0	1,396
anos_de_educacion_jefe	0.000	0.00	2.268	0.000	2.00	31.000	364,607	0	79,291

### 1.2.3 List of numerical diagnosis (zero)

Table 1.6: List of numerical diagnosis (zero)

variables	min	median	max	zero	zero ratio(%)
anos_de_educacion_jefe	0	0	31	364,607	72.248
anos_de_educacion	0	5	31	87,375	17.314
dias_incapacidad_p6134	0	2	120	12,478	2.473
dias_cita_consulta_medico_p6199	0	1	30	5,046	1.000
dias_cita_consulta_odontologo_p6199	0	1	30	458	0.091

### 1.2.4 List of numerical diagnosis (minus)

No numeric variable with negative value





## Chapter 2

# Diagnose Outliers

### 2.1 Overview of Diagnosis

#### 2.1.1 Diagnosis of numerical variable outliers

Table 2.1: Diagnosis of numerical variable outliers

variables	min	median	max	outlier	outlier ratio(%)
anos_de_educacion_jefe	0.000	0.000	31.000	79,291	15.712
fex_c	2.429	343.032	9,401.786	39,776	7.882
dias_incapacidad_p6134	0.000	2.000	120.000	3,300	0.654
dias_cita_consulta_medico_p6199	0.000	1.000	30.000	1,954	0.387
anos_de_educacion	0.000	5.000	31.000	1,396	0.277
controles_crecimiento_desarrollo_numero_p6161	1.000	2.000	24.000	853	0.169
dias_cita_consulta_odontologo_p6199	0.000	1.000	30.000	235	0.047

## 2.2 Detailed outliers diagnosis

variable : anos\_de\_educacion\_jefe

Table 2.2: Outliers information of anos\_de\_educacion\_jefe

Measures	Values
Outliers count	79,291.00
Outliers ratio (%)	15.71
Mean of outliers	11.68
Mean with outliers	2.27
Mean without outliers	0.51

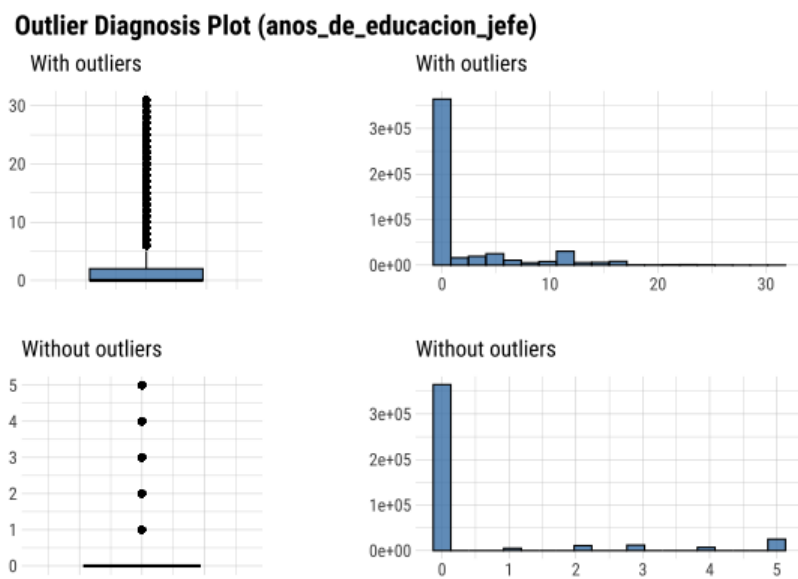


Figure 2.1: Distribution of anos\_de\_educacion\_jefe

variable : fex\_c

Table 2.3: Outliers information of fex\_c

Measures	Values
Outliers count	39,776.00
Outliers ratio (%)	7.88
Mean of outliers	2,382.65
Mean with outliers	584.59
Mean without outliers	430.75

**Outlier Diagnosis Plot (fex\_c)**

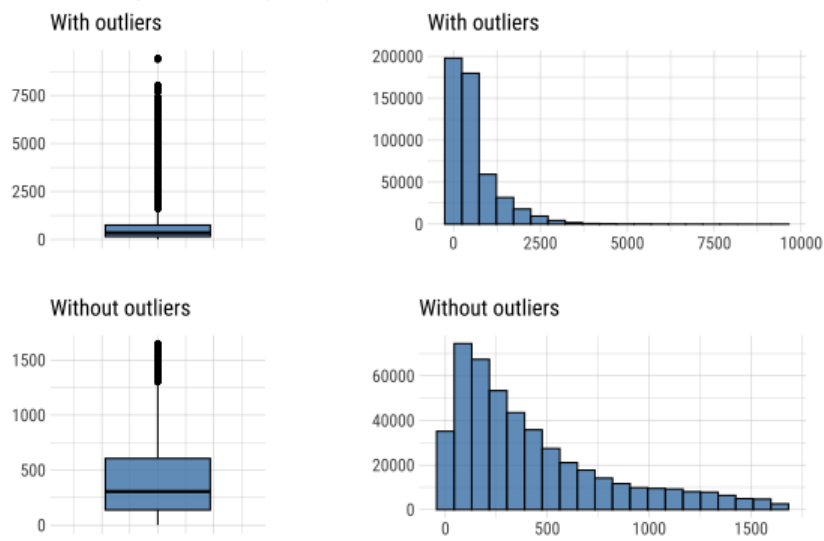


Figure 2.2: Distribution of fex\_c

variable : dias\_incapacidad\_p6134

Table 2.4: Outliers information of dias\_incapacidad\_p6134

Measures	Values
Outliers count	3,300.00
Outliers ratio (%)	0.65
Mean of outliers	48.55
Mean with outliers	6.65
Mean without outliers	3.31

### Outlier Diagnosis Plot (dias\_incapacidad\_p6134)

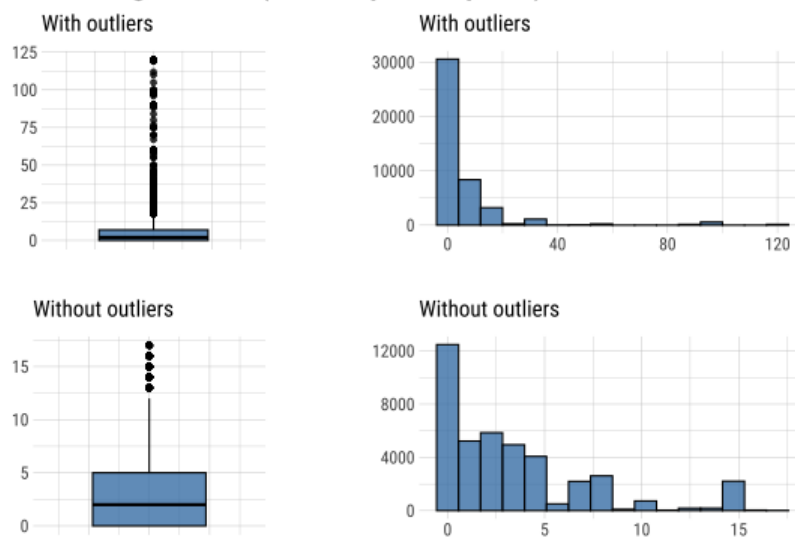


Figure 2.3: Distribution of dias\_incapacidad\_p6134

variable : dias\_cita\_consulta\_medico\_p6199

Table 2.5: Outliers information of dias\_cita\_consulta\_medico\_p6199

Measures	Values
Outliers count	1,954.00
Outliers ratio (%)	0.39
Mean of outliers	14.92
Mean with outliers	3.04
Mean without outliers	1.21

**Outlier Diagnosis Plot (dias\_cita\_consulta\_medico\_p6199)**

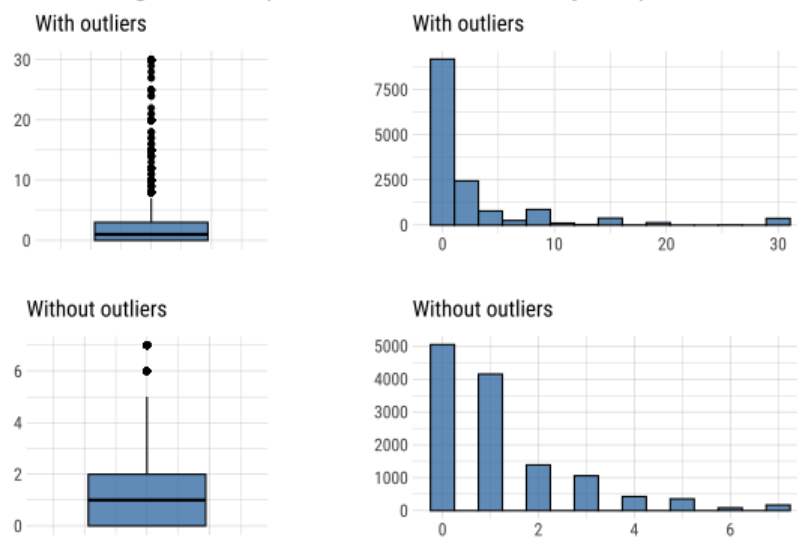


Figure 2.4: Distribution of dias\_cita\_consulta\_medico\_p6199

variable : anos\_de\_educacion

Table 2.6: Outliers information of anos\_de\_educacion

Measures	Values
Outliers count	1,396.00
Outliers ratio (%)	0.28
Mean of outliers	26.36
Mean with outliers	6.53
Mean without outliers	6.47

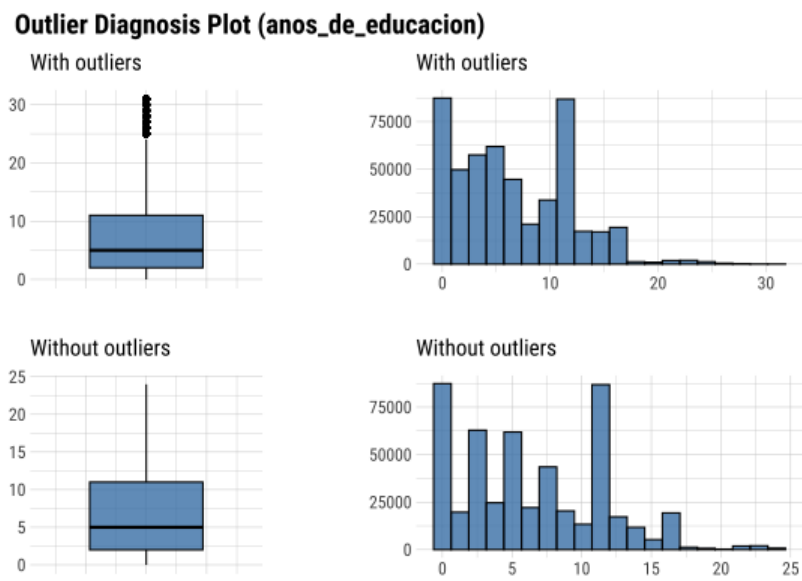


Figure 2.5: Distribution of anos\_de\_educacion

variable : controles\_crecimiento\_desarrollo\_numero\_p6161

Table 2.7: Outliers information of controles\_crecimiento\_desarrollo\_numero\_p6161

Measures	Values
Outliers count	853.00
Outliers ratio (%)	0.17
Mean of outliers	6.72
Mean with outliers	2.57
Mean without outliers	2.24

#### Outlier Diagnosis Plot (controles\_crecimiento\_desarrollo\_numero\_p6161)

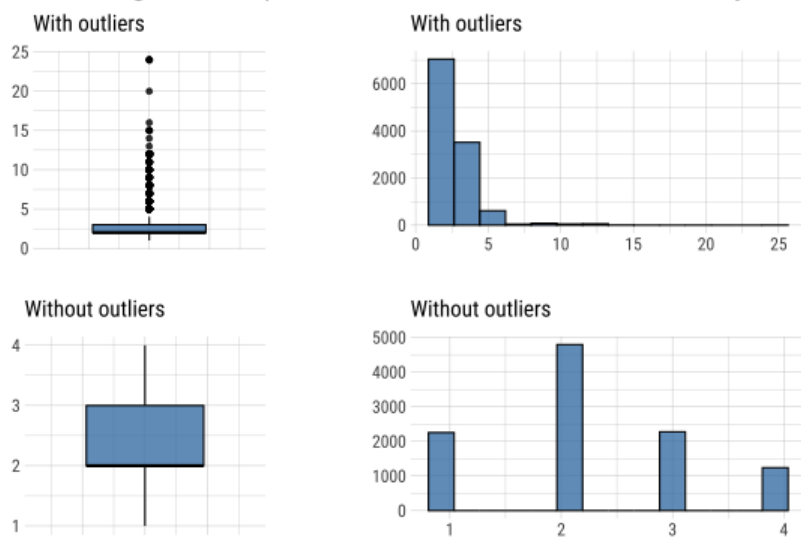


Figure 2.6: Distribution of controles\_crecimiento\_desarrollo\_numero\_p6161

variable : dias\_cita\_consulta\_odontologo\_p6199

Table 2.8: Outliers information of dias\_cita\_consulta\_odontologo\_p6199

Measures	Values
Outliers count	235.00
Outliers ratio (%)	0.05
Mean of outliers	14.64
Mean with outliers	3.51
Mean without outliers	1.16

**Outlier Diagnosis Plot (dias\_cita\_consulta\_odontologo\_p6199)**

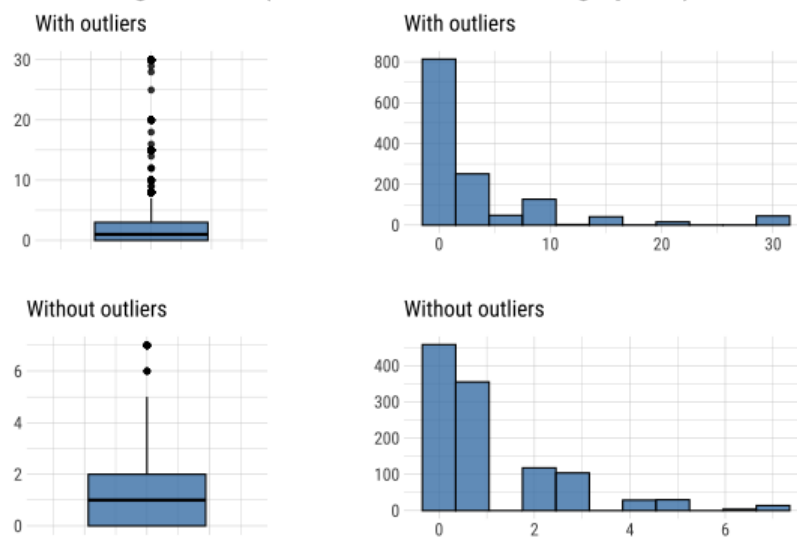


Figure 2.7: Distribution of dias\_cita\_consulta\_odontologo\_p6199