Statistics of the Russian Empire: Digitized Data Tables

Egor Malkov[1]

This version: June 2021

Contents

---

[1] University of Minnesota and the Federal Reserve Bank of Minneapolis.
E-mail: malko017@umn.edu
Website: sites.google.com/a/umn.edu/emalkov
Typos in the original documents are corrected to the best of my knowledge.

Territory

File: **territory.dta**. The data covers 1904-1916.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
area — Area without considerable internal waters (square miles).
n_cities — Number of cities.

Population

File: **population.dta**. The data covers 1851, 1863, 1885, 1897, and 1904-1916.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
male_rural — Rural male population (thousands).
female_rural — Rural female population (thousands).
rural_pop — Rural population (thousands).
male_urban — Urban male population (thousands).
female_urban — Urban female population (thousands).
urban_pop — Urban population (thousands).
male_total — Total male population (thousands).
female_total — Total female population (thousands).
total_pop — Total population (thousands).

Ethnic Composition

File: **ethnicity.dta**. The data is from the Russian Imperial Census of 1897.

Variables

region_id — Region ID.[2]
region — Region.
macroregion — Macroregion.
year — Year.
russians — Russians (%).
tatars — Tatars (%).
poles — Poles (%).
finnish — Finnish (%).
jews — Jews (%).
lithuanians — Lithuanians (%).
germans — Germans (%).
georgians — Georgians (%).
caucasian_highlanders — Caucasian highlanders (%).
armenians — Armenians (%).
mongols — Mongols (%).
other — Other ethnicities (%).

---

[2] Data for Batum Oblast and Kutais Governorate as well as for Primorskaya Oblast and Sakhalin Oblast is reported jointly, so I do not assign region IDs to these observations.

Religious Composition

File: **religion.dta**. The data is from the Russian Imperial Census of 1897.

Variables

region_id — Region ID.[3]
region — Region.
macroregion — Macroregion.
year — Year.
orthodox_christians — Orthodox Christians (%).
muslims — Muslims (%).
catholics — Catholics (%).
protestants — Protestants (%).
judaists — Judaists (%).
other_christians — Other Christians (%).
other_non_christians — Other Non-Christians (%).

---

[3] Data for Batum Oblast and Kutais Governorate as well as for Primorskaya Oblast and Sakhalin Oblast is reported jointly, so I do not assign region IDs to these observations.

Estate Composition

File: **estates.dta**. The data is from the Russian Imperial Census of 1897.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
nobles_civ_serv — Nobles and civil cervants (%).
clergy — Clergy (%).
not_bourg_merch — Notable bourgeois and merchants (%).
bourgeois — Bourgeois (%).
peasants — Peasants (%).
cossacks — Cossacks (%).
indigenous — Indigenous peoples (%).
other — Other (%).

Primary Education and Literacy

File: **education.dta**. The data covers 1897, 1900, 1903, and 1909-1914.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
male_literacy — Literacy (%), males.
female_literacy — Literacy (%), females.
literacy — Literacy (%), total.
literacy_excl_kids — Literacy (%) excluding kids under 9 years old, total.
male_literacy_excl_kids — Literacy (%) excluding kids under 9 years old, males.
n_schools — Total number of elementary schools.
total_students — Total number of students in the elementary schools.
male_students — Number of male students in the elementary schools.
female_students — Number of female students in the elementary schools.

Marriages

File: **marriage.dta**. The data covers 1899-1914.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
n_marriages — Number of marriages.

Internal Migration

File: **internal_migration.dta**. The data covers 1896-1907.

Variables

`region_id` — Region ID.
`region` — Region.
`macroregion` — Macroregion.
`year` — Year.
`n_migrants` — Number of people migrated to Siberia through Chelyabinsk.

Transportation

Files: **railway_transport.dta** and **water_transport.dta**. The data covers 1904-1912. One verst = 1.0668 kilometers.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
main_lines — Length of main railway lines (versts).
branch_lines — Length of branch railway lines (versts).
total_lines — Total length of railway lines (versts).
raft_waterways — Length of waterways for rafts (versts).
ship_down — Length of waterways for shipping down the river (versts).
ship_both — Length of waterways for shipping, both directions (versts).
total_waterways — Total length of waterways (versts).

Identities:

main_lines + branch_lines = total_lines

raft_waterways + ship_down + ship_both = total_waterways

Occupational Composition

File: **occupations.dta**. The data is from the Russian Imperial Census of 1897.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
admin_law — Administration, law, police (%).
military — Military (%).
clergy — Clergy (%).
services — Services, servants, jorneymen (%).
rentiers — Rentiers (%).
agriculture — Agriculture, forestry, fishing, hunting (%).
manufacturing — Manufacturing and mining (%).
transport — Transportation (%).
commerce — Commerce (%).
other — Other occupations (%).

Labor Income

File: **labor_income.dta**. The data covers 1913-1915.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
wage_cotton — Average annual wage income of workers in cotton processing industry (rubles).
wage_wool — Average annual wage income of workers in wool processing industry (rubles).
wage_silk — Average annual wage income of workers in silk processing industry (rubles).
wage_linum — Average annual wage income of workers in linum and hemp processing industry (rubles).
wage_mixed_textile — Average annual wage income of workers in mixed textile processing industry (rubles).
wage_paper — Average annual wage income of workers in paper processing industry (rubles).
wage_wood — Average annual wage income of workers in wood processing industry (rubles).
wage_metal — Average annual wage income of workers in metal processing industry (rubles).
wage_mineral — Average annual wage income of workers in mineral processing industry (rubles).
wage_animal — Average annual wage income of workers in animal products processing industry (rubles).
wage_food — Average annual wage income of workers in food processing industry (rubles).
wage_chemical — Average annual wage income of workers in chemical industry (rubles).
wage_extraction — Average annual wage income of workers in extraction industry (rubles).
wage_other — Average annual wage income of workers in other industries (rubles).
wage_region — Average annual wage income of workers in region (rubles).

Wage Distribution (By Industry and Gender)

File: **wage_distribution.dta**. The data covers 1914 and 1916. One kopeck = 0.01 ruble.

Variables

year — Year.
industry — Industry.
worker_type — Worker type (by gender and age).
wage_under_50 — Number of workers with a daily wage below 50 kopecks.
wage_51_75 — Number of workers with a daily wage between 51 and 75 kopecks.
wage_76_100 — Number of workers with a daily wage between 76 and 100 kopecks.
wage_101_150 — Number of workers with a daily wage between 101 and 150 kopecks.
wage_151_200 — Number of workers with a daily wage between 151 and 200 kopecks.
wage_201_300 — Number of workers with a daily wage between 201 and 300 kopecks.
wage_301_400 — Number of workers with a daily wage between 301 and 400 kopecks.
wage_above_401 — Number of workers with a daily wage above 401 kopecks.
total — Total number of workers.

Employment Distribution (By Region, Industry, and Gender)

File: **employment.dta**. The data covers 1913-1915.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
production — Type of production within industry.
industry — Industry.
n_factories — Number of factories.
child_workers — Number of child workers (age below 14).
teenage_workers — Number of teenage workers (age between 14 and 17).
male_adult_workers — Number of male adult workers (age 18 and above).
female_adult_workers — Number of female adult workers (age 18 and above).
total_workers — Total number of workers.
male_workers — Total number of male workers.

Identities:

child_workers + teenage_workers + male_adult_workers + female_adult_workers = total_workers

Notes: The total number of female workers can be obtained by subtracting male_workers from total_workers.

Working Hours Distribution (By Industry and Gender)

File: **hours_distribution.dta**. The data covers 1913.

Variables

year — Year.
industry — Industry.
work_conditions — Working conditions (for mining).
control — Control agency (Russian Factory Inspection and Russian Mining Inspection).
worker_type — Worker type (by gender and age).
work_shifts — Number of work shifts (1, 2, or 3).
n_factories — Number of factories.
hours_below_8 — Number of workers working less than 8 hours per day.
hours_8 — Number of workers working 8 hours per day.
hours_8_9 — Number of workers working between 8 and 9 hours per day.
hours_9 — Number of workers working 9 hours per day.
hours_9_10 — Number of workers working between 9 and 10 hours per day.
hours_10 — Number of workers working 10 hours per day.
hours_10_11 — Number of workers working between 10 and 11 hours per day.
hours_11 — Number of workers working 11 hours per day.
hours_above_11 — Number of workers working more than 11 hours per day.
total — Total number of workers.
hours_above_8 — Number of workers working more than 8 hours per day (this category is for work_shifts = 3).

Notes: work_shifts = 999 stands for totals over the worker type by industry.

Agriculture

File: **agriculture.dta**. The data covers 1899-1915. One pood = 16.38 kilograms or 36.11 pounds. One desyatina = 10926.512 square meters.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
year_avg — Interval of years (to obtain the data for 1899-1903).
winter_sown_area — Sown area of winter cereals (desyatinas).
winter_total_harvest — Total harvest of winter cereals (thousand poods).
winter_net_harvest — Net harvest (total harvest minus seeding) of winter cereals (thousand poods).
winter_net_harvest_pc — Net harvest (total harvest minus seeding) of winter cereals per capita (poods).
spring_sown_area — Sown area of spring cereals (desyatinas).
spring_total_harvest — Total harvest of spring cereals (thousand poods).
spring_net_harvest — Net harvest (total harvest minus seeding) of spring cereals (thousand poods).
spring_net_harvest_pc — Net harvest (total harvest minus seeding) of spring cereals per capita (poods).
total_sown_area — Sown area of winter and spring cereals (desyatinas).
total_total_harvest — Total harvest of winter and spring cereals (thousand poods).
total_net_harvest — Net harvest (total harvest minus seeding) of winter and spring cereals (thousand poods).
total_net_harvest_pc — Net harvest (total harvest minus seeding) of winter and spring cereals per capita (poods).
winter_rye_total_harvest — Total harvest of winter rye (thousand poods).
winter_wheat_total_harvest — Total harvest of winter wheat (thousand poods).
spring_rye_total_harvest — Total harvest of spring rye (thousand poods).
spring_wheat_total_harvest — Total harvest of spring wheat (thousand poods).
farro_total_harvest — Total harvest of farro (thousand poods).
hordeum_total_harvest — Total harvest of hordeum (thousand poods).
buckwheat_total_harvest — Total harvest of buckwheat (thousand poods).
proso_millet_total_harvest — Total harvest of proso millet (thousand poods).
corn_total_harvest — Total harvest of corn (thousand poods)
peas_total_harvest — Total harvest of peas (thousand poods).
lentils_total_harvest — Total harvest of lentils (thousand poods).
beans_total_harvest — Total harvest of beans (thousand poods).
oat_total_harvest — Total harvest of oat (thousand poods).
potatoes_total_harvest — Total harvest of potatoes (thousand poods).
winter_hordeum_total_harvest — Total harvest of winter hordeum (thousand poods)
spring_hordeum_total_harvest — Total harvest of spring hordeum (thousand poods)

`rice_total_harvest` — Total harvest of rice (thousand poods).
`cotton_total_harvest` — Total harvest of cotton (thousand poods).
`sorghum_total_harvest` — Total harvest of sorghum (thousand poods).
`winter_wheat_sown_area` — Sown area of winter wheat (desyatinas).
`spring_wheat_sown_area` — Sown area of spring wheat (desyatinas).
`winter_hordeum_sown_area` — Sown area of winter hordeum (desyatinas).
`spring_hordeum_sown_area` — Sown area of spring hordeum (desyatinas).
`corn_sown_area` — Sown area of corn (desyatinas).
`oat_sown_area` — Sown area of oat (desyatinas).
`buckwheat_sown_area` — Sown area of buckwheat (desyatinas).
`potatoes_sown_area` — Sown area of potatoes (desyatinas).
`proso_millet_sown_area` — Sown area of proso millet (desyatinas).
`rice_sown_area` — Sown area of rice (desyatinas).
`cotton_sown_area` — Sown area of cotton (desyatinas).
`sorghum_sown_area` — Sown area of sorghum (desyatinas).
`winter_rye_sown_area` — Sown area of winter rye (desyatinas).
`spring_rye_sown_area` — Sown area of spring rye (desyatinas).
`farro_sown_area` — Sown area of farro (desyatinas).
`peas_sown_area` — Sown area of peas (desyatinas).
`lentils_sown_area` — Sown area of lentils (desyatinas).
`beans_sown_area` — Sown area of beans (desyatinas).
`potatoes_net_harvest` — Net harvest (total harvest minus seeding) of potatoes (thousand poods).
`oat_net_harvest` — Net harvest (total harvest minus seeding) of oat (thousand poods).
`potatoes_net_harvest_pc` — Net harvest (total harvest minus seeding) of potatoes per capita (poods).

Russian Statistical Yearbooks provide the data for 1904-1915. However, since they also report 5-year averages for harvest, I could construct the data back to 1899. For example, if I observe $x_{1904}, ..., x_{1907}$ and the average over 1903-1907, $\bar{x}_{1903-1907}$, then I obtain x_{1903} using the following expression:

$$x_{1903} = 5\bar{x}_{1903-1907} - \sum_{i=1904}^{1907} x_i$$

I construct the 1899-1903 data for aggregate categories (winter cereals, spring cereals, total cereals). Data for disaggregated cereals can be constructed using variable `year_avg` and the procedure described above.

Meadows, total area (desyatinas): **1904** – 34227889, **1905** – 33345737, **1906** – 33019363.

Average Harvest

File: **harvest.dta**. The data captures the average harvest over the period of 1895-1904. One pood = 16.38 kilograms or 36.11 pounds. One desyatina = 10926.512 square meters.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
winter_rye_hvst — Average winter rye harvest (poods per desyatina).
winter_wheat_hvst — Average winter wheat harvest (poods per desyatina).
spring_rye_hvst — Average spring rye harvest (poods per desyatina).
spring_wheat_hvst — Average spring wheat harvest (poods per desyatina).
farro_hvst — Average farro harvest (poods per desyatina).
hordeum_hvst — Average hordeum harvest (poods per desyatina).
buckwheat_hvst — Average buckwheat harvest (poods per desyatina).
proso_millet_hvst — Average proso millet harvest (poods per desyatina).
corn_hvst — .Average corn harvest (poods per desyatina).
peas_hvst — Average peas harvest (poods per desyatina).
beans_hvst — Average beans harvest (poods per desyatina).
lentils_hvst — Average lentils harvest (poods per desyatina).
potatoes_hvst — Average potatoes harvest (poods per desyatina).
oat_hvst — Average oat harvest (poods per desyatina).

Cereal Prices

File: **agriculture_prices.dta**. The data covers 1889-1909. One kopeck = 0.01 ruble. One pood = 16.38 kilograms or 36.11 pounds.

Variables

`region_id` — Region ID.
`region` — Region.
`macroregion` — Macroregion.
`year` — Year.[4][5]
`market` — Market (North ports, South ports, or internal markets).
`price_oat` — Average oat price (kopeks per pood).
`price_rye` — Average rye price (kopeks per pood).
`price_wheat` — Average wheat price (kopeks per pood).
`price_hordeum` — Average hordeum price (kopeks per pood).

---

[4] *year = 1905* for observations with non-empty *region_id* corresponds to the average over 1901-1905.
[5] *year = 1899* for observations with empty *region_id* corresponds to the average over 1890-1899.

Livestock

File: **livestock.dta**. The data covers 1901-1915.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
horses — Number of horses.
horses_old — Number of horses with the age of 4 years and older.
horses_young — Number of horses with the age of less than 4 years.
cattle — Number of cattle.
sheeps_goats — Number of sheeps and goats.
pigs — Number of pigs.

Identities:

horses_old + horses_young = horses

Land Ownership

File: **land_ownership.dta**. The data covers 1905, 1907, and 1909. One desyatina = 10926.512 square meters.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
private_land — Privately owned land area (desyatinas).
peasant_land — Area of land owned by peasants: nadel or purchased (desyatinas).
state_total_land — Area of land owned by the government, church, etc. (desyatinas).
region_land — Area of land owned by region or city (posad) (desyatinas).
state_land — Area of state-owned land (desyatinas).
total_land — Total area (desyatinas).

Identities:

private_land + peasant_land + state_total_land = total_land

region_land + state_land = state_total_land

Land Prices

File: **land_prices.dta**. The data covers 1905-1914. One desyatina = 10926.512 square meters.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
land_price_1 — Average price of desyatina of land purchased by peasants from private landowners (rubles).
land_price_2 — Average price of desyatina of land purchased by peasants from the Peasants' Land Bank (rubles).
land_price_3 — Average price of desyatina of land purchased by the Peasants' Land Bank (rubles).

Production (Excise Tax)

File: **production_excise.dta**. The data covers 1903 and 1907-1913. One bucket = 12.299 liters. One pood = 16.38 kilograms or 36.11 pounds.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
n_factory_excise — Number of factories working in industries that are subject to excise tax.
production_a_excise — Volume of production of industries that are subject to excise tax (thousand buckets).
production_b_excise — Volume of production of industries that are subject to excise tax (thousand poods).
matches — Production of matches (thousands).
n_workers_excise — Number of workers employed in industries that are subject to excise tax.

Production (No Excise Tax)

File: **production_no_excise.dta**. The data covers 1900 and 1910.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
cotton_factories — Number of cotton processing factories.
cotton_value — Value of cotton processing production (thousand rubles).
cotton_factory_workers — Number of workers in cotton processing factories.
cotton_home_workers — Number of workers working from home for cotton processing factories.
wool_factories — Number of wool processing factories.
wool_value — Value of wool processing production (thousand rubles).
wool_factory_workers — Number of workers in wool processing factories.
wool_home_workers — Number of workers working from home for wool processing factories.
silk_factories — Number of silk processing factories.
silk_value — Value of silk processing production (thousand rubles).
silk_factory_workers — Number of workers in silk processing factories.
silk_home_workers — Number of workers working from home for silk processing factories.
linum_factories — Number of linum, hemp, and jute processing factories.
linum_value — Value of linum, hemp, and jute processing production (thousand rubles).
linum_factory_workers — Number of workers in linum, hemp, and jute processing factories.
linum_home_workers — Number of workers working from home for linum, hemp, and jute processing factories.
mixed_textile_factories — Number of mixed textile processing factories.
mixed_textile_value — Value of mixed textile processing production (thousand rubles).
mixed_textile_factory_workers — Number of workers in mixed textile processing factories.
mixed_textile_home_workers — Number of workers working from home for mixed textile processing factories.
paper_factories — Number of paper production factories.
paper_value — Value of paper production (thousand rubles).
paper_factory_workers — Number of workers in paper production factories.
paper_home_workers — Number of workers working from home for paper production factories.
wood_factories — Number of wood processing factories.
wood_value — Value of wood processing production (thousand rubles).
wood_factory_workers — Number of workers in wood processing factories.
wood_home_workers — Number of workers working from home for wood processing factories.
metal_factories — Number of metal processing and production factories.
metal_value — Value of metal processing and production (thousand rubles).
metal_factory_workers — Number of workers in metal processing and production factories.

`metal_home_workers` — Number of workers working from home for metal processing and production factories.

`mineral_factories` — Number of mineral processing factories.

`mineral_value` — Value of mineral processing production (thousand rubles).

`mineral_factory_workers` — Number of workers in mineral processing factories.

`mineral_home_workers` — Number of workers working from home for mineral processing factories.

`animal_factories` — Number of animal products processing factories.

`animal_value` — Value of animal products processing (thousand rubles).

`animal_factory_workers` — Number of workers in animal products processing factories.

`animal_home_workers` — Number of workers working from home for animal products processing factories.

`food_factories` — Number of food processing factories.

`food_value` — Value of food processing production (thousand rubles).

`food_factory_workers` — Number of workers in food processing factories.

`food_home_workers` — Number of workers working from home for food processing factories.

`chemical_factories` — Number of chemical production factories.

`chemical_value` — Value of chemical production (thousand rubles).

`chemical_factory_workers` — Number of workers working from home for chemical production factories.

`chemical_home_workers` — Number of workers working from home for chemical production factories.

`total_factories` — Number of factories in industries that are not subject to excise tax.

`total_value` — Value of output in industries that are not subject to excise tax (thousand rubles).

`total_factory_workers` — Number of workers working in industries that are not subject to excise tax.

`total_home_workers` — Number of workers working from home for in industries that are not subject to excise tax.

`n_factories_value` — Number of factories reporting the value of production (variable `total_value`).

`n_factories_workers` — Number of factories reporting the number of workers (variable `total_factory_workers`).

`engines` — Number of engines.

`horsepowers` — Number of (engine) horsepowers.

Industry Characteristics (No Excise Tax)

File: **ind_characteristics_no_excise.dta**. The data covers 1900 and 1910.

Variables

industry — Industry (not subject to excise tax).
year — Year.
n_factories — Number of factories.
output_value — Value of output (thousand rubles).
n_workers — Number of workers.
fact_workers — Number of workers in factories.
home_workers — Number of workers working from home.
equipment_value — Value of equipment (thousand rubles).
n_engines — Number of engines.
n_horsepowers — Number of (engine) horsepowers.
exp_materials — Expenses on materials (thousand rubles).
exp_fuel — Expenses on fuel (thousand rubles).
exp_wages — Wage expenses (thousand rubles).
exp_administration — Administrative expenses (thousand rubles).
exp_depreciation — Depreciation expenses (thousand rubles).
exp_insurance_taxes — Expenses on property insurance and taxes (thousand rubles).
exp_worker_benefits — Expenses on worker benefits (thousand rubles).
n_fact_value — Number of factories reporting the value of output
(variable output_value).
n_fact_workers — Number of factories reporting the number of workers
(variables fact_workers and home_workers).
n_fact_engines — Number of factories reporting the number of engines
(variable n_engines).
n_fact_horsepowers — Number of factories reporting the number of (engine) horsepowers
(variable n_horsepowers).

Identities:

n_workers = fact_workers + home_workers

Trade

File: **trade_countries.dta**. The data covers 1897-1914. One pood = 16.38 kilograms or 36.11 pounds.

Variables

country — Country (exporters and importers).
year — Year.
export — Export (rubles).
import — Import (rubles).
rye_export — Export of rye (poods).
wheat_export — Export of wheat (poods).
hordeum_export — Export of hordeum (poods).
oat_export — Export of oat (poods).
corn_export — Export of corn (poods).
buckwheat_export — Export of buckwheat (poods).
pea_export — Export of pea (poods).
beans_lentils_export — Export of beans and lentils (poods).
buckwheat_groats_export — Export of buckwheat groats (poods).
wheat_flour_export — Export of wheat flour (poods).
rye_flour_export — Export of rye flour (poods).
bran_export — Export of bran (poods).
horses_export — Export of horses (number).
cattle_export — Export of cattle (number).
sheep_export — Export of sheep (number).
pigs_export — Export of pigs (number).
geese_export — Export of geese (number).
poultry_export — Export of poultry (number).
pork_export — Export of pork (poods).
butter_export — Export of butter (poods).
eggs_export — Export of eggs (thousands).
sheep_goat_skins_export — Export of sheep and goat skins (number).
skins_export — Export of skins (number).
wool_export — Export of wool (poods).
horse_tails_export — Export of horse tails (poods).
bristle_export — Export of bristle (poods).
feather_export — Export of feather (poods).
tripe_export — Export of tripe (poods).
horses_import — Import of horses (number).
cattle_import — Import of cattle (number).
sheep_import — Import of sheep (number).
pigs_import — Import of pigs (number).

Notes: Be careful with "Other countries" (variable `country`) because it is a residual category, and may not coincide across the datasets.

Data over 1903-1914 for variables `export` and `import` is published in thousand rubles, hence be aware of rounding.

Total value of export of horses (in rubles): **1911** – 11438649, **1912** – 12197254, **1913** – 13884801, **1914** – 8633030.

Total value of export of cattle (in rubles): **1911** – 258910, **1912** – 182299, **1913** – 688234, **1914** – 229482.

Total value of export of sheep (in rubles): **1911** – 60989, **1912** – 135906, **1913** – 721265, **1914** – 966896.

Total value of export of pigs (in rubles): **1911** – 6294513, **1912** – 8465558, **1913** – 8677644, **1914** – 2374661.

Total value of import of horses (in rubles): **1911** – 1638293, **1912** – 1499718, **1913** – 2590178, **1914** – 2103046.

Total value of import of cattle (in rubles): **1911** – 5725347, **1912** – 5568018, **1913** – 7898513, **1914** – 6986633.

Total value of import of sheep (in rubles): **1911** – 3000516, **1912** – 3973854, **1913** – 5922898, **1914** – 3418734.

Total value of import of pigs (in rubles): **1911** – 431574, **1912** – 604076, **1913** – 705683, **1914** – 580392.

Trade (By Broad Border)

File: **trade_broad_borders.dta**. The data covers 1899-1907.

Variables

broad_border — Broad border (European, Asian, or Finland).
year — Year.
export_land — Export, land border (thousand rubles).
export_sea — Export, sea border (thousand rubles).
export_total — Export, total (thousand rubles).
import_land — Import, land border (thousand rubles).
import_sea — Import, sea border (thousand rubles).
import_total — Import, total (thousand rubles).

Trade (By Border)

File: **trade_borders.dta**. The data covers 1897-1914.

Variables

border — Border.
broad_border — Broad border (European, Asian, or Finland).
year — Year.
im_borders — Import (rubles).
ex_borders — Export (rubles).

Notes: Data over 1903-1914 is published in thousand rubles, hence be aware of rounding.

Trade (By Custom and Product Category)

File: **trade_customs.dta**. The data covers 1902-1908. One pood = 16.38 kilograms or 36.11 pounds.

Variables

customs_id — Customs ID.
customs — Customs.
year — Year.
ex_horses_cust — Export of horses (number).
ex_cattle_cust — Export of cattle (number).
ex_sheep_cust — Export of sheep (number).
ex_pigs_cust — Export of pigs (number).
ex_geese_cust — Export of geese (number).
ex_poultry_cust — Export of poultry (number).
ex_pork_cust — Export of pork (poods).
ex_butter_cust — Export of butter (poods).
ex_eggs_cust — Export of eggs (thousands).
ex_horse_tails_cust — Export of horse tails (poods).
ex_sheep_goat_skins_cust — Export of sheep and goat skins (poods).
ex_leather_cust — Export of leather (poods).
ex_wool_cust — Export of wool (poods).
ex_bristle_cust — Export of bristle (poods).
ex_feather_cust — Export of feather (poods).
ex_tripe_cust — Export of tripe (poods).
ex_horses_cust_value — Export value of horses (rubles).
ex_cattle_cust_value — Export value of cattle (rubles).
ex_sheep_cust_value — Export value of sheep (rubles).
ex_pigs_cust_value — Export value of pigs (rubles).
im_horses_cust — Import of horses (number).
im_cattle_cust — Import of cattle (number).
im_sheep_cust — Import of sheep (number).
im_pigs_cust — Import of pigs (number).
im_horses_cust_value — Import value of horses (rubles).
im_cattle_cust_value — Import value of cattle (rubles).
im_sheep_cust_value — Import value of sheep (rubles).
im_pigs_cust_value — Import value of pigs (rubles).

Identities:

Variable customs: *Total = Main port customs + Land customs + Other customs*

Notes: Customs-level data for the import of pigs in 1908 is not reliable because of the bad quality of original documents. However, the numbers for total quantity and value are correct.

Export and Import (By Custom)

File: **customs_ex_im.dta**. The data covers 1897-1914.

Variables

customs_id — Customs ID.
customs — Customs.
year — Year.
export_cust — Export (rubles).
import_cust — Import (rubles).

Notes: Data over 1903-1914 (some customs—starting from 1901) is published in thousand rubles, hence be aware of rounding.

Be careful with "Other customs" (variable customs) because it is a residual category, and may not coincide across the datasets.

Export (By Agricultural Product)

File: **export_value.dta**. The data covers 1892-1914. One pood = 16.38 kilograms or 36.11 pounds.

Variables

year — Year.
ex_quant — Export (thousand poods).
ex_value — Export (thousand rubles).
wheat_ex_quant — Export of wheat (thousand poods).
wheat_ex_value — Export of wheat (thousand rubles).
rye_ex_quant — Export of rye (thousand poods).
rye_ex_value — Export of rye (thousand rubles).
hordeum_ex_quant — Export of hordeum (thousand poods).
hordeum_ex_value — Export of hordeum (thousand rubles).
oat_ex_quant — Export of oat (thousand poods).
oat_ex_value — Export of oat (thousand rubles).
corn_ex_quant — Export of corn (thousand poods).
corn_ex_value — Export of corn (thousand rubles).
peas_ex_quant — Export of peas (thousand poods).
peas_ex_value — Export of peas (thousand rubles).
wheat_flour_ex_quant — Export of wheat flour (thousand poods).
wheat_flour_ex_value — Export of wheat flour (thousand rubles).
rye_flour_ex_quant — Export of rye flour (thousand poods).
rye_flour_ex_value — Export of rye flour (thousand rubles).
barn_ex_quant — Export of barn (thousand poods).
barn_ex_value — Export of barn (thousand rubles).

Export (By Custom and Agricultural Product)

File: **export_customs.dta**. The data covers 1902-1914. One pood = 16.38 kilograms or 36.11 pounds.

Variables

customs_id — Customs ID.
customs — Customs.
year — Year.
wheat_ex_cust — Export of wheat (poods).
rye_ex_cust — Export of rye (poods).
hordeum_ex_cust — Export of hordeum (poods).
oat_ex_cust — Export of oat (poods).
corn_ex_cust — Export of corn (poods).
buckwheat_ex_cust — Export of buckwheat (poods).
peas_ex_cust — Export of peas (poods).
beans_lentils_ex_cust — Export of beans and lentils (poods).
buckwheat_groat_ex_cust — Export of buckwheat groat (poods).
wheat_flour_ex_cust — Export of wheat flour (poods).
rye_flour_ex_cust — Export of rye flour (poods).
bran_ex_cust — Export of bran (poods).

Identities:

Variable customs: *Total = Main port customs + Land customs + Other customs*

Government Expenditures (Plan)

File: **budget_expenditures.dta**. The data covers 1865-1917.

Variables

item — Item.
year — Year.
regular_exp — Regular expenditures (rubles).
extraordinary_exp — Extraordinary expenditures (rubles).
total_exp — Total expenditures (rubles).
total_exp_comparable — Total expenditures (rubles): Same structure as in the next year.

Identities:

regular_exp + extraordinary_exp = total_exp

Variable total_exp: *Subtotal Expenditures (Agencies and Government Debt) + Extraordinary Expenditures + Spending to Compensate Shortfall in Budget Revenue + Current Expenditures = Total Expenditures*

Notes: Item total_exp = *"Extraordinary Expenditures"* includes expenditures on railways, rearming, food reserves, etc.

Variable total_exp_comparable for year t allows to compare the expenditures with year $t + 1$ as it takes into account the changes in budget structure.

Government Revenue (Plan)

File: **budget_revenue.dta**. The data covers 1865-1917.

Variables

year — Year.
direct_taxes_gross — Direct taxes, gross (rubles).
indirect_taxes_gross — Indirect taxes and duties, gross (rubles).
total_taxes_gross — Tax revenue, gross (rubles).
total_rev_gross — Total government revenue, gross (rubles).
direct_taxes_coll — Direct taxes, cost of collection (rubles).
indirect_taxes_coll — Indirect taxes and duties, cost of collection (rubles).
total_taxes_coll — Tax revenue, cost of collection (rubles).
total_rev_coll — Total government revenue, cost of collection (rubles).
direct_taxes_net — Direct taxes, net (rubles).
indirect_taxes_net — Indirect taxes and duties, net (rubles).
total_taxes_net — Tax revenue, net (rubles).
total_rev_net — Total government revenue, net (rubles).
direct_taxes_comp — Direct taxes, gross (rubles): Same structure as in the next year.
indirect_taxes_comp — Indirect taxes and duties, gross (rubles): Same structure as in the next year.
total_taxes_comp — Tax revenue, gross (rubles): Same structure as in the next year.
total_rev_comp — Total government revenue, gross (rubles): Same structure as in the next year.

Identities:

direct_taxes + indirect_taxes = total_taxes

direct_taxes_net + direct_taxes_coll = direct_taxes_gross

Notes: Variables ***_comp for year t allow to compare the revenue (and sources) with year $t + 1$ as it takes into account the changes in budget structure.

Government Budget (Realized)

File: **government_budget.dta**. The data covers 1832-1914.

Variables

year — Year.
govt_revenue — Government revenue (thousand rubles).
regular_revenue — Regular government revenue (thousand rubles).
extraordinary_revenue — Extraordinary government revenue (thousand rubles).
govt_expenditures — Government expenditures (thousand rubles).
regular_expenditures — Regular government expenditures (thousand rubles).
extraordinary_expenditures — Extraordinary government expenses (thousand rubles).
govt_loans — Government loans (thousand rubles).
state_credit_funds — State credit funds (thousand rubles).
govt_debt_funds — Funds to cover government debt (thousand rubles).
money_emission — Money emission (thousand rubles).
treasury_bonds — Treasury bonds emission (thousand rubles).

Identities:

regular_revenue + extraordinary_revenue = govt_revenue

regular_expenditures + extraordinary_expenditures = govt_expenditures

govt_loans + state_credit_funds + govt_debt_funds + money_emission +
+ treasury_bonds = govt_expenditures - govt_revenue

Household Deposits

File: **deposits.dta**. The data covers 1899-1914.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
n_funds — Number of sberkassas (state saving banks)
n_accounts — Number of accounts.
n_accounts_per_1000 — Accounts per 1000 residents.
total_deposits — Total amount of deposits (thousand rubles).

Strikes: Results

File: **strikes_results.dta**. The monthly data covers 1916.

Variables

year — Year.
month — Month (numeric).
month_name — Month.
n_strikes — Total number of strikes.
n_participants — Total number of strike participants.
lost_days — Total number of days lost due to strikes.
n_strikes_econ — Number of strikes by economic reasons.
n_participants_econ — Number of participants in strikes by economic reasons.
n_strikes_wage — Number of strikes by economic reasons – wage.
n_participants_wage — Number of participants in strikes by economic reasons – wage.
n_strikes_time — Number of strikes by economic reasons – working time.
n_participants_time — Number of participants in strikes by economic reasons – working time.
n_strikes_cond — Number of strikes by economic reasons – working conditions.
n_participants_cond — Number of participants in strikes by economic reasons – working conditions.
lost_days_econ — Number of days lost due to strikes by economic reasons.
n_strikes_full — Number of strikes where workers' claims were fully satisfied.
n_participants_full — Number of participants in strikes where workers' claims were fully satisfied.
n_strikes_part — Number of strikes where workers' claims were partly satisfied.
n_participants_part — Number of participants in strikes where workers' claims were partly satisfied.
n_strikes_unst — Number of strikes where workers' claims were rejected.
n_participants_unst — Number of participants in strikes where workers' claims were rejected.
n_strikes_oth — Number of strikes by political and other reasons.
n_participants_oth — Number of participants in strikes by political and other reasons.
lost_days_oth — Number of days lost due to strikes by political and other reasons.

Identities:

n_strikes_econ + n_strikes_oth = n_strikes

n_strikes_wage + n_strikes_time + n_strikes_cond = n_strikes_econ

Same relationships for the number of strike participants.

Strikes: Results (By Industry and Reason)

File: **strikes_reasons_results.dta**. The data covers 1913-1915.

Variables

year — Year.
reason — Strike reason.
industry — Industry.
n_str_full — Number of strikes where workers' claims were fully satisfied.
n_part_full — Number of participants in strikes where workers' claims were fully satisfied.
n_str_part — Number of strikes where workers' claims were partly satisfied.
n_part_part — Number of participants in strikes where workers' claims were partly satisfied.
n_str_unst — Number of strikes where workers' claims were rejected.
n_part_unst — Number of participants in strikes where workers' claims were rejected.
n_str_nores — Number of strikes with no results.
n_part_nores — Number of participants in strikes with no results.
n_str_group1 — Number of strikes lasting less than 0.5 days.
n_part_group1 — Number of participants in strikes lasting less than 0.5 days.
n_str_group2 — Number of strikes lasting 0.5-2 days.
n_part_group2 — Number of participants in strikes lasting 0.5-2 days.
n_str_group3 — Number of strikes lasting 2-3 days.
n_part_group3 — Number of participants in strikes lasting 2-3 days.
n_str_group4 — Number of strikes lasting 3-5 days.
n_part_group4 — Number of participants in strikes lasting 3-5 days.
n_str_group5 — Number of strikes lasting 5-10 days.
n_part_group5 — Number of participants in strikes lasting 5-10 days.
n_str_group6 — Number of strikes lasting 10-15 days.
n_part_group6 — Number of participants in strikes lasting 10-15 days.
n_str_group7 — Number of strikes lasting 15-20 days.
n_part_group7 — Number of participants in strikes lasting 15-20 days.
n_str_group8 — Number of strikes lasting 20-30 days.
n_part_group8 — Number of participants in strikes lasting 20-30 days.
n_str_group9 — Number of strikes lasting more than 30 days.
n_part_group9 — Number of participants in strikes lasting more than 30 days.
n_str_total — Total number of strikes.
n_part_total — Total number of strike participants.

Strikes (By Region)

File: **strikes_region.dta**. The monthly data covers 1916.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
month — Month (numeric).
month_name — Month.
n_strikes_reg — Number of strikes.
n_participants_reg — Number of strike participants.

Strikes (By Industry)

File: **strikes_industry.dta**. The monthly data covers 1913-1916.

Variables

year — Year.
month — Month (numeric).
month_name — Month.
reason — Strike reason.
n_strikes_cotton — Number of strikes (cotton processing industry).
n_participants_cotton — Number of strike participants (cotton processing industry).
n_strikes_wool — Number of strikes (wool processing industry).
n_participants_wool — Number of strike participants (wool processing industry).
n_strikes_silk — Number of strikes (silk processing industry).
n_participants_silk — Number of strike participants (silk processing industry).
n_strikes_linum — Number of strikes (linum, hemp, and jute processing industry).
n_participants_linum — Number of strike participants (linum, hemp, and jute processing industry).
n_strikes_mixed — Number of strikes (mixed textile processing industry).
n_participants_mixed — Number of strike participants (mixed textile processing industry).
n_strikes_paper — Number of strikes (paper processing industry).
n_participants_paper — Number of strike participants (paper processing industry).
n_strikes_wood — Number of strikes (wood processing industry).
n_participants_wood — Number of strike participants (wood processing industry).
n_strikes_metal — Number of strikes (metal processing industry).
n_participants_metal — Number of strike participants (metal processing industry).
n_strikes_mineral — Number of strikes (mineral processing industry).
n_participants_mineral — Number of strike participants (mineral processing industry).
n_strikes_animal — Number of strikes (animal products processing industry).
n_participants_animal — Number of strike participants (animal products processing industry).
n_strikes_food — Number of strikes (food processing industry).
n_participants_food — Number of strike participants (food processing industry).
n_strikes_chemical — Number of strikes (chemical industry).
n_participants_chemical — Number of strike participants (chemical industry).
n_strikes_extraction — Number of strikes (extraction industry).
n_participants_extraction — Number of strike participants (extraction industry).
n_strikes_other — Number of strikes (other industries).
n_participants_other — Number of strike participants (other industries).
n_strikes_total — Total number of strikes.
n_participants_total — Total number of strike participants.

Strikes (By Region and Industry)

File: **strikes_region_industry.dta**. The data covers 1913-1915.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
n_strikes_reg_cotton — Number of strikes (cotton processing industry).
n_part_reg_cotton — Number of strike participants (cotton processing industry).
n_strikes_reg_wool — Number of strikes (wool processing industry).
n_part_reg_wool — Number of strike participants (wool processing industry).
n_strikes_reg_silk — Number of strikes (silk processing industry).
n_part_reg_silk — Number of strike participants (silk processing industry).
n_strikes_reg_linum — Number of strikes (linum, hemp, and jute processing industry).
n_part_reg_linum — Number of strike participants (linum, hemp, and jute processing industry).
n_strikes_reg_mixed — Number of strikes (mixed textile processing industry).
n_part_reg_mixed — Number of strike participants (mixed textile processing industry).
n_strikes_reg_paper — Number of strikes (paper processing industry).
n_part_reg_paper — Number of strike participants (paper processing industry).
n_strikes_reg_wood — Number of strikes (wood processing industry).
n_part_reg_wood — Number of strike participants (wood processing industry).
n_strikes_reg_metal — Number of strikes (metal processing industry).
n_part_reg_metal — Number of strike participants (metal processing industry).
n_strikes_reg_mineral — Number of strikes (mineral processing industry).
n_part_reg_mineral — Number of strike participants (mineral processing industry).
n_strikes_reg_animal — Number of strikes (animal products processing industry).
n_part_reg_animal — Number of strike participants (animal products processing industry).
n_strikes_reg_food — Number of strikes (food processing industry).
n_part_reg_food — Number of strike participants (food processing industry).
n_strikes_reg_chemical — Number of strikes (chemical industry).
n_part_reg_chemical — Number of strike participants (chemical industry).
n_strikes_reg_extraction — Number of strikes (extraction industry).
n_part_reg_extraction — Number of strike participants (extraction industry).
n_strikes_reg_other — Number of strikes (other industries).
n_part_reg_other — Number of strike participants (other industries).
n_strikes_reg_total — Total number of strikes.
n_part_reg_total — Total number of strike participants.

Diseases

File: **diseases.dta**. The data covers 1903-1914.

Variables

region_id — Region ID.
region — Region.
macroregion — Macroregion.
year — Year.
smallpox — Smallpox (number of cases).
scarletfever — Scarlet Fever (number of cases).
diphtheria — Diphtheria (number of cases).
measles — Measles (number of cases).
whoopingcough — Whooping Cough (number of cases).
flu — Flu (number of cases).
typhus — Typhus (number of cases).
typhoidfever — Typhoid Fever (number of cases).
relapsingfever — Relapsing Fever (number of cases).
unspecifiedtyphus — Unspecified Typhus (number of cases).
dysentery — Dysentery (number of cases).
choleranostras — Cholera Nostras (number of cases).
choleraasiatica — Cholera Asiatica (number of cases).
gastrointestinitis — Gastrointestinitis (number of cases).
mumps — Mumps (number of cases).
erysipelas — Erysipelas (number of cases).
sepsisandpyaemia — Sepsis and Pyaemia (number of cases).
syphilis — Syphilis (number of cases).
pr_syphilis — Primary Syphilis (number of cases).
sec_syphilis — Secondary Syphilis (number of cases).
tertiarysyphilis — Tertiary Syphilis (number of cases).
gonorrhea — Gonorrhea (number of cases).
chancroid — Chancroid (number of cases).
croupouspneumonia — Croupous Pneumonia (number of cases).
tuberculosis — Tuberculosis (number of cases).
malaria — Malaria (number of cases).
scabies — Scabies (number of cases).
scurvy — Scurvy (number of cases).
rheumaticfever — Rheumatic Fever (number of cases).
trachoma — Trachoma (number of cases).
plague — Plague (number of cases).
std — Sexually Transmitted Diseases (number of cases).
traumas — Traumas (number of cases).
mentaldisorder — Mental Disorder (number of cases).
smallpoxvaccines — Number of Smallpox Vaccines (number of cases).

Other (Not Digitized) Data

Meteorological data:

- Normal temperature, average temperature, maximum and minimum temperature. Monthly data is available for the period of 1913-1917 for meteorological stations (with geographical coordinates).

- Normal precipitation, actual precipitation. Monthly data is available for the period of 1913-1917 for meteorological stations (with geographical coordinates).

For further details about non-digitized data (migration, banking statistics, livestock diseases, etc.), please, contact me by malko017@umn.edu.

Sources

Annual Budgets of the Russian Empire, 1866-1917.

Central Statistical Committee Review, 1863-1912.

Russian Statistical Yearbooks, 1904-1916.

Statistical Report, 1913-1917.