

# Legal ID - Multicountry Comparison

Shabrina Mardevi

18/10/2020

## 1. Prepare IPUMS Data

```
# Load latest IPUMS Census Data for Ecuador, Pakistan, Peru, and Zambia
multi_ipums <- read_dta("ipumsi_00009 2.dta")

# Subset population to samples with legal ID age and above
ipums_legal_age <-
  multi_ipums %>%
  filter((country == 218 & age > 16) |
         (country == 586 & age > 17) |
         (country == 604 & age > 17) |
         (country == 894 & age > 15)) %>%

# Set variables as.factor()
mutate_at(.vars = c("country", "sex", "ethnec", "geolev1", "age2",
                   "edattain", "ec2010a_citcard", "ec2010a_race",
                   "pk1998a_relig", "pk1998a_idcard", "pe2007a_idcard",
                   "pe2007a_religion", "zm2010a_relig", "ethniczm",
                   "zm2010a_nidcard"), ~as.factor(.))%>%

# Recode countries
mutate(country = recode(country, "218" = "ecuador",
                        "586" = "pakistan",
                        "604" = "peru",
                        "894" = "zambia"),
       edattain = recode(edattain, "0" = "NIU",
                        "1" = "Less than primary",
                        "2" = "Primary",
                        "3" = "Secondary",
                        "4" = "Post-secondary",
                        "9" = "Unknown"),
       ethniczm = recode(ethniczm,
                        "4" = "Bisa",
                        "18" = "Luano",
                        "34" = "Luyana",
                        "42" = "Mashi",
                        "61" = "Yombe",
                        "68" = "Other",
                        "99" = "Unspecified",
```

```

        "67" = "European",
        "66" = "Asian",
        "65" = "American",
        "64" = "Other African"),
  ethnicec = recode(ethnicec,
    "1" = "Achuar",
    "2" = "Awa",
    "3" = "Cofan",
    "4" = "Chachi",
    "7" = "Siona",
    "9" = "Zapara",
    "18" = "Quitu",
    "31" = "Epera",
    "33" = "Andoa",
    "36" = "Pastos",
    "37" = "Tomabela",
    "34" = "Pastos"),
  zm2010a_relig = recode(zm2010a_relig, "4" = "Hindu", "3" = "Muslim"),
  pk1998a_relig = recode(pk1998a_relig, "3" = "Hindu", "5" = "Scheduled Caste"),
  sex = recode(sex, "1" = "Male", "2" = "Female"),
  age2 = recode(age2, "4" = "15-19",
    "12" = "20-24",
    "13" = "25-29",
    "14" = "30-34",
    "15" = "35-39",
    "16" = "40-44",
    "17" = "45-49",
    "18" = "50-54",
    "19" = "55-59",
    "20" = "60-64",
    "21" = "65-69",
    "22" = "70-74",
    "23" = "75-79",
    "24" = "80+",
    "25" = "85+" ),
  geolev1 = recode(geolev1, "586005" = "Baluchistan",
    "218006" = "Chimborazo",
    "604007" = "Callao",
    "604018" = "Moquegua")
)

```

```

# Create function to summarise Legal ID ownership by subpopulations
subpop_summary <- function(data) {

  # Set CI alpha (95% level) - (1-alpha/2)*100%
  alpha = 0.05

  # Calculate the critical z-score
  z = qnorm(1-alpha/2)

  data_imp <-
  data %>%
  gather(-c(country, id, ), key = "type", value = "subpop") %>%

```

```

group_by(country, type, subpop, id) %>%

# Number of ID possession by Subgroups, weighted (Person Weight = 10)
summarise(n_id=(n()*10)) %>%
mutate (prop_id = n_id/sum(n_id)) %>%
mutate (total_subpop = sum(n_id)) %>%

# Standard Error
mutate (Std_Err = z*sqrt(prop_id*(1-prop_id)/total_subpop)) %>%

# Confidence Intervals
mutate (Lower_CI = prop_id + (-1)*Std_Err) %>%
mutate (Upper_CI = prop_id + 1*Std_Err) %>%
filter(id == "yes") %>%
select(-id)

# Range
Range <- max(data_imp$prop_id)-min(data_imp$prop_id)
data_imp <- cbind(data_imp, Range)

data_imp <- rename(data_imp, "Range" = "...10")

return(data_imp)
}

# Ecuador Legal ID by Education, Ethnicity, Race, and Urban/Rural
coord_ecuador <- runif(44, 0.5, 1.3)

ecuador <-
  ipums_legal_age %>%
  subset(country == "ecuador",
    select = c(country, geolev1, sex, age2, edattain,
      ec2010a_citcard, ec2010a_race)) %>%
  mutate(ec2010a_citcard = recode(ec2010a_citcard, "1" = "yes", "2" = "no")) %>%
  rename_at (.vars = c("ec2010a_race", "ec2010a_citcard"),
    ~c("race", "id")) %>%
  subpop_summary() %>%
  cbind(., coord_ecuador) %>%
  rename(coord = "...11") %>%
  mutate(subpop = recode(subpop, "3" = "Black"))

## Warning: attributes are not identical across measure variables;
## they will be dropped

## `summarise()` regrouping output by 'country', 'type', 'subpop' (override with `.groups` argument)

## New names:
## * NA -> ...10

## New names:
## * NA -> ...11

```

ecuador%>%

kable(format = "pipe", caption = "Legal ID Possession by Subpopulations")

Table 1: Legal ID Possession by Subpopulations

country type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_C	Upper_C	Range	coord
ecuador age2	15-19	779270	0.9255868	841920	0.0005600	0.9250262	0.9261473	0.0744132	0.5967049
ecuador age2	20-24	12650300	0.9817393	1288560	0.0002312	0.9815081	0.9819705	0.0744132	0.6456062
ecuador age2	25-29	11877100	0.9872655	1203030	0.0002004	0.9870651	0.9874659	0.0744132	0.6122436
ecuador age2	30-34	10556700	0.9895020	1066870	0.0001934	0.9893086	0.9896954	0.0744132	0.8606575
ecuador age2	35-39	932950	0.9905295	941870	0.0001956	0.9903339	0.9907251	0.0744132	0.2081918
ecuador age2	40-44	811760	0.9905915	819470	0.0002090	0.9903823	0.9908005	0.0744132	0.5510471
ecuador age2	45-49	746170	0.9908507	753060	0.0002150	0.9906356	0.9910657	0.0744132	0.2491239
ecuador age2	50-54	602800	0.9908444	608370	0.0002393	0.9906051	0.9910837	0.0744132	0.1595913
ecuador age2	55-59	508760	0.9899788	513910	0.0002723	0.9897063	0.9902511	0.0744132	0.9923463
ecuador age2	60-64	392160	0.9876842	397050	0.0003431	0.9873411	0.9880272	0.0744132	0.1337957
ecuador age2	65-69	318760	0.9869647	322970	0.0003912	0.9865736	0.9873559	0.0744132	0.8658178
ecuador age2	70-74	234020	0.9821219	238280	0.0005320	0.9815898	0.9826539	0.0744132	0.9480655
ecuador age2	75-79	161830	0.9834701	164550	0.0006160	0.9828540	0.9840861	0.0744132	0.5004945
ecuador age2	80+	113430	0.9799568	115750	0.0008074	0.9791494	0.9807642	0.0744132	0.8150164
ecuador age2	85+	91410	0.9665856	94570	0.0011454	0.9654402	0.9677310	0.0744132	0.0482614
ecuador edattain	Less than primary	17954700	0.9647981	1860980	0.0002648	0.9645333	0.9650629	0.0744132	0.6394571
ecuador edattain	Post-secondary	761520	1.0000000	761520	0.0000000	1.0000000	1.0000000	0.0744132	0.0149558
ecuador edattain	Primary	36166600	0.9790925	3693890	0.0001459	0.9789466	0.9792384	0.0744132	0.5119528
ecuador edattain	Secondary	27312800	0.9960686	2742060	0.0000741	0.9959946	0.9961427	0.0744132	0.7888524
ecuador edattain	Unknown	296800	0.9519533	311780	0.0007507	0.9512026	0.9527040	0.0744132	0.5833634
ecuador geolev1	218001	456530	0.9830111	464420	0.0003717	0.9826394	0.9833827	0.0744132	0.7928891
ecuador geolev1	218002	110860	0.9768262	113490	0.0008753	0.9759508	0.9777013	0.0744132	0.6598846
ecuador geolev1	218004	101940	0.9477501	107560	0.0013299	0.9464202	0.9490800	0.0744132	0.2393140
ecuador geolev1	218005	249100	0.9826818	253490	0.0005078	0.9821739	0.9831896	0.0744132	0.9023995
ecuador geolev1	218007	387930	0.9825988	394800	0.0004079	0.9821909	0.9830067	0.0744132	0.7739131
ecuador geolev1	218009	63248500	0.9825758	6437010	0.0001011	0.9824747	0.9826768	0.0744132	0.1365450
ecuador geolev1	218010	252570	0.9822276	257140	0.0005107	0.9817169	0.9827383	0.0744132	0.0349533
ecuador geolev1	218011	282980	0.9813428	288360	0.0004939	0.9808489	0.9818366	0.0744132	0.6078435
ecuador geolev1	218014	75300	0.9709865	77550	0.0011813	0.9698051	0.9721678	0.0744132	0.9527406
ecuador geolev1	218016	47200	0.9837432	47980	0.0011316	0.9826117	0.9848748	0.0744132	0.1104156
ecuador geolev1	218018	339130	0.9895539	342710	0.0003404	0.9892133	0.9898942	0.0744132	0.9545477
ecuador geolev1	218019	50910	0.9790385	52000	0.0012313	0.9778072	0.9802697	0.0744132	0.5346340
ecuador geolev1	218021	229820	0.9624356	238790	0.0007626	0.9616730	0.9631982	0.0744132	0.1662686
ecuador geolev1	Chimborazo	292610	0.9921337	294930	0.0003188	0.9918149	0.9924526	0.0744132	0.5044102
ecuador race	1	577870	0.9752093	592560	0.0003959	0.9748134	0.9756052	0.0744132	0.5268060
ecuador race	2	376740	0.9706542	388130	0.0005310	0.9701232	0.9711851	0.0744132	0.6004743
ecuador race	Black	87550	0.9366642	93470	0.0015615	0.9351027	0.9382256	0.0744132	0.6464922
ecuador race	4	169650	0.9674936	175350	0.0008300	0.9666633	0.9683236	0.0744132	0.8836994
ecuador race	5	713180	0.9781246	729130	0.0003358	0.9777889	0.9784604	0.0744132	0.9770122
ecuador race	6	66687900	0.9850793	6769800	0.0000913	0.9849880	0.9851706	0.0744132	0.8347690
ecuador race	7	572530	0.9788009	584930	0.0003691	0.9784317	0.9791700	0.0744132	0.6509509
ecuador race	8	35420	0.9609333	36860	0.0019780	0.9589553	0.9629112	0.0744132	0.8177613
ecuador sex	Female	47084200	0.9834799	4787510	0.0001142	0.9833658	0.9835941	0.0744132	0.2903603
ecuador sex	Male	44933100	0.9804898	4582720	0.0001266	0.9803631	0.9806164	0.0744132	0.0944005

```

# Pakistan Legal ID by Education, Religion, and Urban/Rural

pakistan <-
  ipums_legal_age %>%
  subset(country == "pakistan",
    select = c(country, geolevl, sex, age2, edattain, pk1998a_relig, pk1998a_idcard)) %>%
  mutate(pk1998a_idcard = recode(pk1998a_idcard, "1" = "yes", "2" = "no",
    "8" = "unknown", "9" = "niu")) %>% # There are NIU IDs in Pakistan des
  rename_at (.vars = c("pk1998a_idcard", "pk1998a_relig"),
    ~c("id", "religion") ) %>%
  subpop_summary() %>%
  cbind(., runif(32, 4.7, 5.3)) %>%
  rename(coord = "...11")

## Warning: attributes are not identical across measure variables;
## they will be dropped

## `summarise()` regrouping output by 'country', 'type', 'subpop' (override with `.groups` argument)

## New names:
## * NA -> ...10

## New names:
## * NA -> ...11

pakistan %>%
  kable(format = "pipe", caption = "Legal ID Possession by Subpopulations")

```

Table 2: Legal ID Possession by Subpopulations

country	type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_CI	Upper_CI	Range	coord
pakistan	age2	15-19	807690	0.122751	116579900	0.0002507	0.1225004	0.1230018	0.8078984	5.250194
pakistan	age2	20-24	4403200	0.3744689	11758520	0.0002766	0.3741923	0.3747455	0.8078984	4.878698
pakistan	age2	25-29	5651930	0.5838385	9680640	0.0003105	0.5835280	0.5841490	0.8078984	4.941686
pakistan	age2	30-34	5721280	0.7008555	8163280	0.0003141	0.7005414	0.7011690	0.8078984	4.864712
pakistan	age2	35-39	4788120	0.7655382	6254580	0.0003320	0.7652062	0.7658702	0.8078984	5.031570
pakistan	age2	40-44	4877460	0.8381956	5819000	0.0002992	0.8378963	0.8384940	0.8078984	4.835173
pakistan	age2	45-49	4046660	0.8776957	4610550	0.0002991	0.8773967	0.8779940	0.8078984	4.829169
pakistan	age2	50-54	3690580	0.8801156	4193290	0.0003109	0.8798047	0.8804265	0.8078984	4.721070
pakistan	age2	55-59	2537350	0.9052782	2802840	0.0003428	0.9049354	0.9056210	0.8078984	5.068662
pakistan	age2	60-64	2337470	0.8740526	2674290	0.0003977	0.8736549	0.8744502	0.8078984	5.202524
pakistan	age2	65-69	1411310	0.8975801	1572350	0.0004739	0.8971061	0.8980540	0.8078984	5.102901
pakistan	age2	70-74	1226290	0.8569462	1431000	0.0005737	0.8563725	0.8575199	0.8078984	5.217880
pakistan	age2	75-79	509380	0.8788323	579610	0.0008401	0.8779922	0.8796724	0.8078984	5.091288
pakistan	age2	80+	477540	0.8048608	593320	0.0010084	0.8038524	0.8058692	0.8078984	5.107677
pakistan	age2	85+	322090	0.7904437	407480	0.0012496	0.7891941	0.7916933	0.8078984	4.942650
pakistan	edattain	Less than primary	262905900	0.5962463	44093510	0.0001448	0.5961014	0.5963911	0.8078984	5.035534
pakistan	edattain	Post-secondary	2198920	0.9306495	2362780	0.0003239	0.9303256	0.9309734	0.8078984	4.727057
pakistan	edattain	Primary	121708700	0.6809160	17874260	0.0002161	0.6806999	0.6811321	0.8078984	5.206252

country	type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_C	Upper_C	Range	coord
pakistan	edattain	Secondary	2058570	0.7748044	2656890	0.0005023	0.7743021	0.7753060	0.8078984	5.148806
pakistan	edattain	Unknown	89400	0.6711208	133210	0.0025229	0.6685979	0.6736437	0.8078984	5.012330
pakistan	geolev1	586001	5462060	0.6297833	8672920	0.0003214	0.6294619	0.6301040	0.8078984	4.787196
pakistan	geolev1	586003	2536001	0.6483387	39115370	0.0001496	0.6481891	0.6484884	0.8078984	5.282214
pakistan	geolev1	586004	102852700	0.6402803	16063700	0.0002347	0.6400456	0.6405149	0.8078984	4.983045
pakistan	geolev1	Baluchistan	1701010	0.5203998	3268660	0.0005416	0.5198582	0.5209414	0.8078984	5.099534
pakistan	religion	1	413441200	0.6395477	64645870	0.0001170	0.6394307	0.6396640	0.8078984	5.000746
pakistan	religion	2	693290	0.6367293	1088830	0.0009034	0.6358260	0.6376327	0.8078984	4.845817
pakistan	religion	4	111040	0.7109290	156190	0.0022482	0.7086808	0.7131772	0.8078984	4.898153
pakistan	religion	6	29950	0.5605465	53430	0.0042084	0.5563381	0.5647549	0.8078984	5.036797
pakistan	religion	Hindu	545370	0.5343517	1020620	0.0009677	0.5333839	0.5353194	0.8078984	5.007551
pakistan	religion	Scheduled Caste	84580	0.5431893	155710	0.0024742	0.5407151	0.5456633	0.8078984	5.258038
pakistan	sex	Female	171036500	0.5295859	32296270	0.0001721	0.5294138	0.5297581	0.8078984	5.176194
pakistan	sex	Male	257047000	0.7381237	34824380	0.0001460	0.7379777	0.7382697	0.8078984	4.973777

```
# Peru Legal ID by Education, Religion, and Urban/Rural
peru <-
  ipums_legal_age %>%
  subset(country == "peru",
    select = c(country, geolev1, sex, age2, edattain, pe2007a_idcard, pe2007a_religion)) %>%
  mutate(pe2007a_idcard = recode(pe2007a_idcard, "1" = "yes", "2" = "no", "9" = "niu")) %>%
  rename_at (.vars = c("pe2007a_idcard", "pe2007a_religion"),
    ~c("id", "religion") ) %>%
  subpop_summary() %>%
  cbind(., runif(50, 1.7, 2.3)) %>%
  rename(coord = "...11")

## Warning: attributes are not identical across measure variables;
## they will be dropped

## `summarise()` regrouping output by 'country', 'type', 'subpop' (override with `.groups` argument)

## New names:
## * NA -> ...10

## New names:
## * NA -> ...11

peru %>%
  kable(format = "pipe", caption = "Legal ID Possession by Subpopulations")
```

Table 3: Legal ID Possession by Subpopulations

country	type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_C	Upper_C	Range	coord
peru	age2	15-19	917210	0.8507810	1078080	0.0006726	0.8501084	0.8514536	0.1683057	2.135483
peru	age2	20-24	2452740	0.9646772	2542550	0.0002269	0.9644503	0.9649041	0.1683057	1.801400
peru	age2	25-29	2256850	0.9791913	2304810	0.0001843	0.9790071	0.9793756	0.1683057	2.137607
peru	age2	30-34	2044530	0.9859191	2073730	0.0001604	0.9857587	0.9860793	0.1683057	2.191682

country	type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_CI	Upper_CI	Range	coord
peru	age2	35-39	1854670	0.986154	21880710	0.000167	0.985987	0.986321	0.168305	7.222323
peru	age2	40-44	1631610	0.986982	31653130	0.000172	0.986809	0.987155	0.168305	7.2294979
peru	age2	45-49	1356260	0.986772	81374440	0.000191	0.986581	0.986963	0.168305	7.2092374
peru	age2	50-54	1134510	0.984450	21152430	0.000225	0.984224	0.984676	0.168305	7.2113520
peru	age2	55-59	868420	0.981088	2885160	0.000283	0.980804	0.981371	0.168305	7.2076329
peru	age2	60-64	710200	0.971134	7731310	0.000383	0.970750	0.971517	0.168305	7.1953366
peru	age2	65-69	557820	0.962904	579310	0.000486	0.962417	0.963390	0.168305	7.2032287
peru	age2	70-74	427140	0.944728	3452130	0.000666	0.944062	0.945394	0.168305	7.1853211
peru	age2	75-79	317610	0.926652	342750	0.000872	0.925779	0.927524	0.168305	7.1906042
peru	age2	80+	180000	0.896816	3200710	0.001330	0.895485	0.898147	0.168305	7.1948101
peru	age2	85+	153900	0.827552	185970	0.001716	0.825835	0.829269	0.168305	7.1738549
peru	edattain	Less than primary	4286420	0.934035	64589140	0.000227	0.933808	0.934262	0.168305	7.1710991
peru	edattain	Post-secondary	1861160	0.995858	51868900	0.000092	0.995766	0.995950	0.168305	7.2097081
peru	edattain	Primary	2963280	0.951614	53113950	0.000238	0.951376	0.951852	0.168305	7.1740565
peru	edattain	Secondary	7752610	0.985681	37865230	0.000083	0.985598	0.985764	0.168305	7.2291096
peru	geolev1	604001	189700	0.909090	208670	0.001233	0.907857	0.910324	0.168305	7.129073
peru	geolev1	604002	633820	0.965850	656230	0.000439	0.965411	0.966289	0.168305	7.2240801
peru	geolev1	604003	220210	0.972487	226440	0.000673	0.971813	0.973160	0.168305	7.1983519
peru	geolev1	604004	764590	0.979904	780270	0.000311	0.979593	0.980215	0.168305	7.2276642
peru	geolev1	604005	341680	0.973835	350860	0.000528	0.973307	0.974363	0.168305	7.2269842
peru	geolev1	604006	758280	0.941611	805300	0.000512	0.941099	0.942123	0.168305	7.2004619
peru	geolev1	604008	673390	0.966431	696780	0.000422	0.966008	0.966854	0.168305	7.1801762
peru	geolev1	604009	237700	0.968464	245440	0.000691	0.967773	0.969156	0.168305	7.2248433
peru	geolev1	604010	398880	0.942310	423300	0.000702	0.941608	0.943012	0.168305	7.2164512
peru	geolev1	604011	451780	0.981405	460340	0.000390	0.981014	0.981795	0.168305	7.1738849
peru	geolev1	604012	723010	0.959357	753640	0.000445	0.958911	0.959803	0.168305	7.2261640
peru	geolev1	604013	982240	0.960719	1022400	0.000376	0.960343	0.961096	0.168305	7.2131258
peru	geolev1	604014	678190	0.966110	701980	0.000423	0.965686	0.966534	0.168305	7.1772458
peru	geolev1	604015	5767360	0.980241	75883610	0.000112	0.980129	0.980354	0.168305	7.1959269
peru	geolev1	604016	444560	0.915259	485720	0.000783	0.914476	0.916043	0.168305	7.2286839
peru	geolev1	604017	65050	0.941525	69090	0.001749	0.939775	0.943275	0.168305	7.2018880
peru	geolev1	604019	164470	0.967129	170060	0.000847	0.966281	0.967976	0.168305	7.1839170
peru	geolev1	604020	967510	0.951421	151016910	0.000417	0.951003	0.951839	0.168305	7.1943694
peru	geolev1	604021	762110	0.977628	779550	0.000328	0.977299	0.977956	0.168305	7.1982882
peru	geolev1	604022	403840	0.936484	431230	0.000727	0.935756	0.937211	0.168305	7.2273465
peru	geolev1	604023	189980	0.983842	193100	0.000562	0.983280	0.984404	0.168305	7.2047081
peru	geolev1	604024	123300	0.960056	128430	0.001071	0.958985	0.961127	0.168305	7.2262361
peru	geolev1	604025	226230	0.933330	242390	0.000993	0.932337	0.934323	0.168305	7.1993609
peru	geolev1	Callao	584860	0.985525	593450	0.000303	0.985221	0.985829	0.168305	7.2142773
peru	geolev1	Moquegua	110730	0.988396	112030	0.000627	0.987768	0.989023	0.168305	7.1799231
peru	religion	1	137850300	0.969240	314222510	0.000089	0.969150	0.969330	0.168305	7.1783756
peru	religion	2	2065080	0.960354	2150330	0.000260	0.960094	0.960615	0.168305	7.1906035
peru	religion	3	548250	0.965756	567690	0.000473	0.965282	0.966229	0.168305	7.2046300
peru	religion	4	465110	0.936419	496690	0.000678	0.935740	0.937097	0.168305	7.1896815
peru	sex	Female	8545700	0.963723	28867380	0.000123	0.963600	0.963846	0.168305	7.2029180
peru	sex	Male	8317770	0.970586	48569840	0.000113	0.970473	0.970699	0.168305	7.1833105

# Zambia Legal ID by Education, Ethnicity, and Religion

```
zambia <-
  ipums_legal_age %>%
```



```

subset(country == "zambia",
       select = c(country, geolevl, sex, age2, edattain, zm2010a_relig, ethniczm, zm2010a_nidcard)) %>%
mutate(zm2010a_nidcard = recode(zm2010a_nidcard, "1" = "yes", "2" = "no",
                               "8" = "unknown", "9" = "niu")) %>%
rename_at (.vars = c("zm2010a_relig", "ethniczm", "zm2010a_nidcard"),
          ~c("religion", "ethnic", "id") ) %>%
subpop_summary() %>%
cbind(., runif(103, 2.7, 3.25)) %>%
rename(coord = "...11")

```

```

## Warning: attributes are not identical across measure variables;
## they will be dropped

```

```

## `summarise()` regrouping output by 'country', 'type', 'subpop' (override with `.groups` argument)

```

```

## New names:
## * NA -> ...10

```

```

## New names:
## * NA -> ...11

```

```

zambia %>%
  kable(format = "pipe", caption = "Legal ID Possession by Subpopulations")

```

Table 4: Legal ID Possession by Subpopulations

country	type	subpop	n_id	prop_id	total_subpop	Std Err	Lower CI	Upper CI	Range	coord
zambia	age2	15-19	564900	0.4646209	1215830	0.0008865	0.4637343	0.4655074	0.7333504	0.119873
zambia	age2	20-24	928500	0.7668104	1210860	0.0007532	0.7660572	0.7675635	0.7333504	0.804571
zambia	age2	25-29	905800	0.8464709	1070090	0.0006830	0.8457878	0.8471539	0.7333504	0.051003
zambia	age2	30-34	735710	0.8652154	850320	0.0007258	0.8644890	0.8659413	0.7333504	0.951858
zambia	age2	35-39	597550	0.8657510	690210	0.0008043	0.8649467	0.8665553	0.7333504	0.056466
zambia	age2	40-44	413540	0.8663789	477320	0.0009652	0.8654137	0.8673442	0.7333504	0.125794
zambia	age2	45-49	331360	0.8719082	380040	0.0010625	0.8708457	0.8729707	0.7333504	0.057504
zambia	age2	50-54	251470	0.8716464	288500	0.0012205	0.8704259	0.8728670	0.7333504	0.734579
zambia	age2	55-59	169580	0.8731336	194220	0.0014802	0.8716534	0.8746137	0.7333504	0.015915
zambia	age2	60-64	149080	0.8731916	170730	0.0015784	0.8716132	0.8747700	0.7333504	0.184072
zambia	age2	65-69	108080	0.8647092	124990	0.0018962	0.8628130	0.8666054	0.7333504	0.178358
zambia	age2	70-74	83320	0.8647639	96350	0.0021593	0.8626040	0.8669232	0.7333504	0.827685
zambia	age2	75-79	56480	0.8507305	66390	0.0027107	0.8480199	0.8534412	0.7333504	0.715880
zambia	age2	80+	56710	0.8223608	68960	0.0028527	0.8195081	0.8252134	0.7333504	0.716510
zambia	edattain	Less than primary	16702500	0.8117112	2057690	0.0005342	0.8111770	0.8122454	0.7333504	0.199149
zambia	edattain	Post-secondary	44780	0.8698524	51480	0.0029065	0.8669459	0.8727589	0.7333504	0.210455
zambia	edattain	Primary	25593900	0.8033164	3186030	0.0004365	0.8028799	0.8037528	0.7333504	0.082931
zambia	edattain	Secondary	10776600	0.9480101	1136760	0.0004081	0.9476020	0.9484182	0.7333504	0.215688
zambia	ethnic	1	11654600	0.8011907	1454660	0.0006480	0.8005421	0.8018392	0.7333504	0.881040
zambia	ethnic	10	36170	0.7579631	47720	0.0038429	0.7541202	0.7618061	0.7333504	0.156599
zambia	ethnic	11	18990	0.7608173	24960	0.0052921	0.7555252	0.7661094	0.7333504	0.819350
zambia	ethnic	12	2200	0.7638889	2880	0.0155105	0.7483784	0.7793994	0.7333504	0.158355



country	type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_CI	Upper_CI	Range	coord
zambia	ethnic	13	1060	0.7517730	1410	0.0225479	0.7292251	0.7743210	0.7333502	8.20006
zambia	ethnic	14	2960	0.8000000	3700	0.0128886	0.7871114	0.8128886	0.7333502	9.65018
zambia	ethnic	15	8650	0.7972350	10850	0.0075652	0.7896698	0.8048003	0.7333502	0.74230
zambia	ethnic	16	10330	0.7612380	13570	0.0071730	0.7540650	0.7684110	0.7333502	0.719246
zambia	ethnic	17	23270	0.7632010	30490	0.0047718	0.7584293	0.7679728	0.7333502	0.863103
zambia	ethnic	19	675940	0.7652093	883340	0.0008839	0.7643254	0.7660932	0.7333502	0.037839
zambia	ethnic	2	57000	0.8252497	69070	0.0028321	0.8224177	0.8280818	0.7333502	0.160127
zambia	ethnic	20	85900	0.7813353	109940	0.0024433	0.7788920	0.7837780	0.7333502	0.916807
zambia	ethnic	21	39350	0.8012625	49110	0.0035293	0.7977332	0.8047918	0.7333502	0.181509
zambia	ethnic	22	40060	0.7510311	53340	0.0036696	0.7473615	0.7547008	0.7333502	0.715836
zambia	ethnic	23	25520	0.8148148	31320	0.0043020	0.8105128	0.8191168	0.7333502	0.846673
zambia	ethnic	24	6890	0.7521834	9160	0.0088415	0.7433419	0.7610249	0.7333502	0.965201
zambia	ethnic	25	10130	0.8235772	12300	0.0067364	0.8168409	0.8303136	0.7333502	0.068497
zambia	ethnic	26	113430	0.7768114	146020	0.0021357	0.7746757	0.7789471	0.7333502	0.245358
zambia	ethnic	27	135650	0.7850570	172790	0.0019369	0.7831201	0.7869939	0.7333502	0.739543
zambia	ethnic	28	61490	0.7375555	83370	0.0029863	0.7345690	0.7405420	0.7333502	0.046364
zambia	ethnic	29	21510	0.7816134	27520	0.0048813	0.7767321	0.7864946	0.7333502	0.982174
zambia	ethnic	3	166760	0.7804558	213670	0.0017551	0.7787007	0.7822110	0.7333502	0.873722
zambia	ethnic	30	6170	0.8075916	7640	0.0088391	0.7987523	0.8164307	0.7333502	0.918086
zambia	ethnic	31	800	0.8080808	990	0.0245311	0.7835497	0.8326119	0.7333502	0.712102
zambia	ethnic	32	25510	0.7690684	33170	0.0045352	0.7645332	0.7736037	0.7333502	0.033573
zambia	ethnic	33	154710	0.7821932	197790	0.0018190	0.7803742	0.7840123	0.7333502	0.727982
zambia	ethnic	35	10160	0.7720365	13160	0.0071676	0.7648689	0.7792040	0.7333502	0.783001
zambia	ethnic	36	2380	0.7437500	3200	0.0151258	0.7286242	0.7588758	0.7333502	0.097888
zambia	ethnic	37	6160	0.7315914	8420	0.0094651	0.7221264	0.7410565	0.7333502	0.041272
zambia	ethnic	38	7110	0.8163031	8710	0.0081323	0.8081708	0.8244354	0.7333502	0.236352
zambia	ethnic	39	1900	0.7011070	2710	0.0172351	0.6838719	0.7183421	0.7333502	0.166808
zambia	ethnic	40	2210	0.7700348	2870	0.0153953	0.7546394	0.7854303	0.7333502	0.791457
zambia	ethnic	41	1650	0.7710280	2140	0.0178019	0.7532261	0.7888300	0.7333502	0.243990
zambia	ethnic	43	322350	0.7700853	418590	0.0012747	0.7688106	0.7713600	0.7333502	0.014231
zambia	ethnic	44	3890	0.8070539	4820	0.0111402	0.7959137	0.8181942	0.7333502	0.865823
zambia	ethnic	45	2610	0.7435897	3510	0.0144454	0.7291444	0.7580351	0.7333502	0.813003
zambia	ethnic	46	25420	0.7549748	33670	0.0045941	0.7503807	0.7595688	0.7333502	0.173484
zambia	ethnic	47	480	0.7058824	680	0.0342468	0.6716353	0.7401292	0.7333502	0.795745
zambia	ethnic	48	396890	0.7800358	508810	0.0011382	0.7788976	0.7811739	0.7333502	0.930887
zambia	ethnic	49	289750	0.7888647	367300	0.0013198	0.7875449	0.7901845	0.7333502	0.716196
zambia	ethnic	5	104200	0.8136176	128070	0.0021327	0.8114848	0.8157503	0.7333502	0.828932
zambia	ethnic	50	226030	0.7875610	287000	0.0014963	0.7860643	0.7890574	0.7333502	0.993526
zambia	ethnic	51	20860	0.7720207	27020	0.0050023	0.7670183	0.7770230	0.7333502	0.915752
zambia	ethnic	52	38910	0.7973361	48800	0.0035663	0.7937693	0.8009026	0.7333502	0.241622
zambia	ethnic	53	15020	0.8293760	18110	0.0054788	0.8238972	0.8348548	0.7333502	0.839874
zambia	ethnic	54	41310	0.7482340	55210	0.0036204	0.7446136	0.7518544	0.7333502	0.813179
zambia	ethnic	55	134290	0.7761979	173010	0.0019640	0.7742340	0.7781619	0.7333502	0.771666
zambia	ethnic	56	152260	0.7970476	191030	0.0018036	0.7952440	0.7988512	0.7333502	0.028658
zambia	ethnic	57	1150	0.7770270	1480	0.0212061	0.7558209	0.7982332	0.7333502	0.038399
zambia	ethnic	58	2250	0.7839721	2870	0.0150561	0.7689160	0.7990282	0.7333502	0.766108
zambia	ethnic	59	237610	0.7764779	306010	0.0014761	0.7750018	0.7779540	0.7333502	0.997162
zambia	ethnic	6	28870	0.8315092	34720	0.0039371	0.8275721	0.8354463	0.7333502	0.704401
zambia	ethnic	60	45910	0.7910062	58040	0.0033078	0.7876984	0.7943140	0.7333502	0.775891
zambia	ethnic	7	36280	0.7938731	45700	0.0037088	0.7901643	0.7975819	0.7333502	0.915392
zambia	ethnic	8	114730	0.8182142	140220	0.0020186	0.8161956	0.8202329	0.7333502	0.987249
zambia	ethnic	9	19330	0.7997518	24170	0.0050451	0.7947066	0.8047969	0.7333502	0.931215

country	type	subpop	n_id	prop_id	total_subpop	Std_Err	Lower_CI	Upper_CI	Range	coord
zambia	ethnic	American	330	0.2374101	1390	0.0223684	0.2150410	0.2597785	0.7333504	2.840067
zambia	ethnic	Asian	2680	0.3055872	8770	0.0096411	0.2959462	0.3152283	0.7333504	3.133225
zambia	ethnic	Bisa	84800	0.7645839	110910	0.0024969	0.7620870	0.7670808	0.7333504	3.092108
zambia	ethnic	European	410	0.2146597	1910	0.0184135	0.1962462	0.2330732	0.7333504	3.069628
zambia	ethnic	Luano	530	0.6022727	880	0.0323367	0.5699360	0.6346095	0.7333504	2.866870
zambia	ethnic	Luyana	370	0.6379310	580	0.0391120	0.5988184	0.6770430	0.7333504	2.760195
zambia	ethnic	Mashi	16730	0.6673315	25070	0.0058324	0.6614991	0.6731639	0.7333504	2.811963
zambia	ethnic	Other	25600	0.6380857	40120	0.0047023	0.6333834	0.6427880	0.7333504	2.818822
zambia	ethnic	Other	18850	0.4292872	43910	0.0046297	0.4246575	0.4339168	0.7333504	3.249000
		African								
zambia	ethnic	Unspecified	13440	0.2249372	59750	0.0033479	0.2215893	0.2282852	0.7333504	3.038298
zambia	ethnic	Yombe	460	0.6666667	690	0.0351730	0.6314930	0.7018403	0.7333504	2.739841
zambia	geolev1	894001	496870	0.7393789	672010	0.0010495	0.7383293	0.7404284	0.7333504	2.887687
zambia	geolev1	894002	932760	0.8284867	1125860	0.0006963	0.8277904	0.8291830	0.7333504	3.151350
zambia	geolev1	894003	12879400	0.7535735	1709110	0.0006461	0.7529274	0.7542190	0.7333504	2.790933
zambia	geolev1	894004	396490	0.7946965	498920	0.0011208	0.7935757	0.7958174	0.7333504	3.209790
zambia	geolev1	894005	10205100	0.8051297	1267510	0.0006890	0.8044402	0.8058193	0.7333504	2.780847
zambia	geolev1	894008	268890	0.7468337	360040	0.0014203	0.7454134	0.7482540	0.7333504	2.877690
zambia	geolev1	894009	608590	0.7576218	803290	0.0009371	0.7566847	0.7585589	0.7333504	2.870095
zambia	geolev1	894010	340030	0.7264512	468070	0.0012771	0.7251741	0.7277282	0.7333504	2.808313
zambia	religion	1	11311800	0.8402514	1346240	0.0006189	0.8396325	0.8408703	0.7333504	2.848552
zambia	religion	2	39657800	0.8327517	4762260	0.0003352	0.8324165	0.8330869	0.7333504	3.154692
zambia	religion	5	3660	0.8206278	4460	0.0112598	0.8093680	0.8318870	0.7333504	2.829695
zambia	religion	6	1580	0.8540541	1850	0.0160880	0.8379661	0.8701420	0.7333504	2.788056
zambia	religion	7	106690	0.8044183	132630	0.0021347	0.8022830	0.8065530	0.7333504	2.749134
zambia	religion	8	119830	0.8040123	149040	0.0020153	0.8019970	0.8060277	0.7333504	2.875273
zambia	religion	Hindu	930	0.2915361	3190	0.0157709	0.2757651	0.3073070	0.7333504	3.120063
zambia	religion	Muslim	22430	0.6946423	32290	0.0050234	0.6896189	0.6996657	0.7333504	3.058370
zambia	sex	Female	27931900	0.7884510	3542630	0.0004253	0.7880257	0.7888762	0.7333504	3.155592
zambia	sex	Male	25588900	0.7610806	3362180	0.0004558	0.7606248	0.7615364	0.7333504	2.773713

#2. Prepare (India) Data

## Set Up

```
library(dplyr)
library(ggplot2)
library(here)
```

```
## here() starts at /Users/sjmardevi/Documents/R Projects Here
```

```
library(magrittr)
```

```
##
```

```
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
## set_names
```

```
## The following object is masked from 'package:tidyr':
##
##      extract
```

```
pulse_mem <- read.csv("../dataverse_files/Inputs/Datasets/20191127_State of Aadhaar_Pulse survey_mem.csv")
```

## Country Level ID Ownership Averages

```
pulse_mem <-
  pulse_mem %>%
  mutate(education = recode(education, "1st Std. Pass" = "At least some primary",
    "1st Std. Pass" = "At least some primary",
    "2nd Std. Pass" = "At least some primary",
    "3rd Std. Pass" = "At least some primary",
    "4th Std. Pass" = "At least some primary",
    "5th Std. Pass" = "At least some primary",
    "6th Std. Pass" = "At least some primary",
    "7th Std. Pass" = "At least some primary",
    "8th Std. Pass" = "At least some primary",
    "9th Std. Pass" = "At least some secondary",
    "11th Std. Pass" = "At least some secondary",
    "12th Std. Pass" = "At least some secondary",
    "10th Std. Pass" = "At least some secondary",
    "Graduate" = "Post-secondary",
    "Ph.D / M.Phil" = "Post-secondary",
    "Post Graduate" = "Post-secondary"))
```

```
tab_overall <-
  pulse_mem %>%
  group_by(has_aadhaar) %>%
  summarise(weighted_n =
    sum(mem_weight_extn_all_hr_regtype_strata_within_state_sample_wave)) %>%
  mutate(population = sum(weighted_n),
    weighted_proportion = weighted_n / population)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
aadhaar_enrollment_overall <- tab_overall %>% filter(has_aadhaar == "Yes") %>%
  select(weighted_proportion) %>%
  pull()
```

```
tab_gender <-
  pulse_mem %>%
  group_by(gender, has_aadhaar) %>%
  summarise(weighted_n =
    sum(mem_weight_extn_all_hr_regtype_strata_within_state_sample_wave)) %>%
  mutate(population = sum(weighted_n),
    weighted_proportion = weighted_n / population) %>%
  as.data.frame()
```

```
## `summarise()` regrouping output by 'gender' (override with `.groups` argument)
```

```

aadhaar_enrollment_males <- tab_gender %>%
  filter(has_aadhaar == "Yes" & gender == "M") %>%
  select(weighted_proportion) %>%
  pull()
aadhaar_enrollment_females <- tab_gender %>%
  filter(has_aadhaar == "Yes" & gender == "F") %>%
  select(weighted_proportion) %>%
  pull()

```

## Population Subgroup Level ID Ownership Tabulations

```

#By Religion
tab_religion <-

pulse_mem %>%
  group_by(religion, has_aadhaar) %>%
  summarise(weighted_n =
    sum(mem_weight_extn_all_hr_regtype_strata_within_state_sample_wave)) %>%
  mutate(population = sum(weighted_n),
    weighted_proportion = weighted_n / population) %>%
  filter(has_aadhaar == "Yes") %>%
  rename("subpop" = "religion") %>%
  mutate("type" = "religion", "country" = "India") %>%
  as.data.frame()

## `summarise()` regrouping output by 'religion' (override with `.groups` argument)

```

```

#By State (ADM 1)
tab_state <-

pulse_mem %>%
  group_by(state, has_aadhaar) %>%
  summarise(weighted_n =
    sum(mem_weight_extn_all_hr_regtype_strata_within_state_sample_wave)) %>%
  mutate(population = sum(weighted_n),
    weighted_proportion = weighted_n / population) %>%
  filter(has_aadhaar == "Yes") %>%
  rename("subpop" = "state") %>%
  mutate("type" = "state", "country" = "India") %>%
  as.data.frame()

## `summarise()` regrouping output by 'state' (override with `.groups` argument)

```

```

#By Gender
tab_gender <-

pulse_mem %>%
  group_by(gender, has_aadhaar) %>%
  summarise(weighted_n =
    sum(mem_weight_extn_all_hr_regtype_strata_within_state_sample_wave)) %>%

```

```
mutate(population = sum(weighted_n),
       weighted_proportion = weighted_n / population) %>%
filter(has_aadhaar == "Yes") %>%
rename("subpop" = "gender") %>%
mutate("type" = "gender", "country" = "India") %>%
as.data.frame()
```

## `summarise()` regrouping output by 'gender' (override with `.groups` argument)

*#By Education*

tab\_education <-

```
pulse_mem %>%
  group_by(education, has_aadhaar) %>%
  summarise(weighted_n =
    sum(mem_weight_extn_all_hr_regtype_strata_within_state_sample_wave)) %>%
  mutate(population = sum(weighted_n),
         weighted_proportion = weighted_n / population) %>%
  filter(has_aadhaar == "Yes") %>%
  rename("subpop" = "education") %>%
  mutate("type" = "edattain", "country" = "India") %>%
  as.data.frame()
```

## `summarise()` regrouping output by 'education' (override with `.groups` argument)

*# Bind all tabs by row*

india <- rbind(tab\_state, tab\_gender, tab\_religion, tab\_education)

*# Clean data for Merging*

```
india <-
  india %>%
  select(-has_aadhaar) %>%
  rename_at(vars(weighted_n, population, weighted_proportion), ~c("n_id", "total_subpop", "prop_id")) %>%
  mutate("Std_Err" = NA, "Lower_CI" = NA, "Upper_CI" = NA) %>%
  select(country, type, subpop, n_id, prop_id, total_subpop, everything())
```

range\_india <- max(india\$prop\_id) - min(india\$prop\_id)

```
india <-
  india %>%
  cbind(., range_india) %>%
  rename("Range" = range_india) %>%
  cbind(., runif(44, 5.7, 6.3)) %>%
  rename(coord = "runif(44, 5.7, 6.3)")
```

### 3. Prepare Myanmar Data

```
G1_Myanmar <-
  read_excel(here::here("./Myanmar Project/G1_Myanmar.xlsx"))
```

```

G2_Myanmar <-
  read_excel(here::here("./Myanmar Project/G2_Myanmar.xlsx"))

# For Sex Dis. Data
union <- read_excel("./Myanmar Project/union_myanmar_modif.xlsx")
union <-
  union %>%
  clean_names() %>%
  select(-age)
union <- aggregate(. ~sex, data=union, sum)

agg_G1 <-
  G1_Myanmar %>%
  select(-c(state_area_name)) %>%
  filter(age_group!="Total") %>%
  mutate(age_group = gsub(" ", "", age_group))
agg_G1 <- aggregate(. ~age_group, data = agg_G1, sum)

myanmar_age <-
  agg_G1 %>%
  select(c(age_group:citizenship_scrutiny_card, national_registration_card)) %>%
  mutate("n_id" = citizenship_scrutiny_card + national_registration_card,
         "prop_id" = n_id/total,
         "country" = "myanmar",
         "type" = "age2",
         "Std_Err" = NA,
         "Lower_CI" = NA,
         "Upper_CI" = NA) %>%
  rename_at(vars(age_group, total), ~c("subpop", "total_subpop")) %>%
  select(c(country, type, subpop, n_id, prop_id, total_subpop, Std_Err, Lower_CI, Upper_CI))

agg_state <-
  G2_Myanmar %>%
  select(2,7:17)
agg_state <- aggregate(. ~state_region_name+urban_rural, data=agg_state, sum)

myanmar_state <-
  agg_state %>%
  select(c(state_region_name:citizenship_scrutiny_card, national_registration_card)) %>%
  filter(urban_rural == "Total") %>%
  mutate("n_id" = citizenship_scrutiny_card + national_registration_card,
         "prop_id" = n_id/total,
         "country" = "myanmar",
         "type" = "geolev1",
         "Std_Err" = NA,
         "Lower_CI" = NA,
         "Upper_CI" = NA) %>%
  rename_at(vars(state_region_name, total), ~c("subpop", "total_subpop")) %>%
  select(c(country, type, subpop, n_id, prop_id, total_subpop, Std_Err, Lower_CI, Upper_CI))

```

```

myanmar_sex <-
  union %>%
  clean_names() %>%
  select(c(sex, total:citizenship_scrutiny_card, national_registration_card)) %>%
  mutate("n_id" = citizenship_scrutiny_card + national_registration_card,
         "prop_id" = n_id/total,
         "country" = "myanmar",
         "type" = "sex",
         "Std_Err" = NA,
         "Lower_CI" = NA,
         "Upper_CI" = NA) %>%
  rename_at(vars(sex, total), ~c("subpop", "total_subpop")) %>%
  select(c(country, type, subpop, n_id, prop_id, total_subpop, Std_Err, Lower_CI, Upper_CI))

```

```

avg_myanmar <-
  agg_state %>%
  filter(urban_rural=="Total") %>%
  select(c(urban_rural, total, citizenship_scrutiny_card, national_registration_card)) %>%
  mutate(id = citizenship_scrutiny_card + national_registration_card)

avg_m <- sum(avg_myanmar$id)/sum(avg_myanmar$total)

```

```

myanmar<-bind_rows(myanmar_state, myanmar_sex, myanmar_age)

range_myanmar <-max(myanmar$prop_id)-min(myanmar$prop_id)

myanmar <-
  myanmar %>%
  cbind(., range_myanmar) %>%
  rename("Range" = range_myanmar) %>%
  cbind(., runif(34, 3.8, 4.3)) %>%
  rename(coord = "runif(34, 3.8, 4.3)")

```

## Bind all Countries

```

multi_emppzi <-
  bind_rows(ecuador, myanmar, pakistan, peru, zambia, india)

```

```

multi_emppzi <-
  multi_emppzi %>%
  mutate(type = recode(type, "ethnic" = "ethnicity/race", "race" = "ethnicity/race", "state" = "geolevel"))
  mutate(country = reorder(country, Range))

```

```

ggplot(multi_emppzi, aes(x = prop_id,
                        y = coord,
                        #factor(country,
                        #levels = c("peru","ecuador", "zambia",
                        #"myanmar", "pakistan", "India")),
                        color = country,
                        shape = type,

```



```

        size = total_subpop)) +
geom_jitter() +
theme_minimal() +
scale_y_continuous(breaks = c(1, 2, 3, 4, 5, 6)) +
geom_vline(xintercept = 0, color = "#999999") +
#Overall country proportions of ID ownership
geom_vline(xintercept = 0.98, color = "#D5E00", linetype = "dashed") +
geom_vline(xintercept = 0.97, color = "#0072B2", linetype = "dashed") +
geom_vline(xintercept = aadhaar_enrollment_overall, color = "#999900", linetype = "dashed") +
geom_vline(xintercept = 0.64, color = "#33CCCC", linetype = "dashed") +
geom_vline(xintercept = 0.78, color = "purple", linetype = "dashed") +
geom_vline(xintercept = avg_m, color = "#009E73", linetype = "dashed") +
geom_text_repel(data=subset(multi_emppzi, ((country == "ecuador" & prop_id<0.94)|
      (country == "ecuador" & prop_id>0.992)|
      (country == "peru" & prop_id<0.90)|
      (country == "peru" & prop_id>0.987)|
      (country == "zambia" & prop_id<0.68 & !(subpop %in% c("Asian", "American"
      (country == "zambia" & prop_id>0.872)|
      (country == "myanmar" & prop_id<0.64)|
      (country == "myanmar" & prop_id>0.88)|
      (country == "pakistan" & prop_id>0.89)|
      (country == "pakistan" & prop_id<0.530)|
      (country == "india" & prop_id<0.75)|
      (country == "india" & prop_id>0.995)|
      (country == "pakistan" & subpop == "Male"))),
      aes(prop_id, coord, label=subpop),
      size = 2.75, vnuage = 1.5, hnuage = 0.275, color = "Black", fontface = "bold") +
scale_size_continuous(name = "Population Subgroup Total",
      breaks = c(1000, 100000, 100000000, 1000000000),
      labels = c("1,000", "100,000", "10,000,000",
        "1,000,000,000")) +
labs(title = "Percentage of Population with Proof of Legal Identity by Population Subgroup, Various C
      x = "% Population Subgroup with Proof of Legal Identity", y = element_blank()) +
scale_shape_discrete(name = "Population Subgroup Type",
      labels = c("Age Group", "Education Attainment", "Ethnicity/Race",
        "Adm. 1 Region", "Religion", "Sex")) +
theme(axis.text.y = element_blank(),
      axis.title.y = element_blank()) +
annotate("text", x = -0.15:0, y = 1, label = "Ecuador\n(Census, 2010)", fontface = "bold", color = "#
annotate("text", x = -0.15:0, y = 2, label = "Peru\n(Census, 2007)", fontface = "bold", color = "#007
annotate("text", x = -0.15:0, y = 3, label = "Zambia\n(Census, 2010)", fontface = "bold", color = "pu
annotate("text", x = -0.15:0, y = 4, label = "Myanmar\n(Census, 2014)", fontface = "bold", color = "#
annotate("text", x = -0.15:0, y = 5, label = "Pakistan\n(Census, 1998)", fontface = "bold", color = "
annotate("text", x = -0.15:0, y = 6, label = "India\n(Household Survey, 2019)", fontface = "bold", co
scale_x_continuous(labels = function(prop_id) paste0(prop_id*100, "%"), limits = c(-0.25, 1), breaks =
guides(color=FALSE)

```

```
## Warning: Ignoring unknown parameters: vnuage, hnuage
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Percentage of Population with Proof of Legal Identity by Population Subgroup, Various Countries (1998–2019)



## Multi-methods/data sources Comparison Analysis

```
comp<-read_excel("comp.xlsx")
comp_base<-
  comp %>%
  gather(-Country, key = "data", value = "prop") %>%
  mutate(x = ifelse(Country == "Zambia", 1,
    ifelse(Country == "Pakistan", 3,
      ifelse(Country == "Myanmar", 2,
        ifelse(Country == "India", 4,
          ifelse(Country == "Peru", 5,
            ifelse(Country == "Ecuador", 6, NA)))))),
    x1 = x-0.25,
    x2 = x+0.25,
    prop1 = prop-0.008,
    prop2 = prop+0.008) %>%
  filter(!(data %in% c("INDEX_recalc", "birth_reg", "voters_reg")))
```

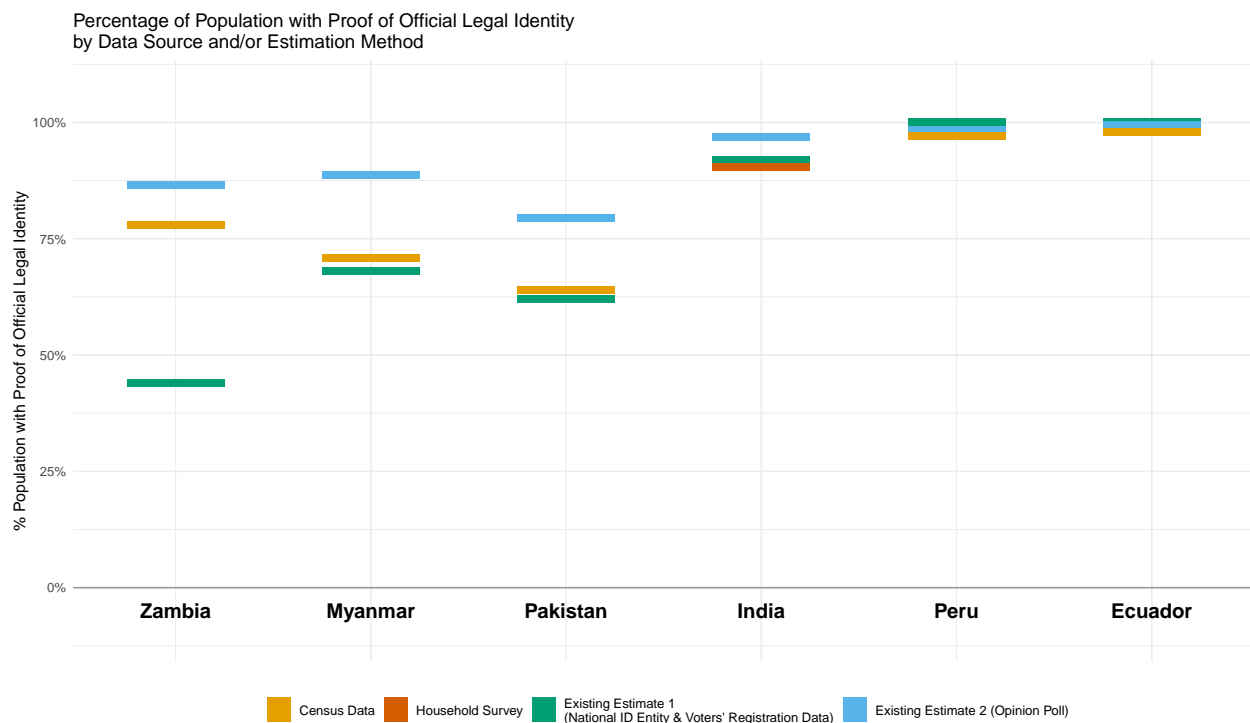
```
ggplot(comp_base, aes(xmin = x1, xmax = x2, ymin = prop1, ymax = prop2, fill = data), color = "black",
  geom_rect() +
  scale_y_continuous(labels = function(prop) paste0(prop*100, "%"), limits = c(-0.1, 1.075), breaks = c(
    labs(title = "Percentage of Population with Proof of Official Legal Identity\nby Data Source and/or L",
    x = element_blank(),
    y = "% Population with Proof of Official Legal Identity") +
  scale_fill_manual(name = element_blank(),
    breaks=c("Census", "Household Survey", "INDEX", "FINDEX"),
    labels = c("Census Data", "Household Survey",
```

```

    "Existing Estimate 1\n(National ID Entity & Voters' Registration Data)",
    "Existing Estimate 2 (Opinion Poll)",
    values = c("#E69F00", "#D55E00", "#009E73", "#56B4E9")) +
  annotate("text", x = 1, y = -0.05, label = "Zambia", fontface = "bold", size=5) +
  annotate("text", x = 2, y = -0.05, label = "Myanmar", fontface = "bold", size=5) +
  annotate("text", x = 3, y = -0.05, label = "Pakistan", fontface = "bold", size=5) +
  annotate("text", x = 4, y = -0.05, label = "India", fontface = "bold", size=5) +
  annotate("text", x = 5, y = -0.05, label = "Peru", fontface = "bold", size=5) +
  annotate("text", x = 6, y = -0.05, label = "Ecuador", fontface = "bold", size=5) +
  geom_hline(yintercept = 0, color = "#999999") +
  theme_minimal() +
  guides(size=FALSE) +
  theme(legend.position="bottom")+
  theme(axis.text.x = element_blank())

```

## Warning: Removed 6 rows containing missing values (geom\_rect).



With additional estimates

```

comp_extra<-
  comp %>%
  gather(-Country, key = "data", value = "prop") %>%
  mutate(x = ifelse(Country == "Zambia", 1,
    ifelse(Country == "Pakistan", 3,
      ifelse(Country == "Myanmar", 2,
        ifelse(Country == "India", 4,
          ifelse(Country == "Peru", 5,

```

```

                                                    ifelse(Country == "Ecuador", 6, NA))))),
  x1 = x-0.25,
  x2 = x+0.25,
  prop1 = prop-0.008,
  prop2 = prop+0.008)

ggplot(comp_extra, aes(xmin = x1, xmax = x2, ymin = prop1, ymax = prop2, fill = data), color = "black",
  geom_rect() +
  scale_y_continuous(labels = function(prop) paste0(prop*100, "%"), limits = c(-0.1, 1.075), breaks = c(
    labs(title = "Percentage of Population with Proof of Official Legal Identity\nby Data Source and/or I
      x = element_blank(),
      y = "% Population with Proof of Official Legal Identity") +
  scale_fill_manual(name = element_blank(),
    breaks=c("Census", "Household Survey", "INDEX", "FINDEX", "INDEX_recalc", "birth_reg",
    labels = c("Census Data", "Household Survey",
      "ID4D-INDEX (UNICEF Birth Registration Database\nNational ID Entity & Vo
      "ID4D-FINDEX (Gallup Opinion Poll)",
      "Census & ID4D-INDEX Estimate Approach",
      "Census-Birth Registration",
      "Census-Voters' Registration"),
    values = c("#E69F00", "#D55E00", "#009E73", "#56B4E9", "#999999", "#CC79A7", "#0072B2")
  annotate("text", x = 1, y = -0.05, label = "Zambia", fontface = "bold", size=5) +
  annotate("text", x = 2, y = -0.05, label = "Myanmar", fontface = "bold", size=5) +
  annotate("text", x = 3, y = -0.05, label = "Pakistan", fontface = "bold", size=5) +
  annotate("text", x = 4, y = -0.05, label = "India", fontface = "bold", size=5) +
  annotate("text", x = 5, y = -0.05, label = "Peru", fontface = "bold", size=5) +
  annotate("text", x = 6, y = -0.05, label = "Ecuador", fontface = "bold", size=5) +
  geom_hline(yintercept = 0, color = "#999999") +
  theme_minimal() +
  guides(size=FALSE) +
  theme(legend.position="bottom")+
  theme(axis.text.x = element_blank())

```

```
## Warning: Removed 20 rows containing missing values (geom_rect).
```

