Uzbekistan Health Reforms: Shymkent/Tashkent Comparison
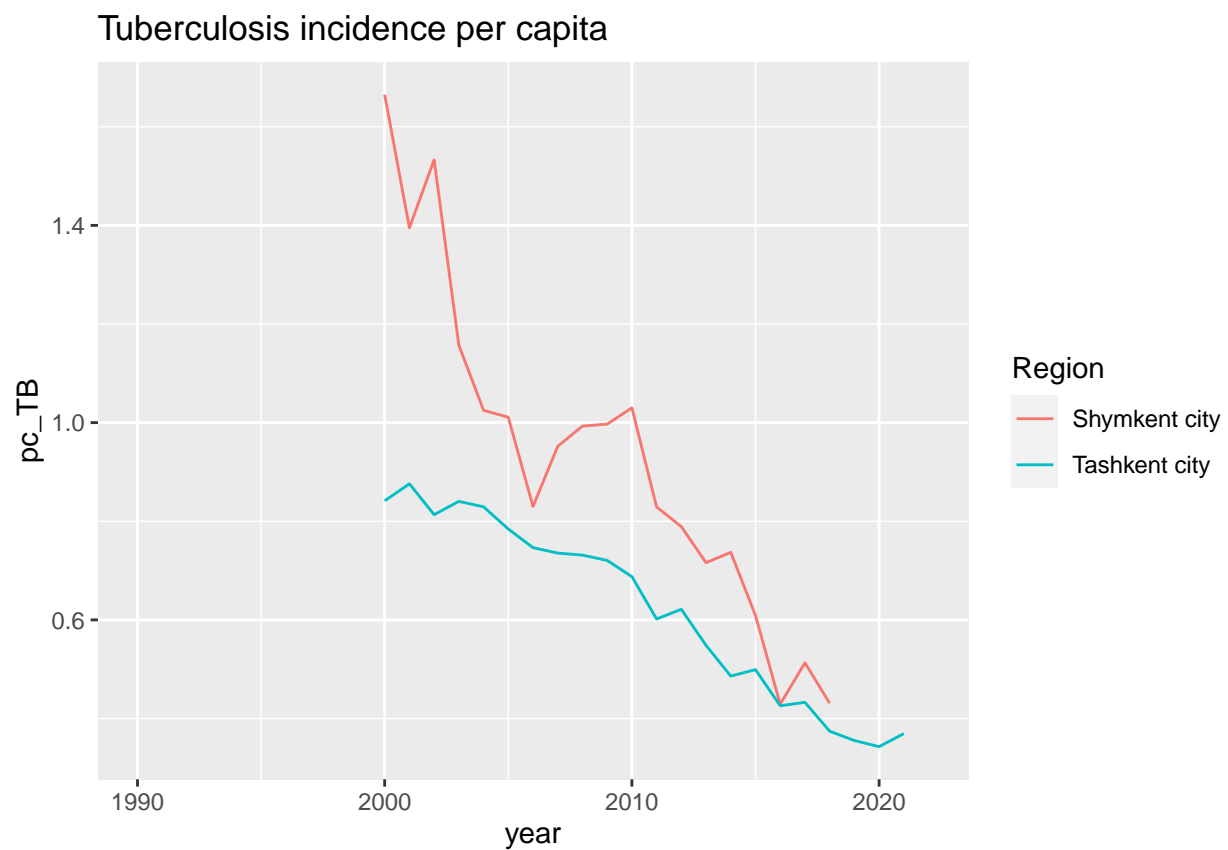
Michelle Schultze

2023-02-23

DISEASES

TB

```
ggplot(comparisons, aes(x = year, y = pc_TB)) +
  geom_line(aes(color = Region)) +
  labs(title = "Tuberculosis incidence per capita")
```

```
## Warning: Removed 11 rows containing missing values (`geom_line()`).
```
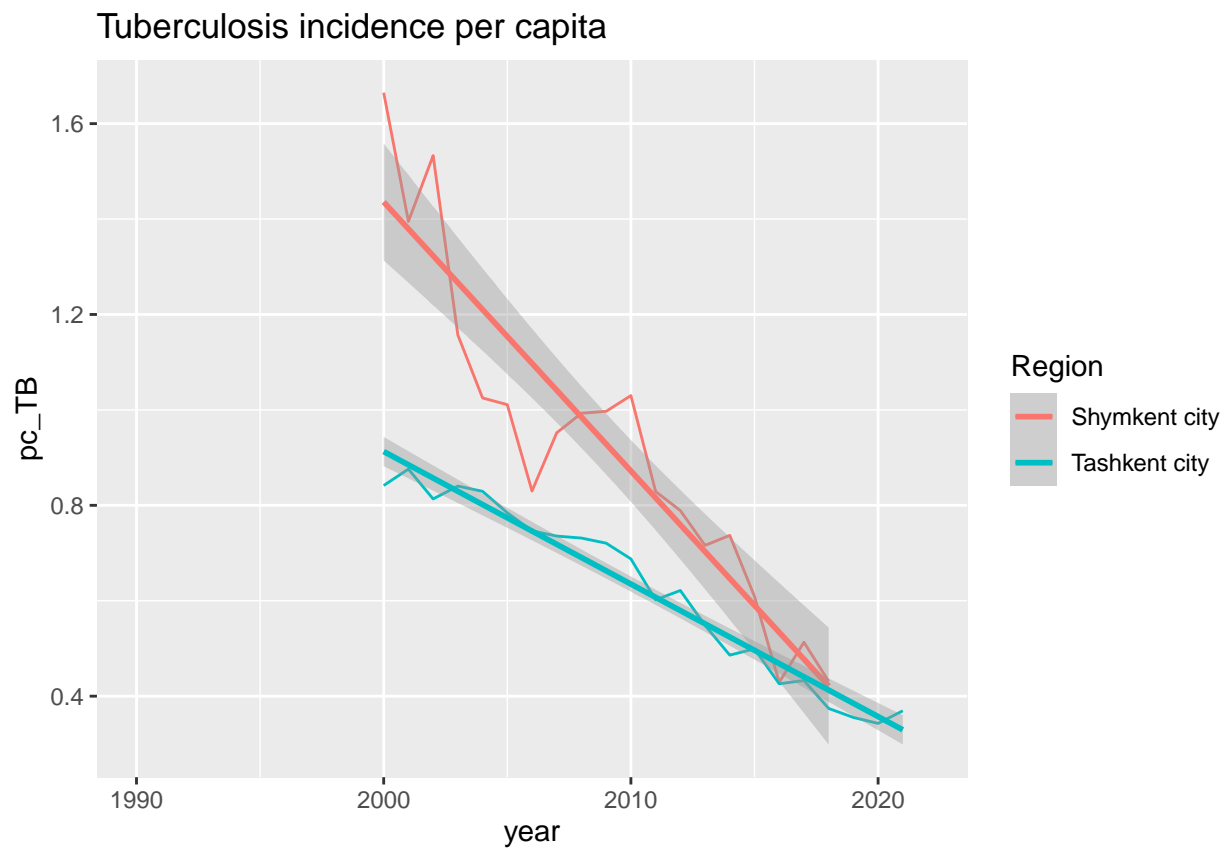
```
ggplot(comparisons, aes(x = year, y = pc_TB)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Tuberculosis incidence per capita")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 11 rows containing non-finite values (`stat_smooth()`).
## Removed 11 rows containing missing values (`geom_line()`).
```

Tuberculosis incidence per capita


```
TBmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_TB ~ year + Region, data = comparisons)
tidy(TBmodel)
```

```
## # A tibble: 3 x 5
##   term                 estimate std.error statistic  p.value
##   <chr>                   <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)             79.2       6.64      11.9  2.02e-14
## 2 year                 -0.0390    0.00330     -11.8  2.88e-14
## 3 RegionTashkent city  -0.249     0.0398      -6.26  2.54e- 7
```
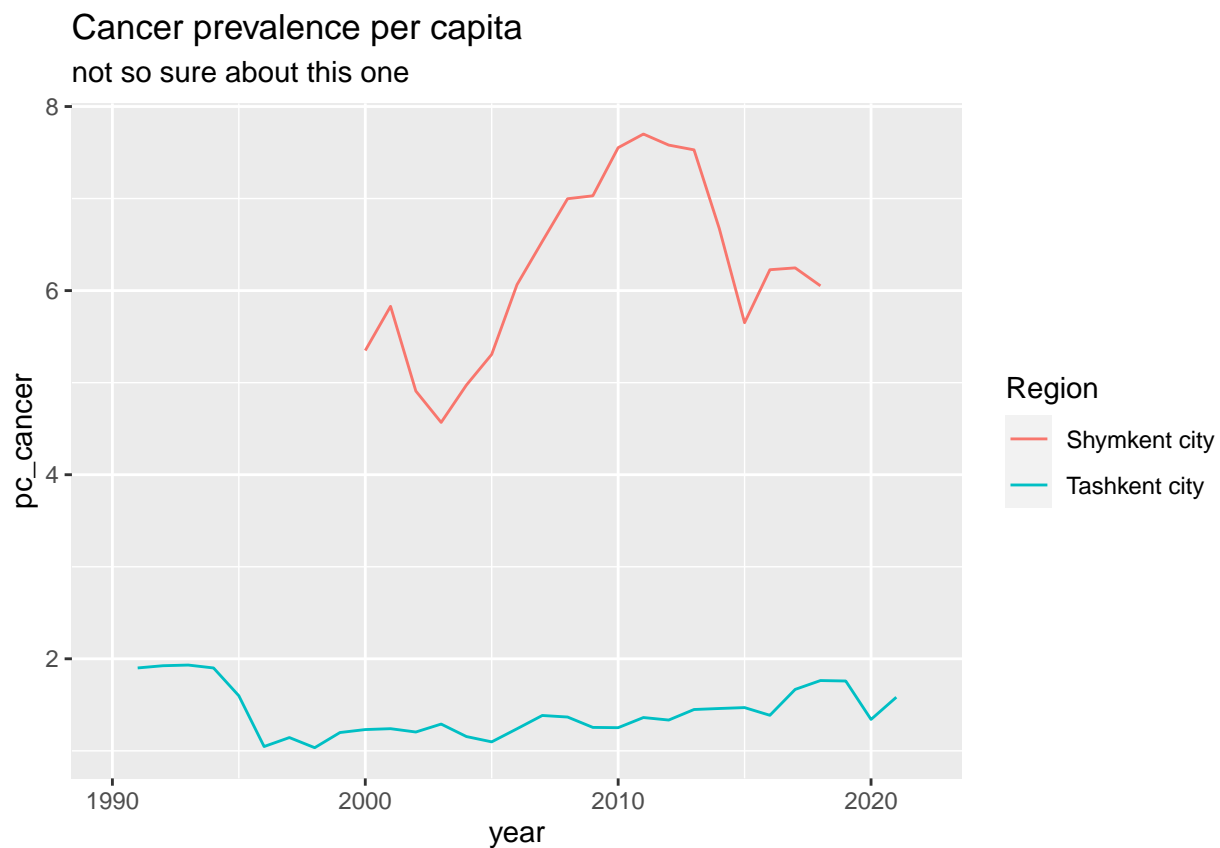
```
glance(TBmodel)$p.value < 0.01 #significant with and without regional effects
```

```
## value
##  TRUE
```

Cancer

```
ggplot(comparisons, aes(x = year, y = pc_cancer)) +
  geom_line(aes(color = Region)) +
  labs(title = "Cancer prevalence per capita",
       subtitle = "not so sure about this one")
```

```
## Warning: Removed 2 rows containing missing values (`geom_line()`).
```



```
ggplot(comparisons, aes(x = year, y = pc_cancer)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Cancer prevalence per capita")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Removed 2 rows containing missing values (`geom_line()`).
```

Cancer prevalence per capita
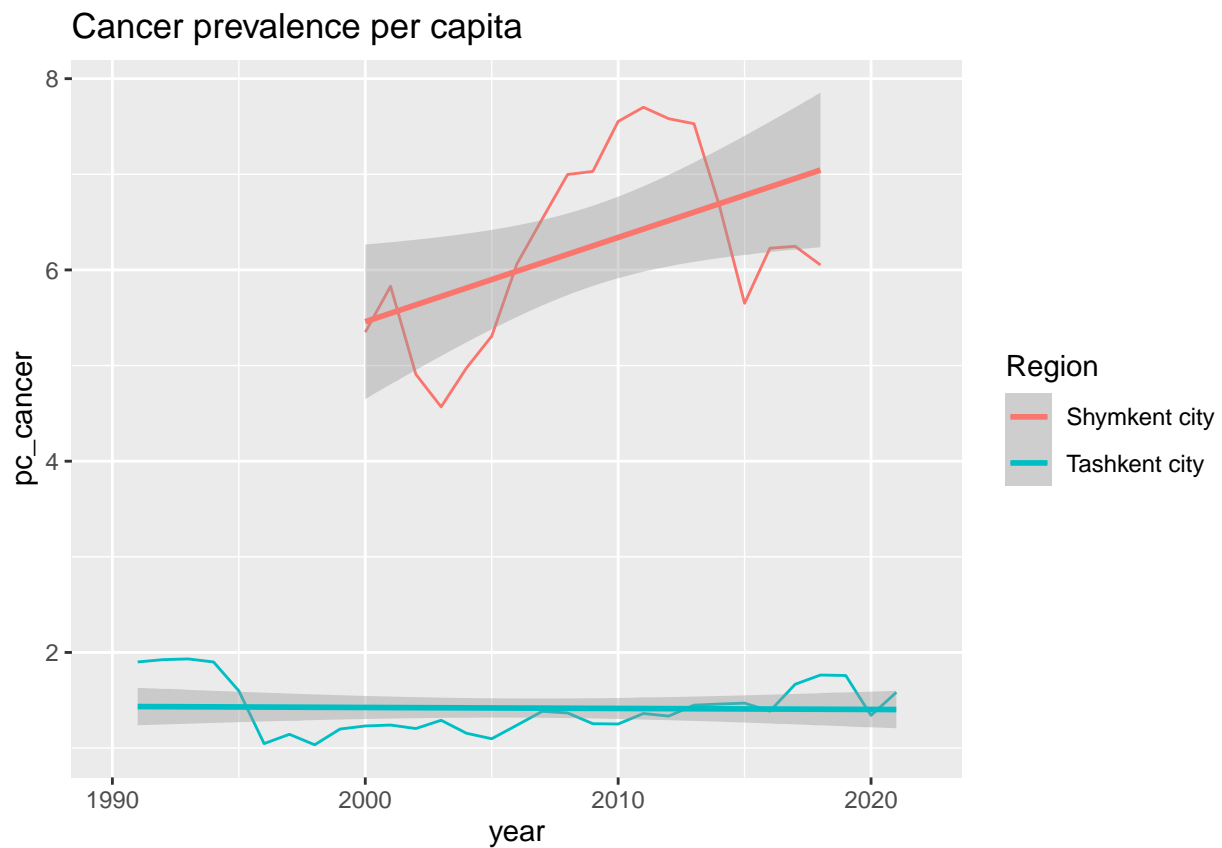

```
cancermodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_cancer ~ year + Region, data = comparisons)
tidy(cancermodel)
```

```
## # A tibble: 3 x 5
##   term               estimate std.error statistic  p.value
##   <chr>                 <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)          -25.1     22.9      -1.10  2.78e- 1
## 2 year                 0.0156   0.0114      1.37  1.77e- 1
## 3 RegionTashkent city  -4.79    0.187     -25.7   2.75e-29
```
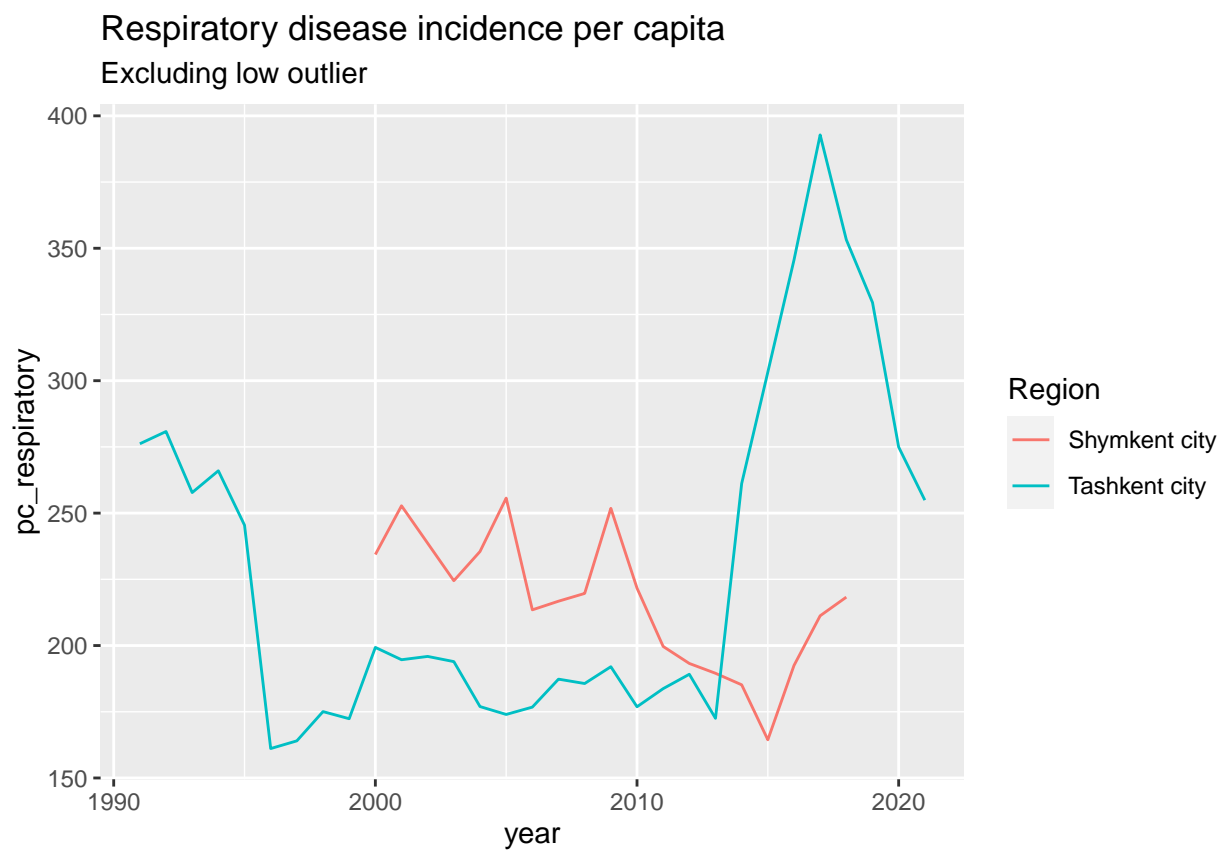
```
glance(cancermodel)$p.value < 0.01 #significant with fixed effects
```

```
## value
##  TRUE
```

Respiratory diseases

```
ggplot(filter(comparisons, pc_respiratory >100), aes(x = year, y = pc_respiratory)) +
  geom_line(aes(color = Region)) +
  labs(title = "Respiratory disease incidence per capita",
       subtitle = "Excluding low outlier")
```
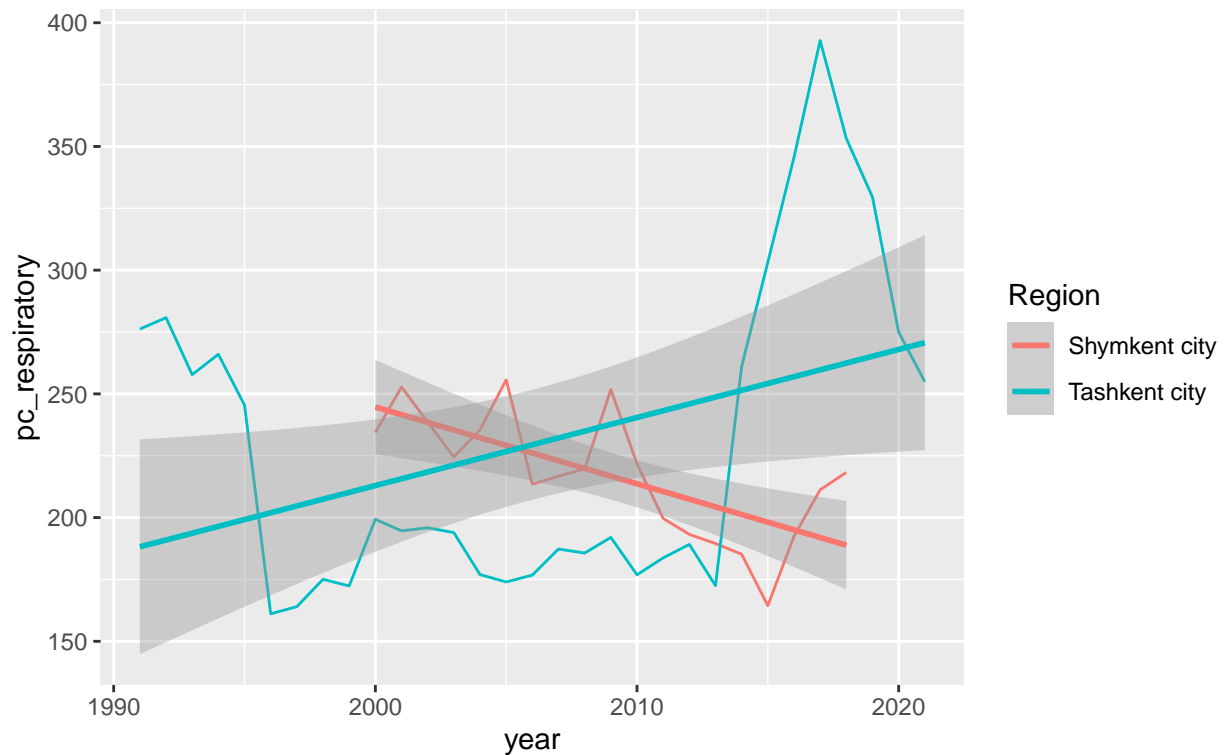


```
ggplot(filter(comparisons, pc_respiratory >100), aes(x = year, y = pc_respiratory)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Respiratory disease incidence per capita",
       subtitle = "Excluding low outlier")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Respiratory disease incidence per capita
Excluding low outlier

```
respmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_respiratory ~ year + Region, data = filter(comparisons, pc_respiratory >100))
tidy(respmodel)
```

```
## # A tibble: 3 x 5
##   term                estimate std.error statistic p.value
##   <chr>                  <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)          -3277.    1926.      -1.70  0.0955
## 2 year                   1.74     0.958      1.81  0.0762
## 3 RegionTashkent city   19.8     15.9        1.24  0.220
```
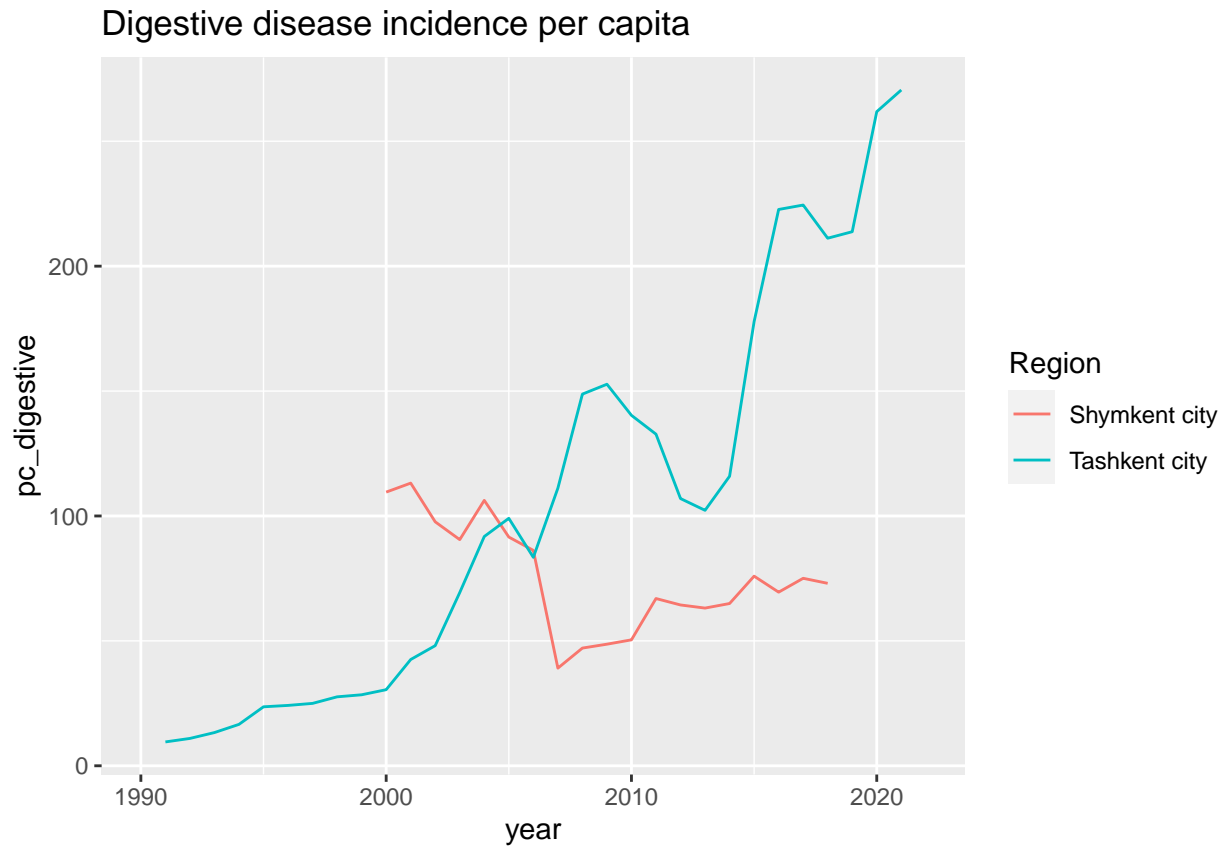
```
glance(respmodel)$p.value < 0.01 #not significant
```

```
## value
## FALSE
```

Digestive diseases

```
ggplot(comparisons, aes(x = year, y = pc_digestive)) +
  geom_line(aes(color = Region)) +
  labs(title = "Digestive disease incidence per capita")
```

```
## Warning: Removed 2 rows containing missing values (`geom_line()`).
```

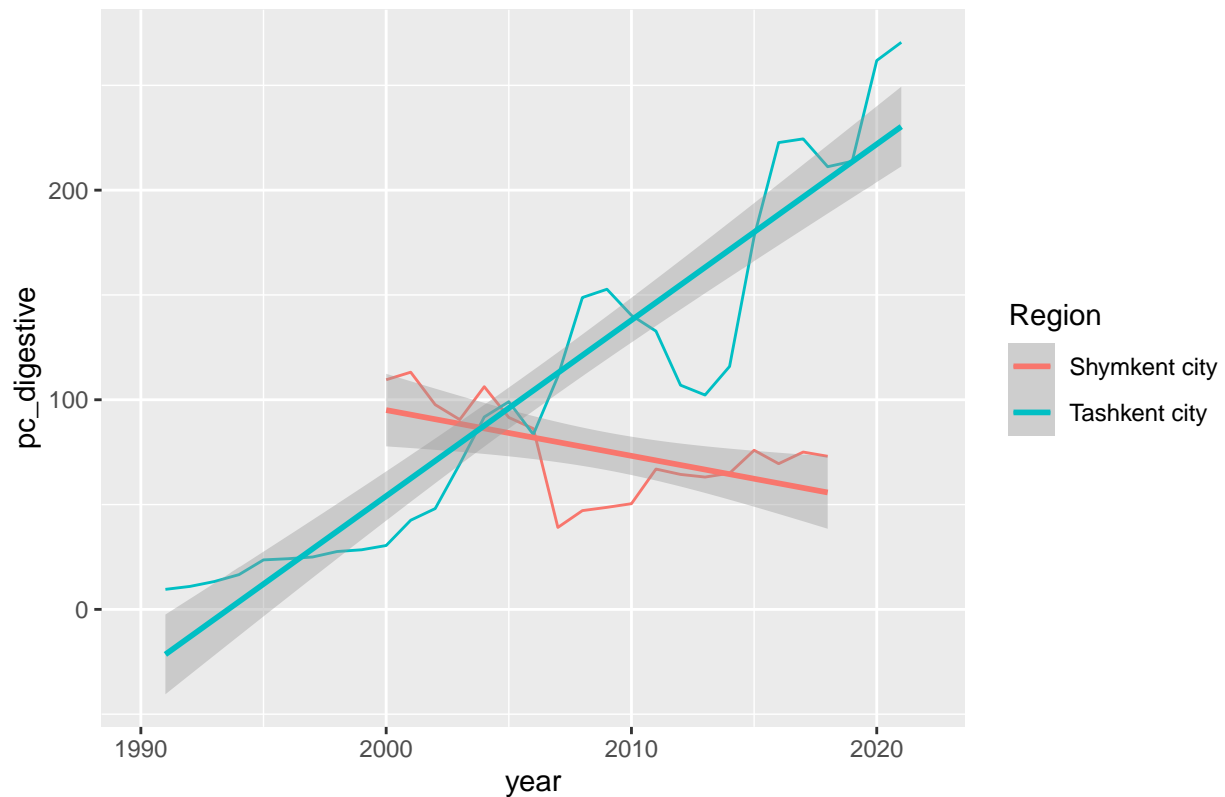## Digestive disease incidence per capita


```
ggplot(comparisons, aes(x = year, y = pc_digestive)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Digestive disease incidence per capita")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Removed 2 rows containing missing values (`geom_line()`).
```

Digestive disease incidence per capita


```r
digmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_digestive ~ year + Region, data = comparisons)
tidy(digmodel)
```

```
## # A tibble: 3 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)          -12817.     1483.     -8.64 2.87e-11
## 2 year                    6.42      0.738     8.69 2.41e-11
## 3 RegionTashkent city    48.2      12.1       3.99 2.28e- 4
```

```r
glance(digmodel)$p.value < 0.01 #significant with and without fixed effects
```
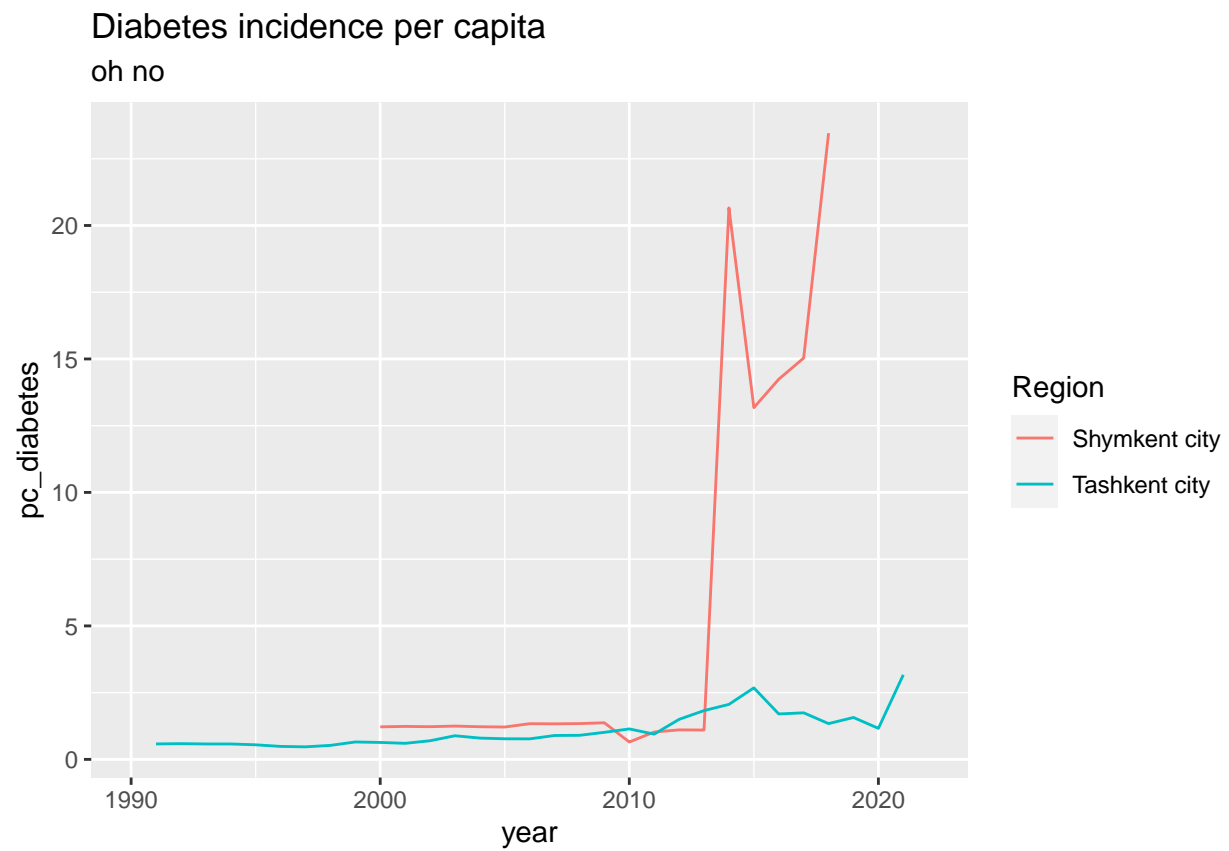
```
## value
##  TRUE
```

Diabetes

```r
ggplot(comparisons, aes(x = year, y = pc_diabetes)) +
  geom_line(aes(color = Region)) +
  labs(title = "Diabetes incidence per capita",
       subtitle = "oh no")
```
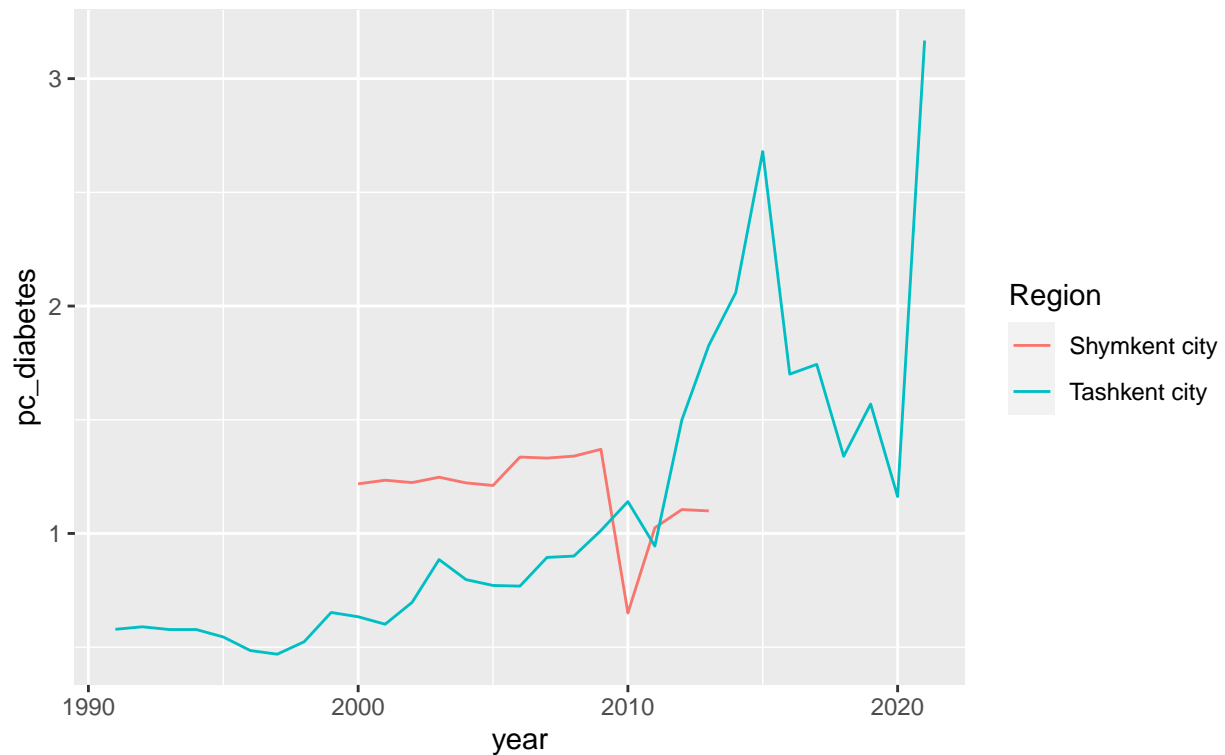
```
## Warning: Removed 2 rows containing missing values (`geom_line()`).
```

Diabetes incidence per capita
oh no


```
ggplot(filter(comparisons, pc_diabetes < 5), aes(x = year, y = pc_diabetes)) +
  geom_line(aes(color = Region)) +
  labs(title = "Diabetes incidence per capita",
       subtitle = "Excluding years with error or changed counting practices")
```
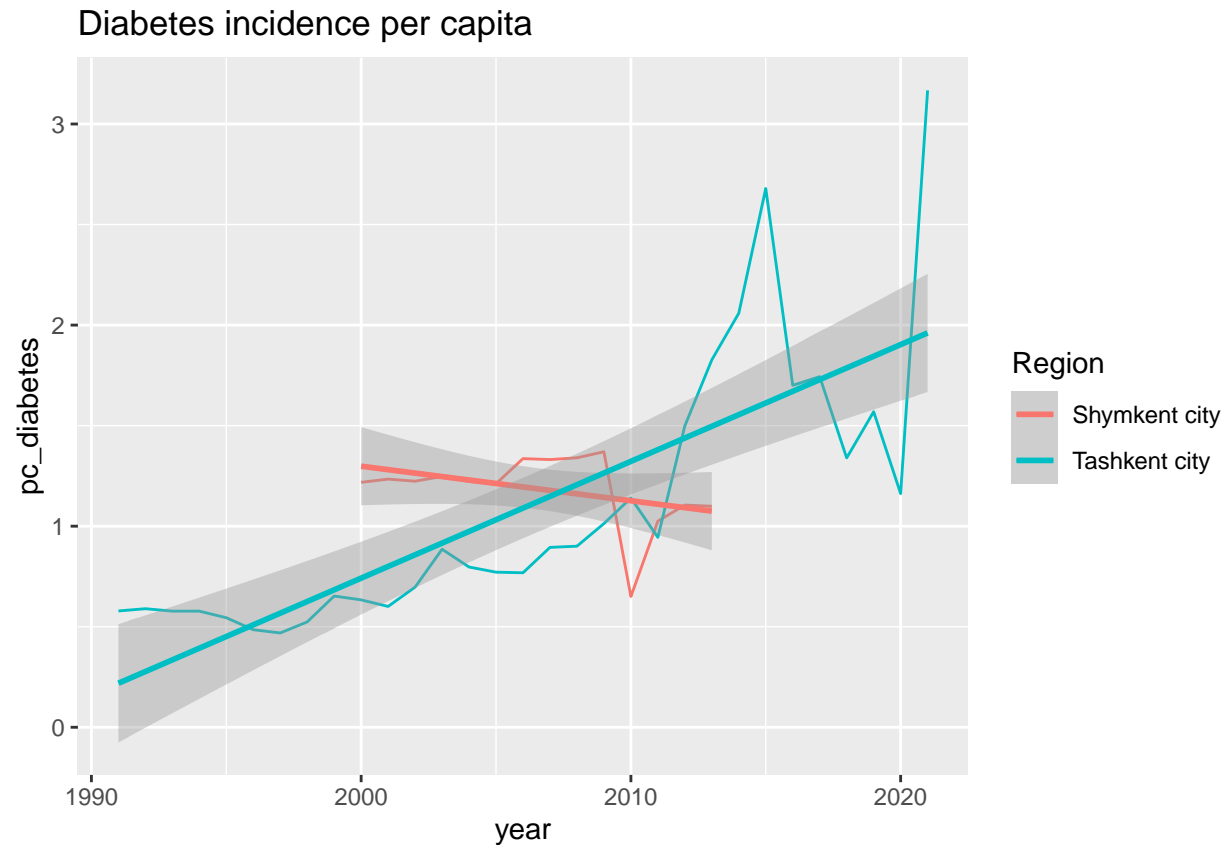
Diabetes incidence per capita
Excluding years with error or changed counting practices


```
ggplot(filter(comparisons, pc_diabetes < 5), aes(x = year, y = pc_diabetes)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Diabetes incidence per capita")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Diabetes incidence per capita


```
diabetesmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_digestive ~ year + Region, data = filter(comparisons, pc_diabetes < 5))
tidy(diabetesmodel)
```

```
## # A tibble: 3 x 5
##   term                 estimate std.error statistic  p.value
##   <chr>                   <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)          -14548.      1454.     -10.0 1.11e-12
## 2 year                    7.29     0.725      10.1 9.44e-13
## 3 RegionTashkent city    31.3      12.1        2.58 1.36e- 2
```

```
glance(diabetesmodel)$p.value < 0.01 #significant with and
```
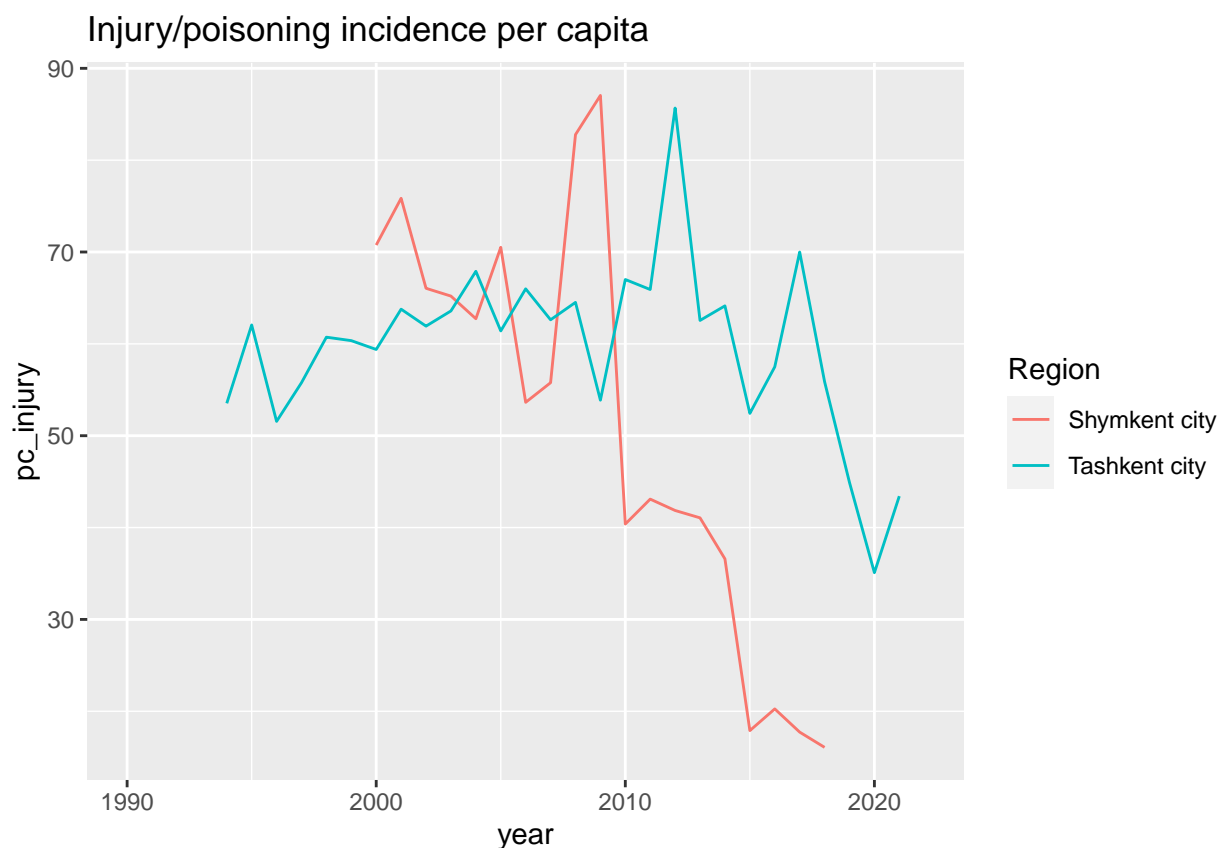
```
## value
##  TRUE
```

```
#without fixed effects, when filtering out measurement error
```

Injury / poisoning

```r
ggplot(comparisons, aes(x = year, y = pc_injury)) +
  geom_line(aes(color = Region)) +
  labs(title = "Injury/poisoning incidence per capita")
```

Warning: Removed 5 rows containing missing values (`geom_line()`).

Injury/poisoning incidence per capita

```r
ggplot(comparisons, aes(x = year, y = pc_injury)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Injury/poisoning incidence per capita")
```

`geom_smooth()` using formula = 'y ~ x'

Warning: Removed 5 rows containing non-finite values (`stat_smooth()`).
Removed 5 rows containing missing values (`geom_line()`).

Injury/poisoning incidence per capita
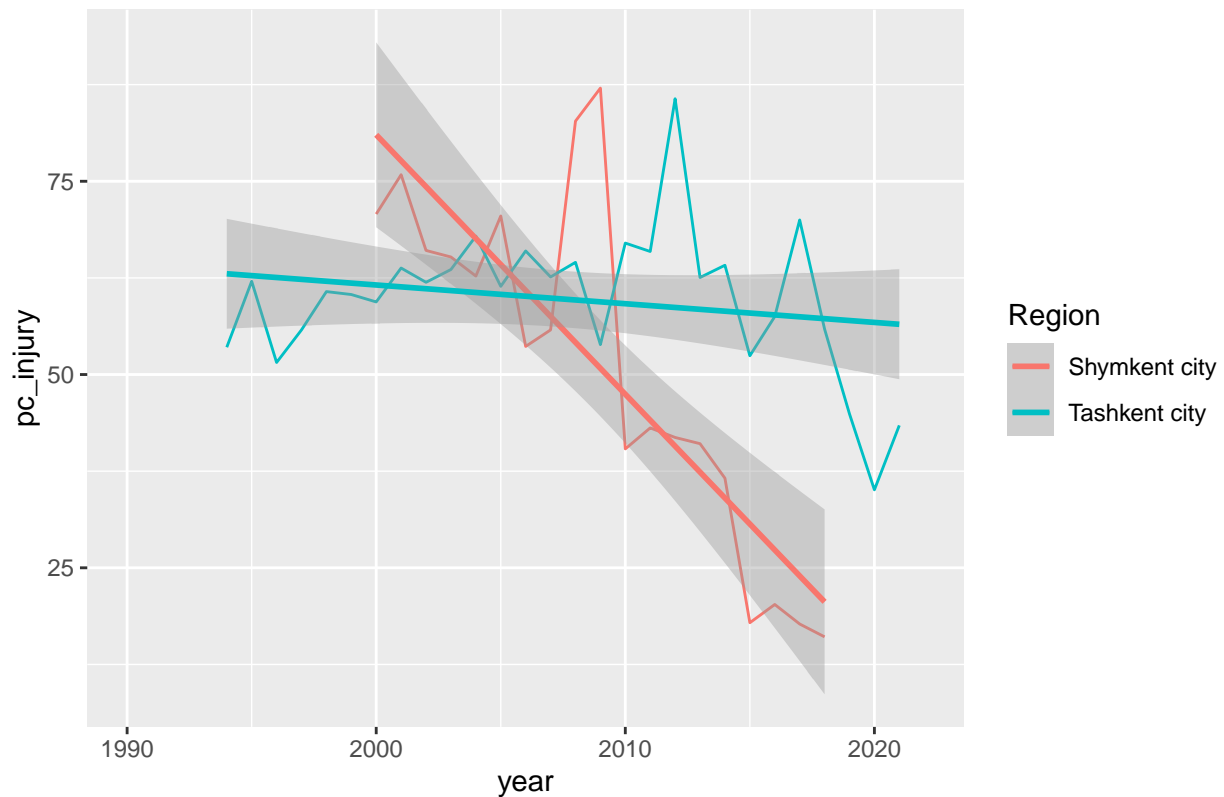

```r
injmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_injury ~ year + Region, data = comparisons)
tidy(injmodel)
```

```
## # A tibble: 3 x 5
##   term               estimate std.error statistic p.value
##   <chr>                 <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)          2025.      597.       3.39 0.00148
## 2 year                 -0.983     0.297     -3.31 0.00188
## 3 RegionTashkent city   7.49      4.35       1.72 0.0919
```

```r
glance(injmodel)$p.value < 0.01 #regional effects not significant, at least for Tashkent city?
```

```
## value
##  TRUE
```

Genitourinary diseases

```r
ggplot(comparisons, aes(x = year, y = pc_genitourinary)) +
  geom_line(aes(color = Region)) +
  labs(title = "Genitourinary diseases incidence per capita")
```
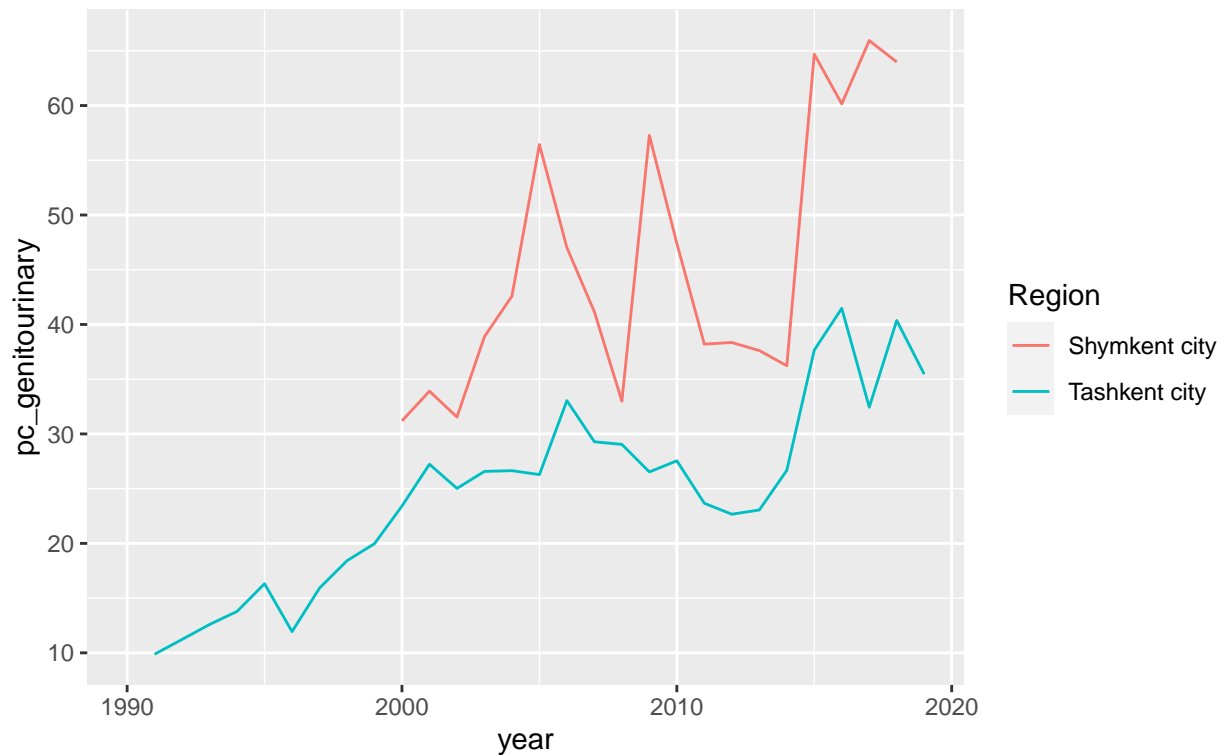
```
## Warning: Removed 2 rows containing missing values ('geom_line()').
```

Genitourinary diseases incidence per capita
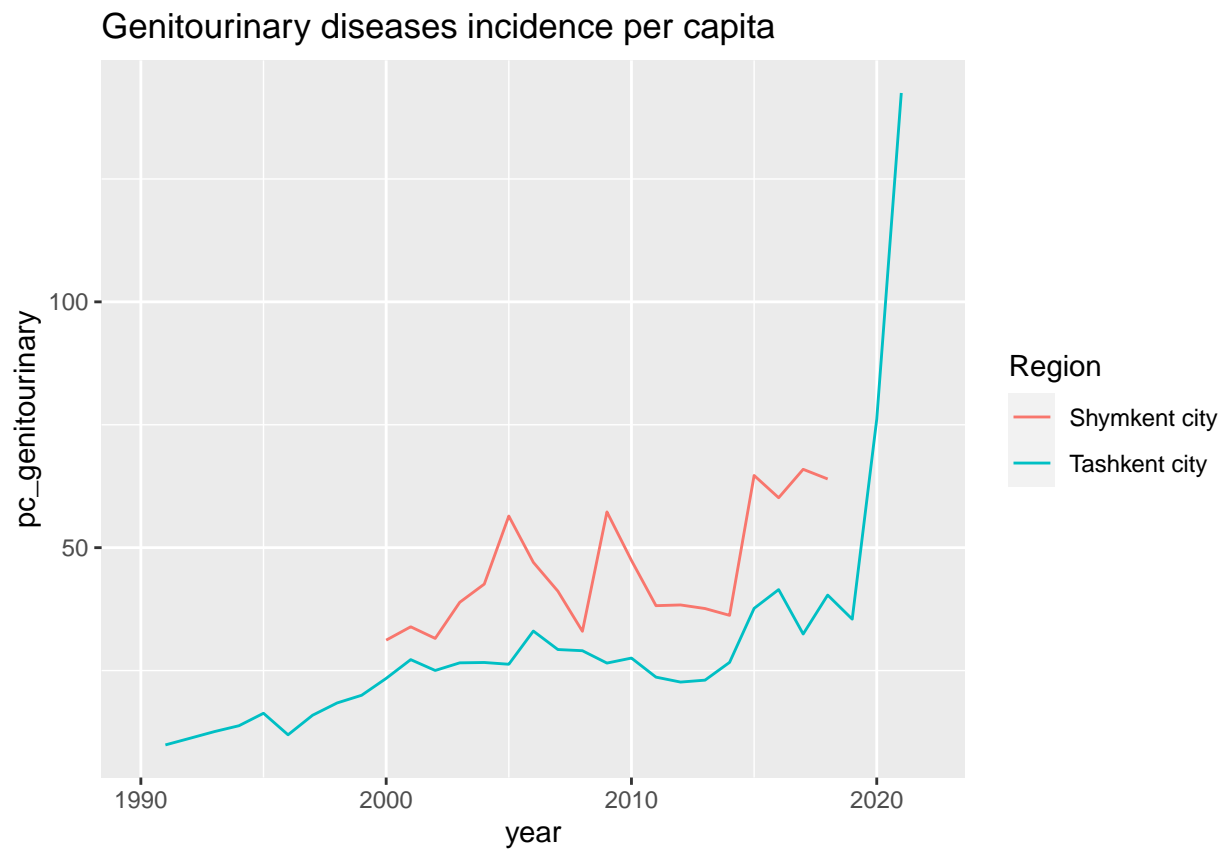

```
ggplot(filter(comparisons, year < 2020), aes(x = year, y = pc_genitourinary)) +
  geom_line(aes(color = Region)) +
  labs(title = "Genitourinary diseases incidence per capita",
       subtitle = "excluding 2020 outlier")
```

```
## Warning: Removed 1 row containing missing values ('geom_line()').
```

## Genitourinary diseases incidence per capita
excluding 2020 outlier


```
ggplot(filter(comparisons, year < 2020), aes(x = year, y = pc_genitourinary)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Genitourinary diseases incidence per capita",
       subtitle = "excluding 2020 outlier")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 1 row containing missing values ('geom_line()').
```

Genitourinary diseases incidence per capita
excluding 2020 outlier
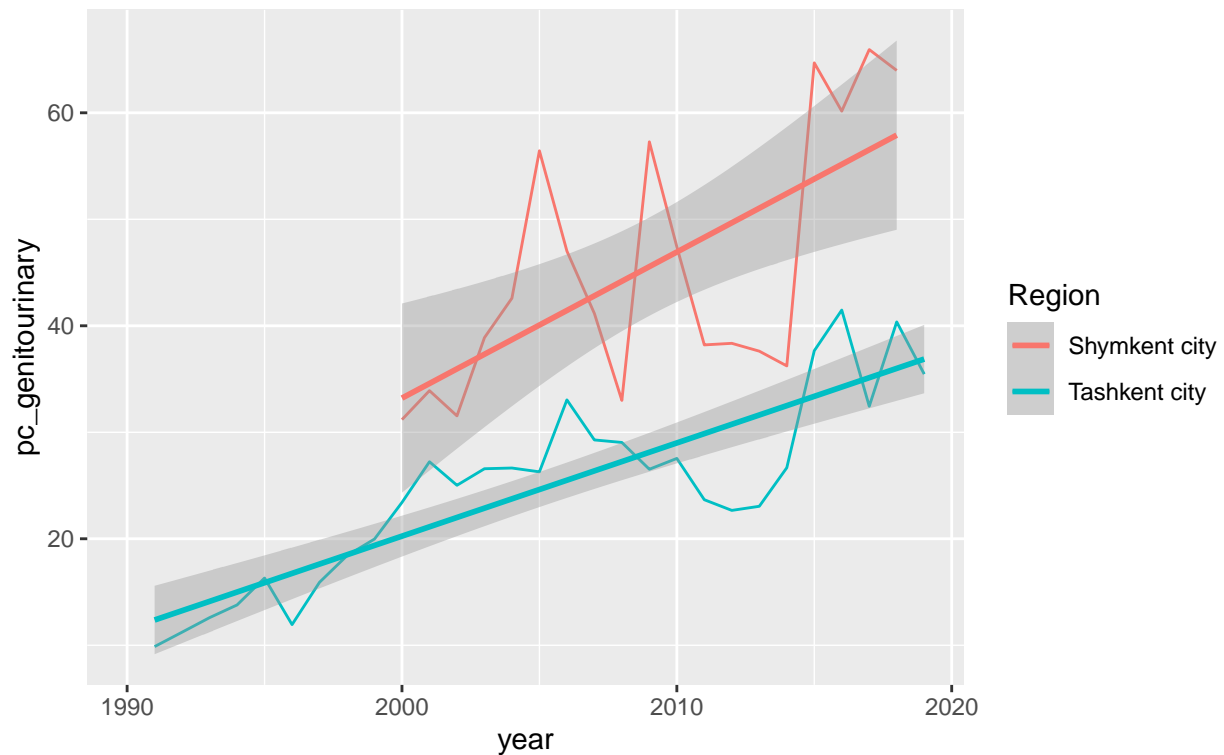

```r
genimodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_genitourinary ~ year + Region, data = filter(comparisons, year < 2020))
tidy(genimodel)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic  p.value
##   <chr>                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)         -1931.      273.     -7.07 8.02e- 9
## 2 year                 0.984     0.136      7.24 4.54e- 9
## 3 RegionTashkent city -17.0       2.12     -8.03 3.14e-10
```

```r
glance(genimodel)$p.value < 0.01 #regional effects significant
```
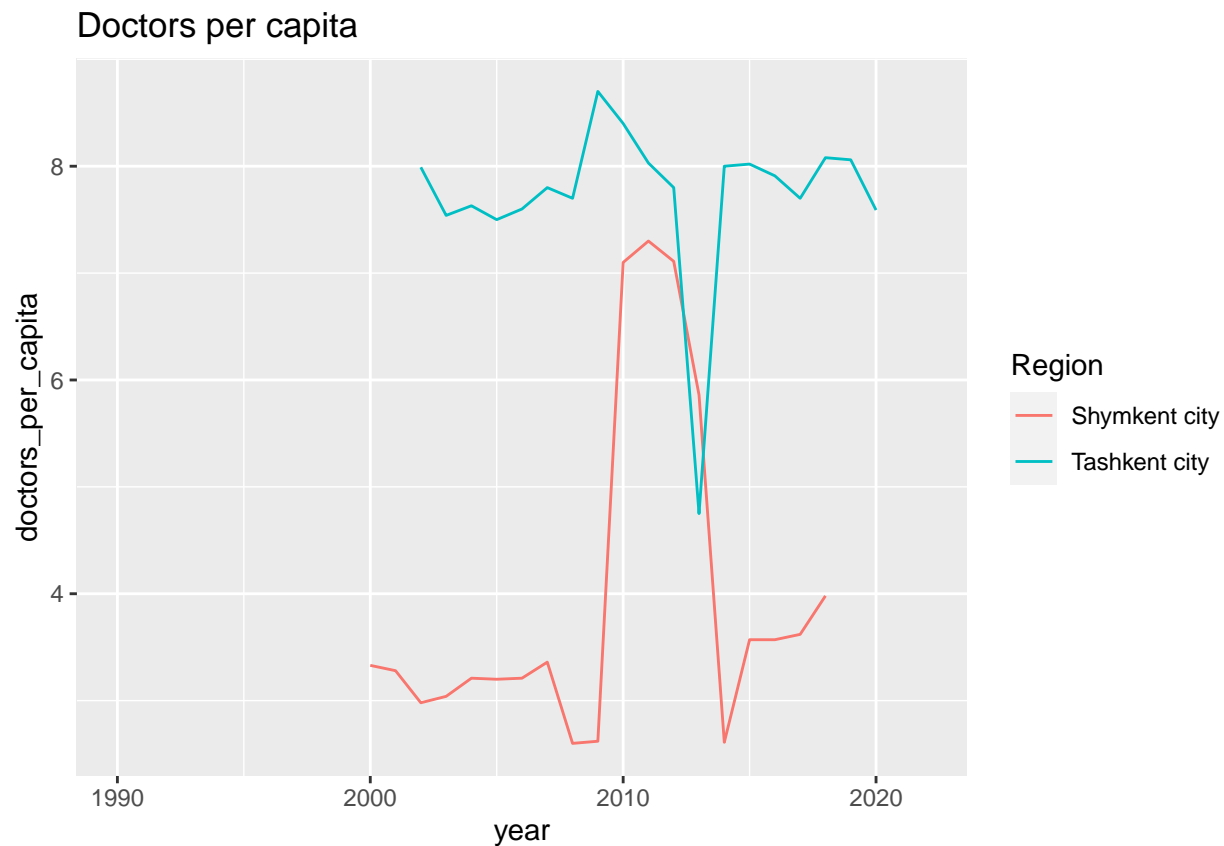
```
## value
##  TRUE
```

HEALTHCARE STRUCTURE

Doctors per Capita

```
ggplot(comparisons, aes(x = year, y = doctors_per_capita)) +
  geom_line(aes(color = Region)) +
  labs(title = "Doctors per capita")
```

## Warning: Removed 14 rows containing missing values ('geom_line()').

Doctors per capita

```
# not sure what's due to error. I will filter out all points
#between 4.5-7.5 because I believe these may be due to changes in
#measurement practice or data collection error.

comparisons %>%
  filter(doctors_per_capita <= 4.5 | doctors_per_capita >= 7.5) %>%
ggplot(aes(x = year, y = doctors_per_capita)) +
  geom_line(aes(color = Region)) +
  labs(title = "Doctors per capita", subtitle = "filtering out potential measurement error or changes i
```

Doctors per capita

filtering out potential measurement error or changes in measurement practice
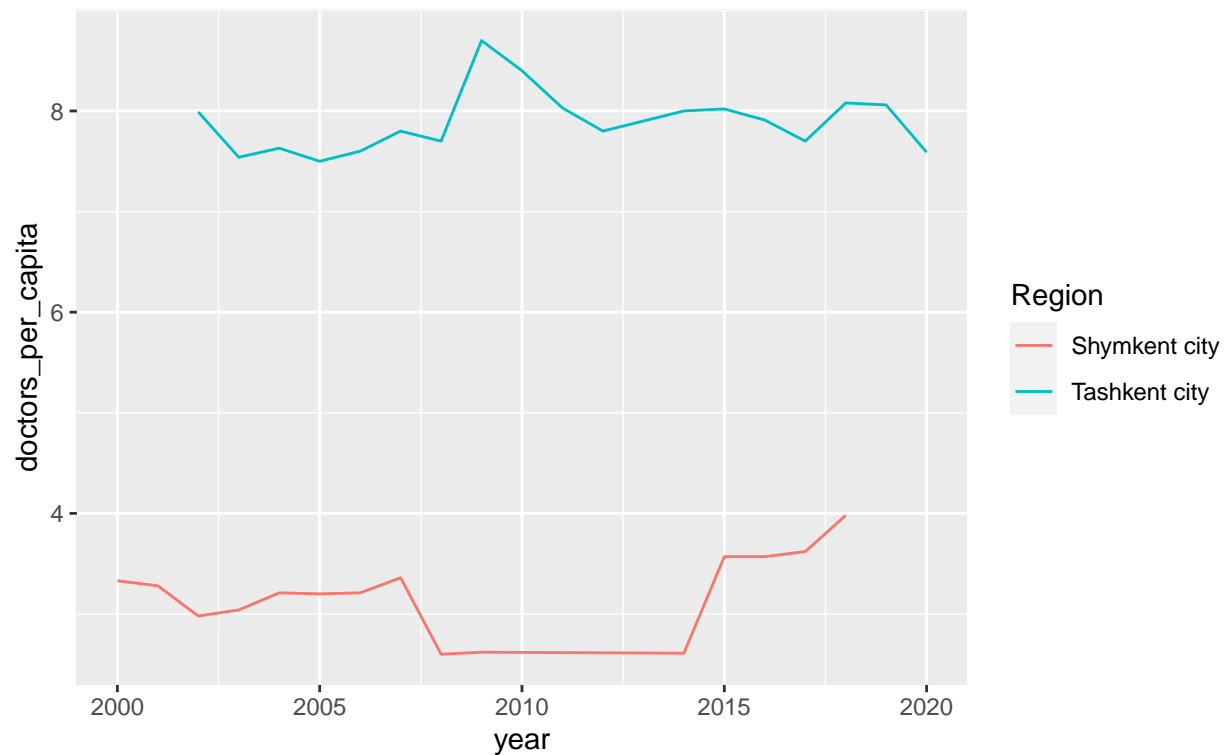


```
comparisons %>%
  filter(doctors_per_capita <= 4.5 | doctors_per_capita >= 7.5) %>%
ggplot(aes(x = year, y = doctors_per_capita)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Doctors per capita", subtitle = "filtering out potential changes in measurement practice
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Doctors per capita

filtering out potential changes in measurement practice or error
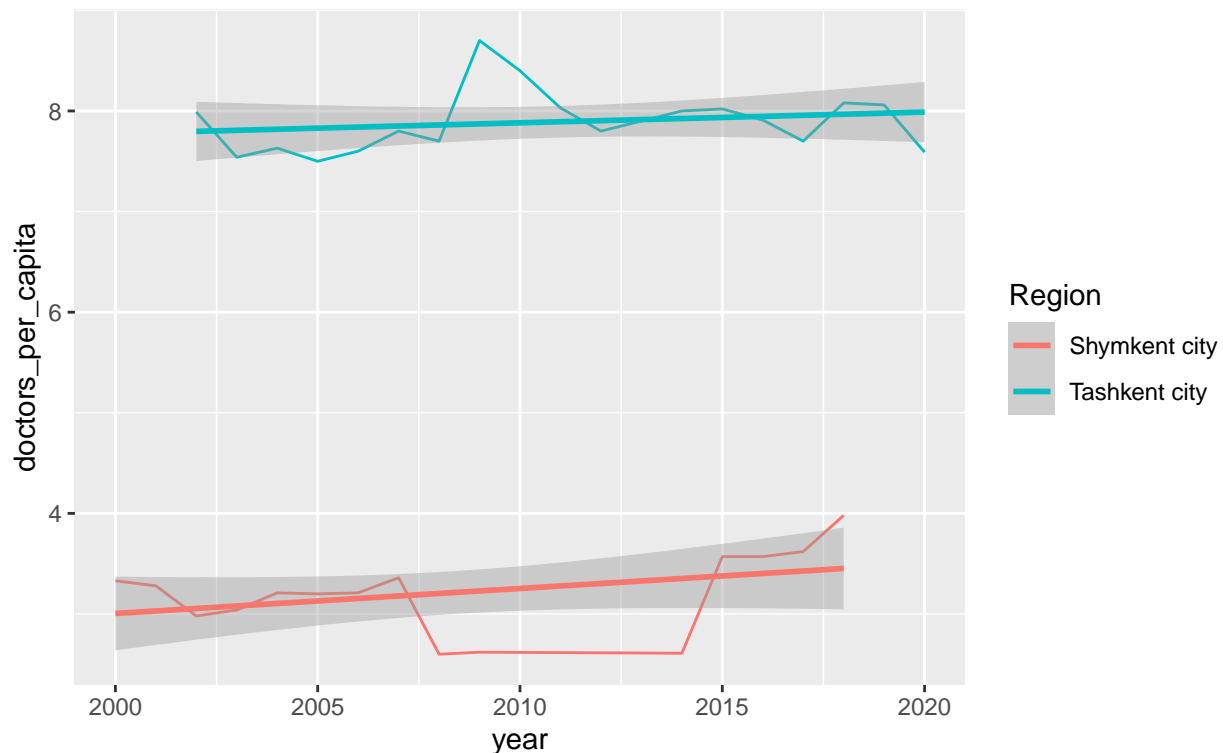
```
doctorsmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(doctors_per_capita ~ year + Region, data = filter(comparisons, doctors_per_capita <= 4.5 | doctors
tidy(doctorsmodel)
```

```
## # A tibble: 3 x 5
##   term               estimate std.error statistic  p.value
##   <chr>                 <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)          -32.0      20.8     -1.54 1.34e- 1
## 2 year                0.0175    0.0103      1.69 1.01e- 1
## 3 RegionTashkent city   4.63     0.123     37.7  7.62e-27
```
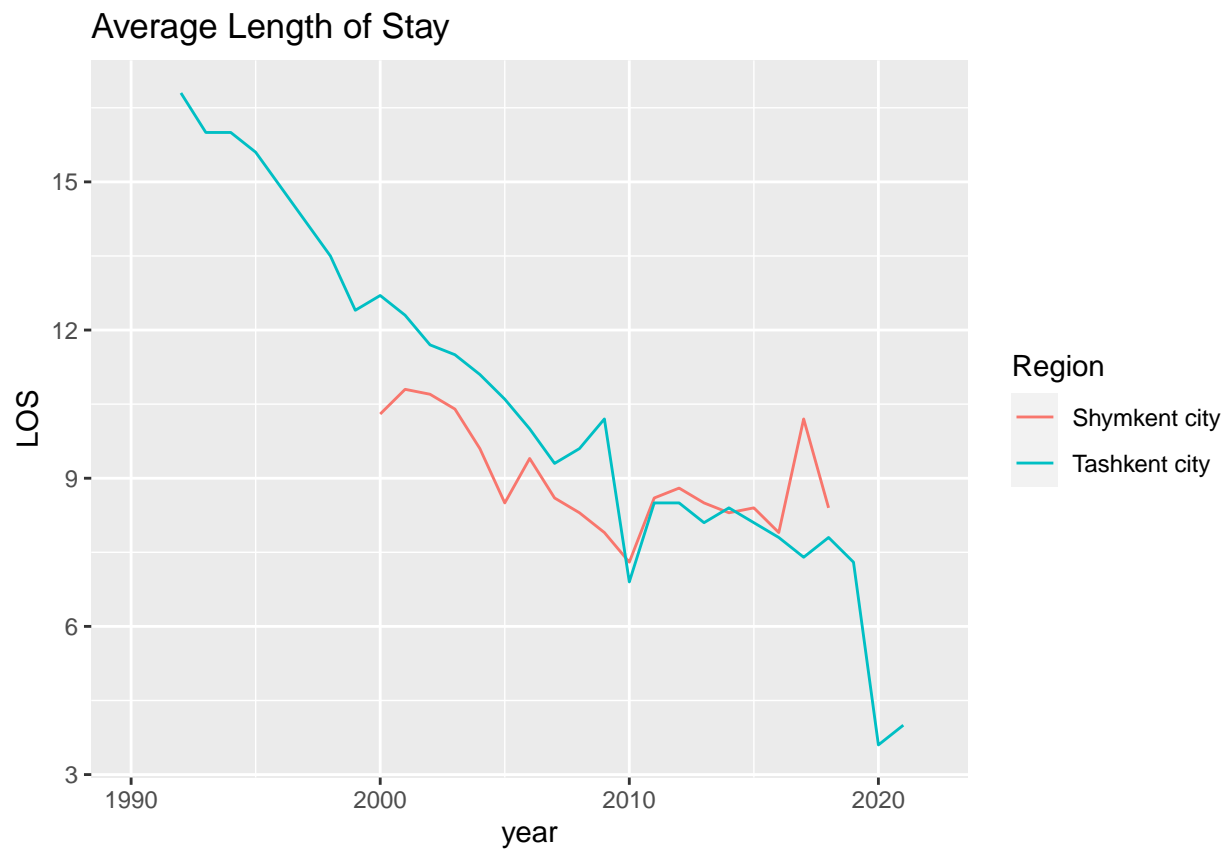
```
glance(doctorsmodel)$p.value < 0.01 #regional effects significant, when filtering out error
```

```
## value
##  TRUE
```

Average Length of Stay

```
ggplot(comparisons, aes(x = year, y = LOS)) +
  geom_line(aes(color = Region)) +
  labs(title = "Average Length of Stay")
```

```
## Warning: Removed 3 rows containing missing values (`geom_line()`).
```

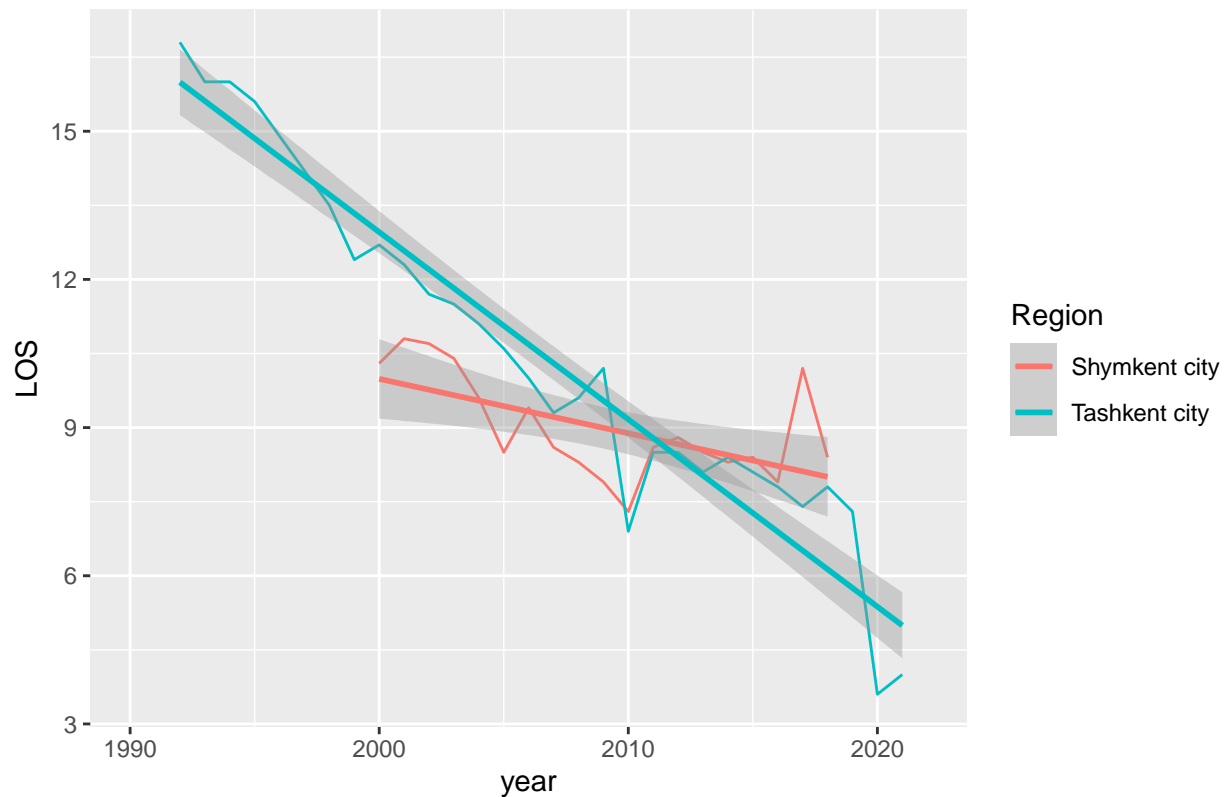## Average Length of Stay


```
ggplot(comparisons, aes(x = year, y = LOS)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Average Length of Stay")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).
## Removed 3 rows containing missing values (`geom_line()`).
```

Average Length of Stay

```r
LOSmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(LOS ~ year + Region, data = comparisons)
tidy(LOSmodel)
```

```
## # A tibble: 3 x 5
##   term               estimate std.error statistic  p.value
##   <chr>                 <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)           662.      46.3      14.3  1.60e-18
## 2 year                 -0.325    0.0231    -14.1  2.68e-18
## 3 RegionTashkent city   0.687    0.363       1.89 6.52e- 2
```
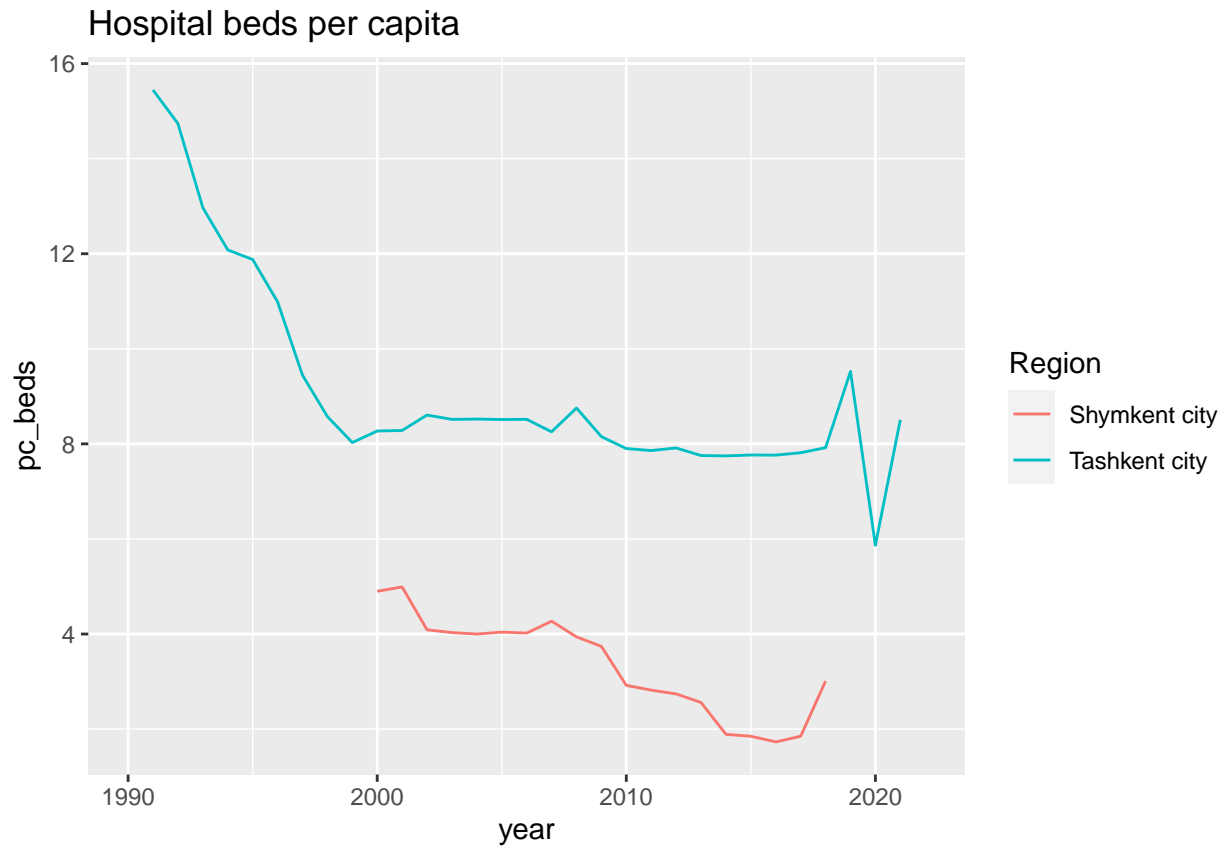
```r
glance(LOSmodel)$p.value < 0.01 #regional effects not significant but this makes no sense
```

```
## value
##  TRUE
```

Hospital beds per capita

```r
ggplot(comparisons, aes(x = year, y = pc_beds)) +
  geom_line(aes(color = Region)) +
  labs(title = "Hospital beds per capita")
```

Warning: Removed 2 rows containing missing values (`geom_line()`).
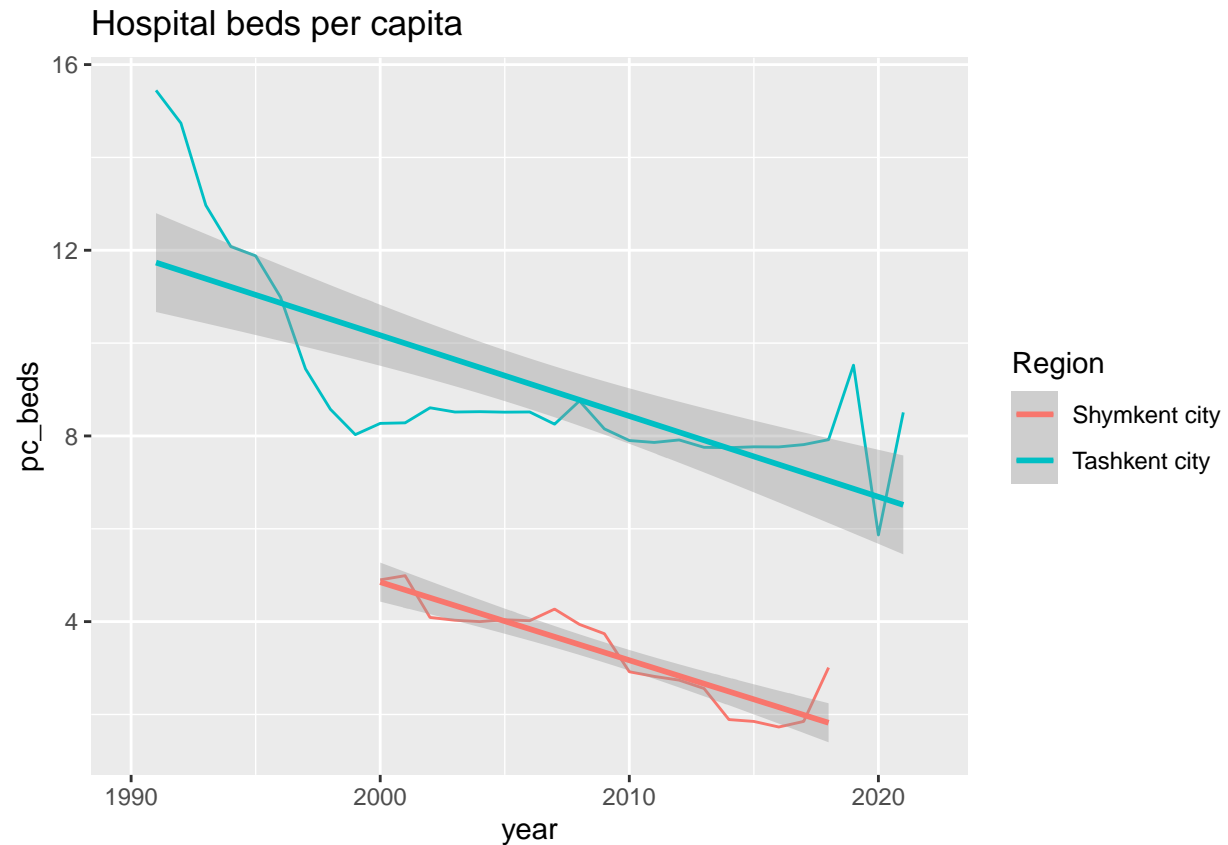
Hospital beds per capita


```
ggplot(comparisons, aes(x = year, y = pc_beds)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Hospital beds per capita")
```

`geom_smooth()` using formula = 'y ~ x'

Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
Removed 2 rows containing missing values (`geom_line()`).

Hospital beds per capita


```r
bedmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_beds ~ year + Region, data = comparisons)
tidy(bedmodel)
```

```
## # A tibble: 3 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)            351.      43.6       8.05 2.14e-10
## 2 year                  -0.173    0.0217     -7.97 2.78e-10
## 3 RegionTashkent city     5.27     0.355     14.8  2.29e-19
```
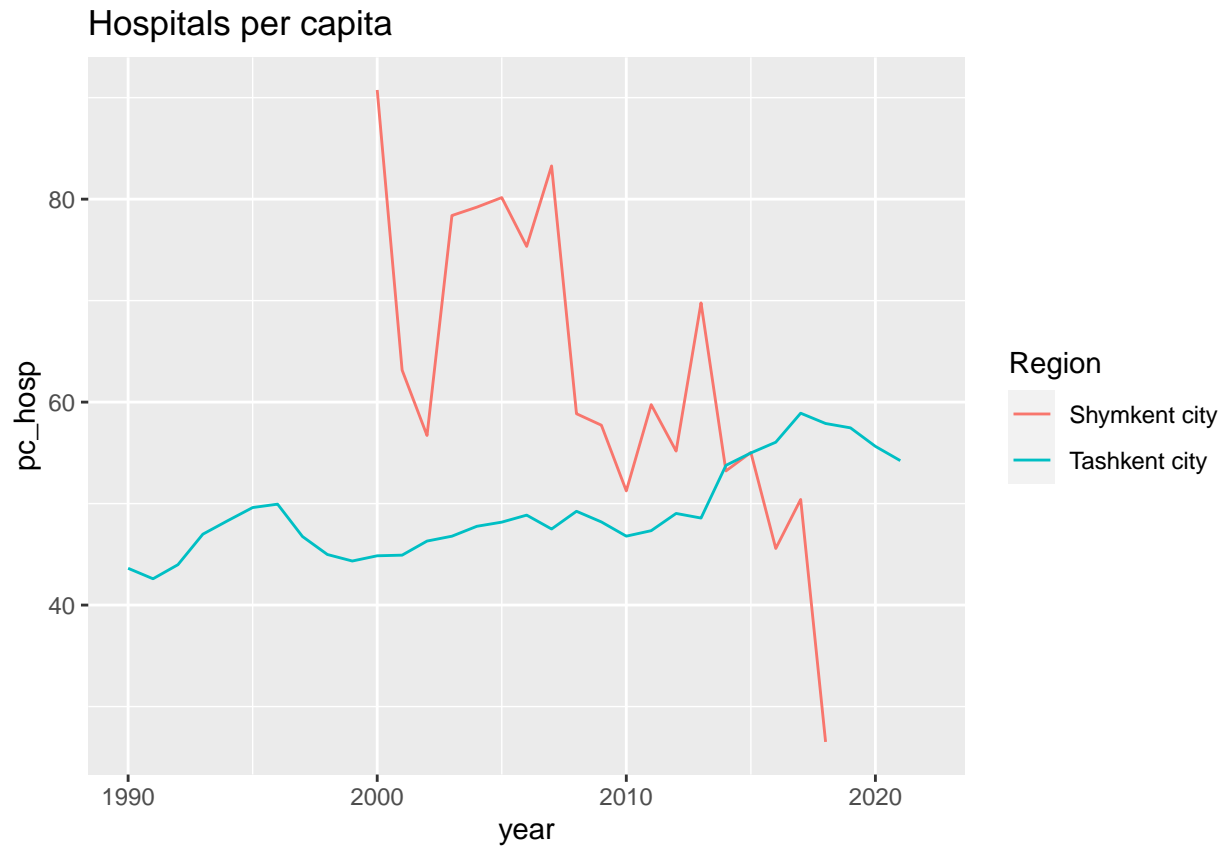
```r
glance(bedmodel)$p.value < 0.01 #regional effects significant
```

```
## value
##  TRUE
```

Hospitals per capita

```r
ggplot(comparisons, aes(x = year, y = pc_hosp)) +
  geom_line(aes(color = Region)) +
  labs(title = "Hospitals per capita")
```

```
## Warning: Removed 1 row containing missing values (`geom_line()`).
```
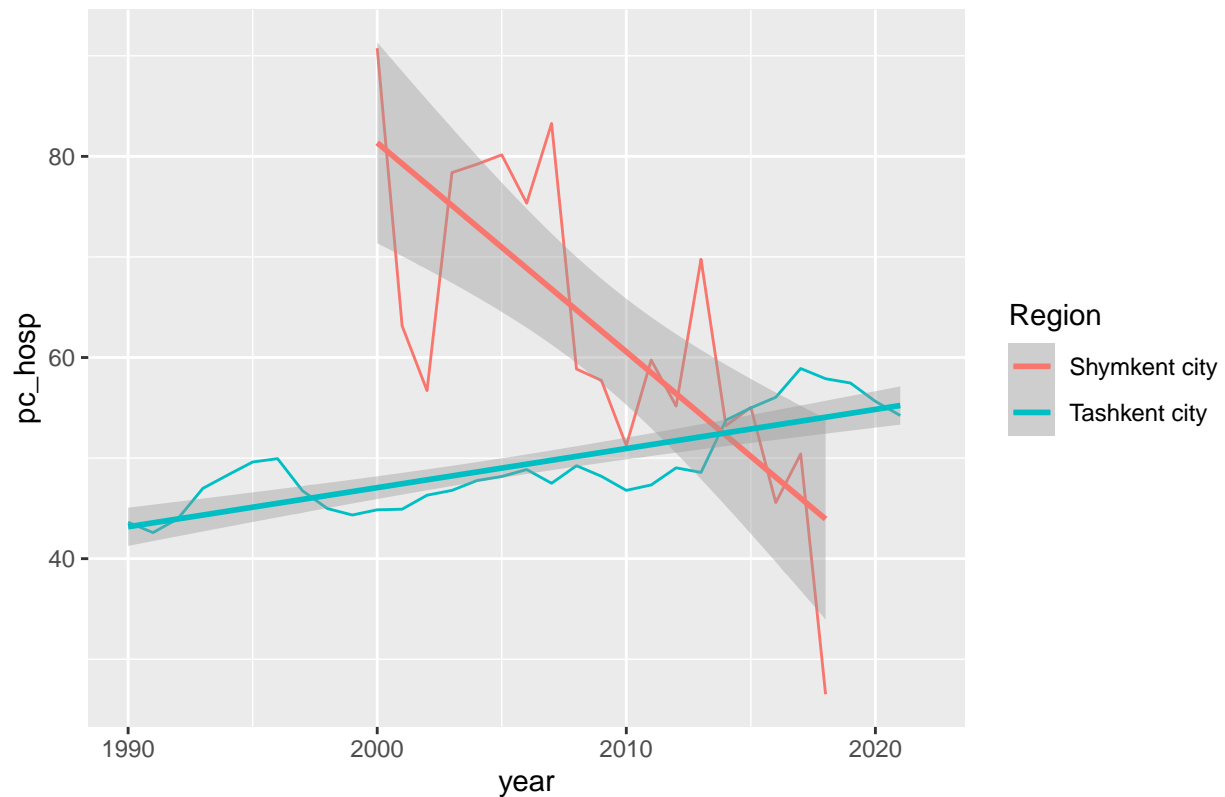
Hospitals per capita


```
ggplot(comparisons, aes(x = year, y = pc_hosp)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Hospitals per capita")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
## Removed 1 row containing missing values (`geom_line()`).
```

Hospitals per capita


```
hospmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pc_hosp ~ year + Region, data = comparisons)
tidy(hospmodel)
```

```
## # A tibble: 3 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)          138.      359.        0.385 0.702
## 2 year                  -0.0375    0.178    -0.210 0.834
## 3 RegionTashkent city  -13.6       3.03     -4.48  0.0000469
```

```
glance(hospmodel)$p.value < 0.01 #regional effects just barely not significant
```
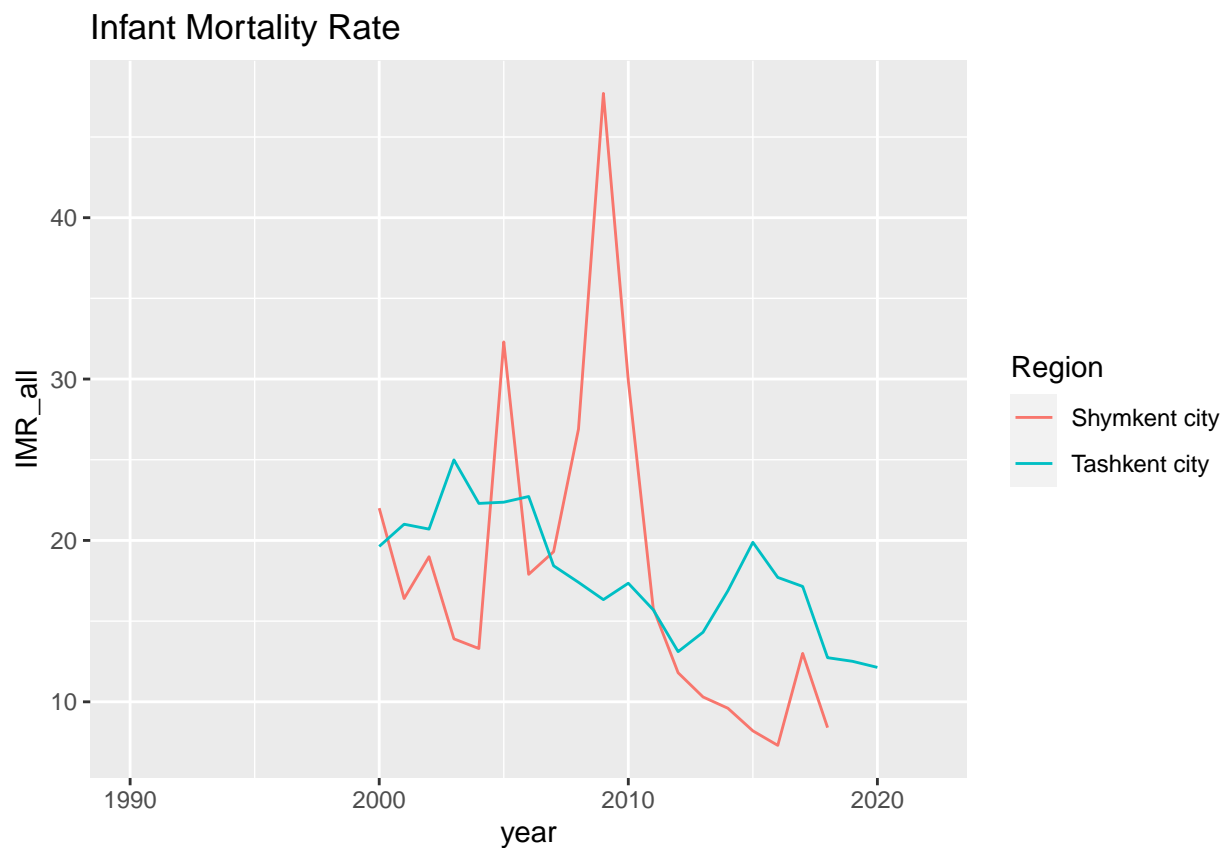
```
## value
##  TRUE
```

Mortality and demography

IMR

```
ggplot(comparisons, aes(x = year, y = IMR_all)) +
  geom_line(aes(color = Region)) +
  labs(title = "Infant Mortality Rate")
```

```
## Warning: Removed 12 rows containing missing values ('geom_line()').
```
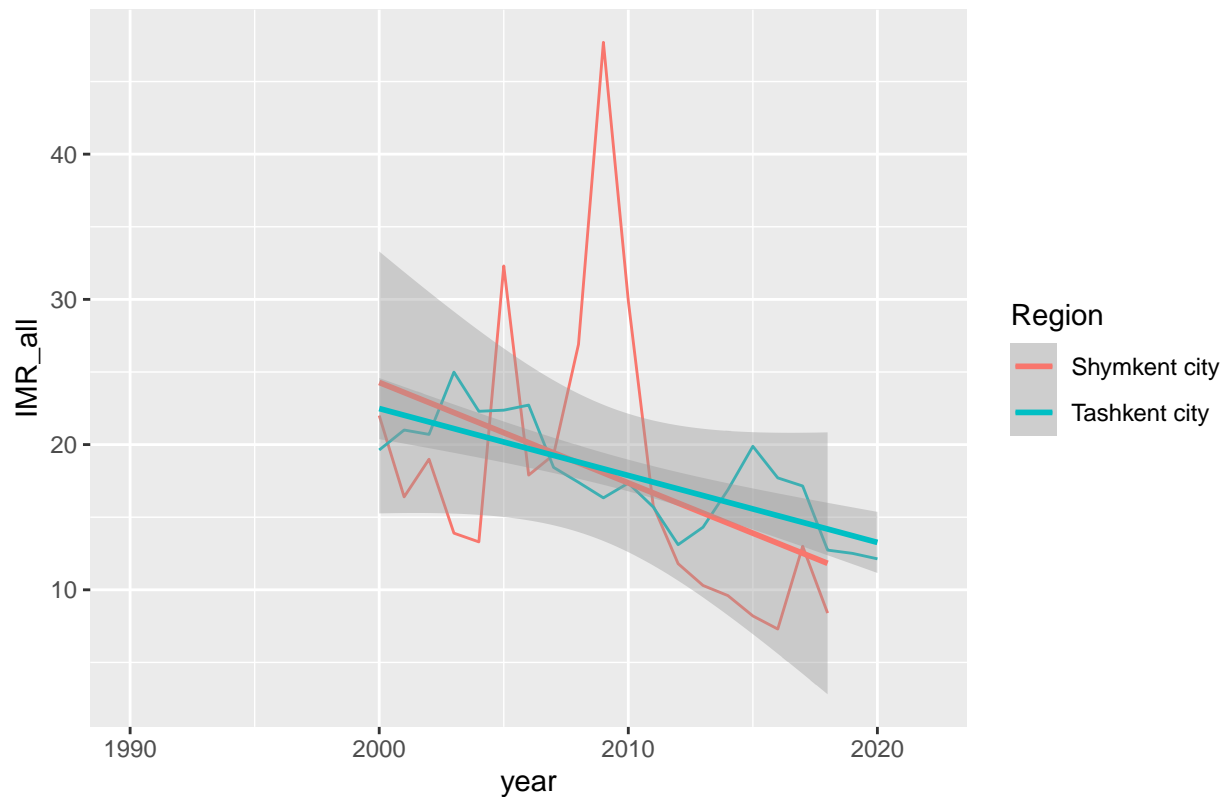

Infant Mortality Rate

```
ggplot(comparisons, aes(x = year, y = IMR_all)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Infant Mortality Rate")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 12 rows containing non-finite values ('stat_smooth()').
## Removed 12 rows containing missing values ('geom_line()').
```

Infant Mortality Rate


```
IMRmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(IMR_all ~ year + Region, data = comparisons)
tidy(IMRmodel)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic p.value
##   <chr>                <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)         1142.      374.       3.05 0.00422
## 2 year                -0.559      0.186    -3.00 0.00479
## 3 RegionTashkent city  0.381      2.17      0.176 0.862
```

```
glance(IMRmodel)$p.value < 0.01 #regional effects not significant
```

```
## value
## FALSE
```

```
IMRmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(IMR_all ~ year, data = comparisons)
tidy(IMRmodel)
```

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic p.value
```

```
##    <chr>          <dbl>     <dbl>      <dbl>    <dbl>
## 1 (Intercept) 1136.        368.        3.09 0.00378
## 2 year          -0.556      0.183     -3.04 0.00431
```
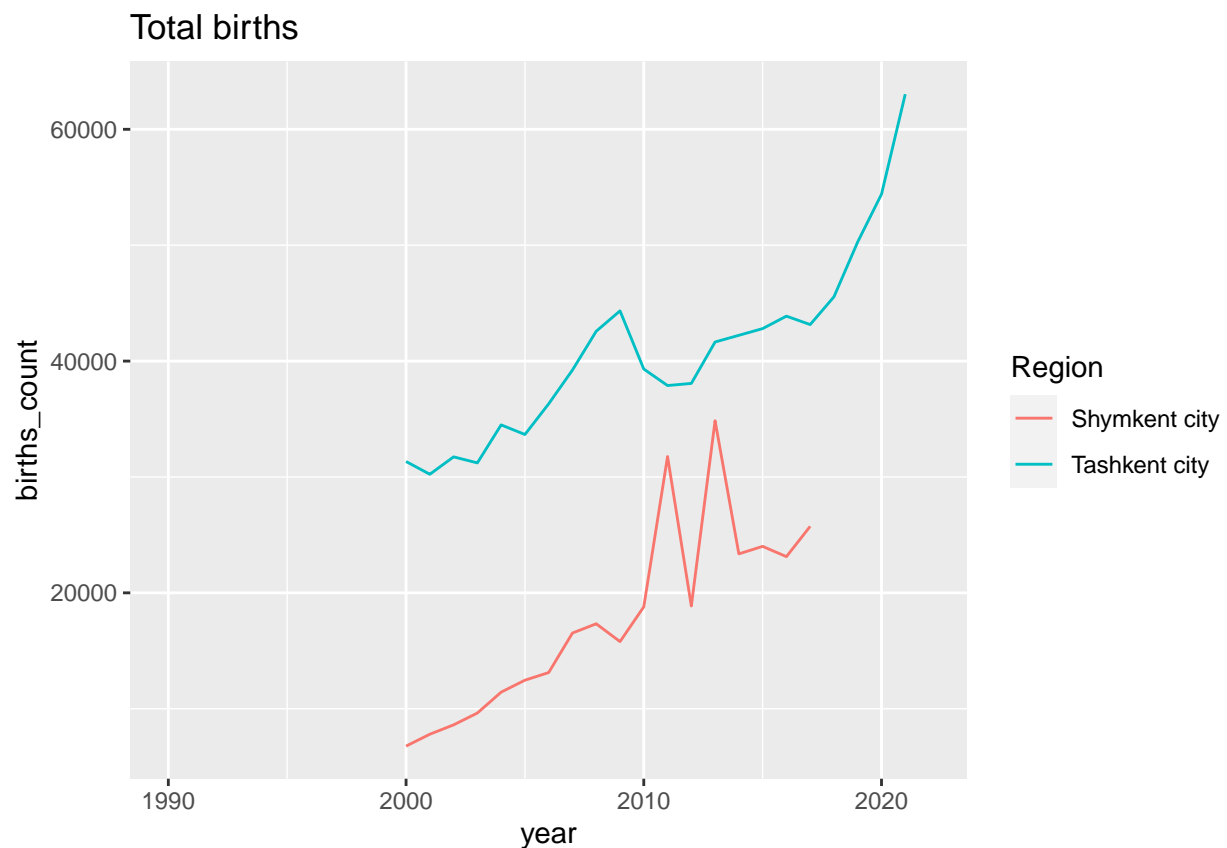
```
glance(IMRmodel)$p.value < 0.01 #regular model without regional effects is significant
```

```
## value
##  TRUE
```

Births and Birth Rate

```
ggplot(comparisons, aes(x = year, y = births_count)) +
  geom_line(aes(color = Region)) +
  labs(title = "Total births")
```
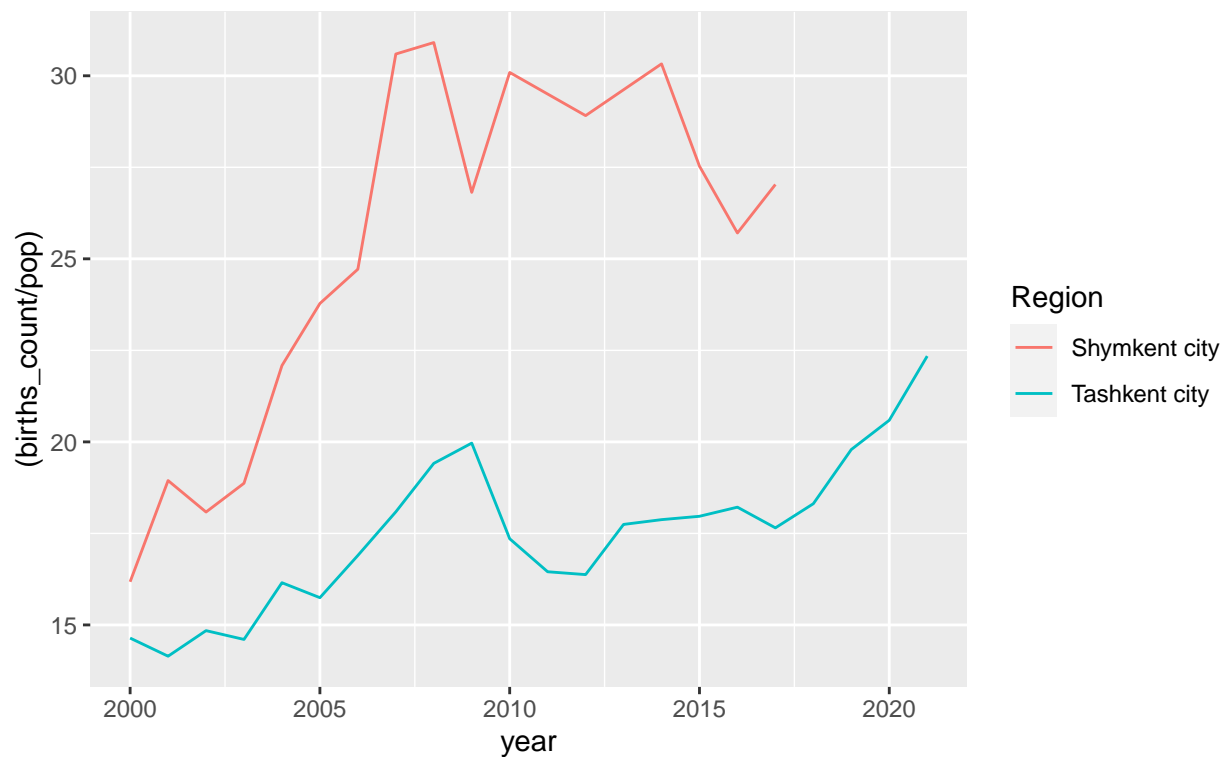
```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
```



```
ggplot(filter(comparisons, (births_count / pop) <= 40), aes(x = year, y = (births_count / pop))) +
  geom_line(aes(color = Region)) +
  labs(title = "Total birth rate (live births per capita)",
       subtitle = "filtering out probable error")
```

Total birth rate (live births per capita)
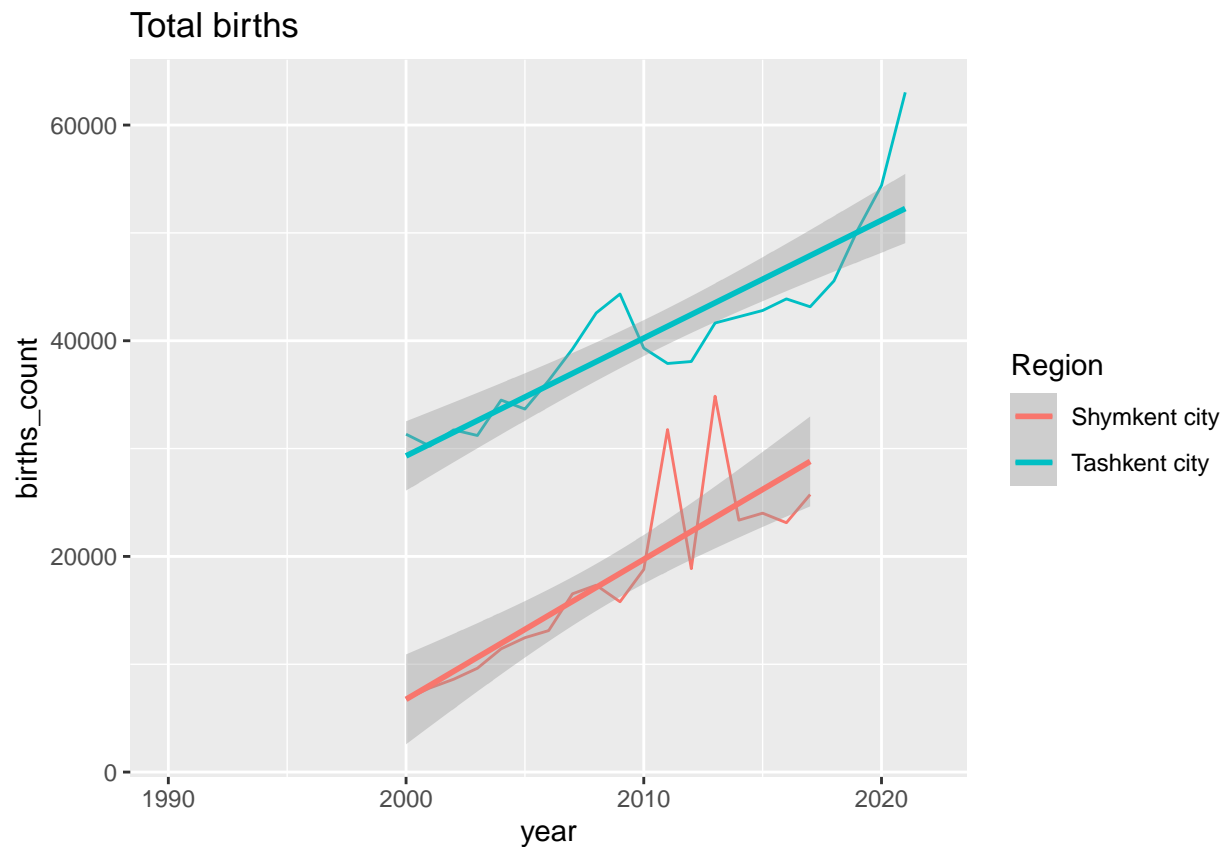filtering out probable error


```
ggplot(comparisons, aes(x = year, y = births_count)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Total births")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 12 rows containing non-finite values (`stat_smooth()`).
## Removed 12 rows containing missing values (`geom_line()`).
```
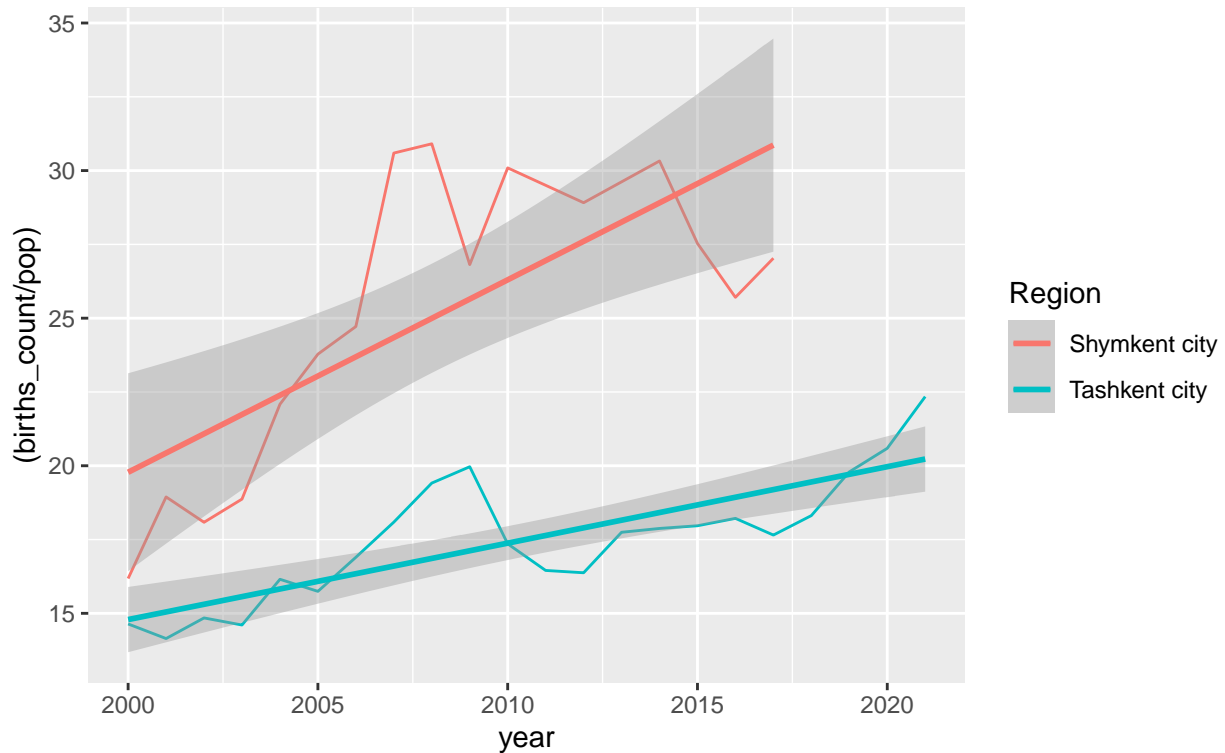
Total births


```r
ggplot(filter(comparisons, (births_count / pop) <= 40), aes(x = year, y = (births_count / pop))) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Total birth rate (live births per capita)",
       subtitle = "filtering out probable error")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Total birth rate (live births per capita)
filtering out probable error


```r
birthmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(births_count ~ year + Region, data = comparisons)
tidy(birthmodel)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic  p.value
##   <chr>                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)      -2322339.   217559.     -10.7 7.52e-13
## 2 year                 1165.      108.      10.8 6.06e-13
## 3 RegionTashkent city 20682.     1293.      16.0 3.28e-18
```

```r
glance(birthmodel)$p.value < 0.01 #regional effects significant
```

```
## value
##  TRUE
```

```r
CBRmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit((births_count / pop) ~ year + Region, data = comparisons)
tidy(CBRmodel)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic    p.value
```

```
##    <chr>                    <dbl>    <dbl>    <dbl>         <dbl>
## 1 (Intercept)            -1011.     318.     -3.18 0.00300
## 2 year                      0.517    0.158    3.26 0.00236
## 3 RegionTashkent city     -11.4      1.89    -6.05 0.000000542
```
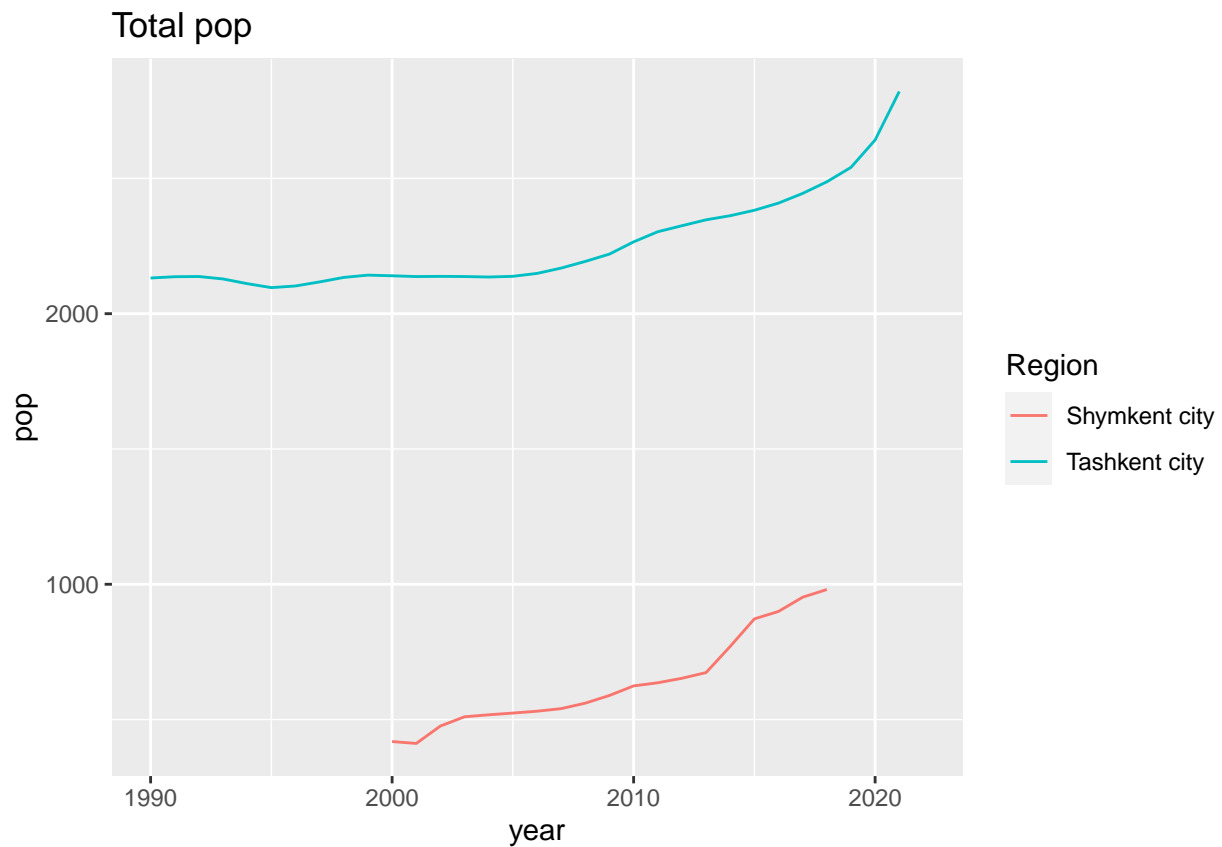
```
glance(CBRmodel)$p.value < 0.01 #regional effects significant
```

```
## value
##  TRUE
```

Population

```
ggplot(comparisons, aes(x = year, y = pop)) +
  geom_line(aes(color = Region)) +
  labs(title = "Total pop")
```
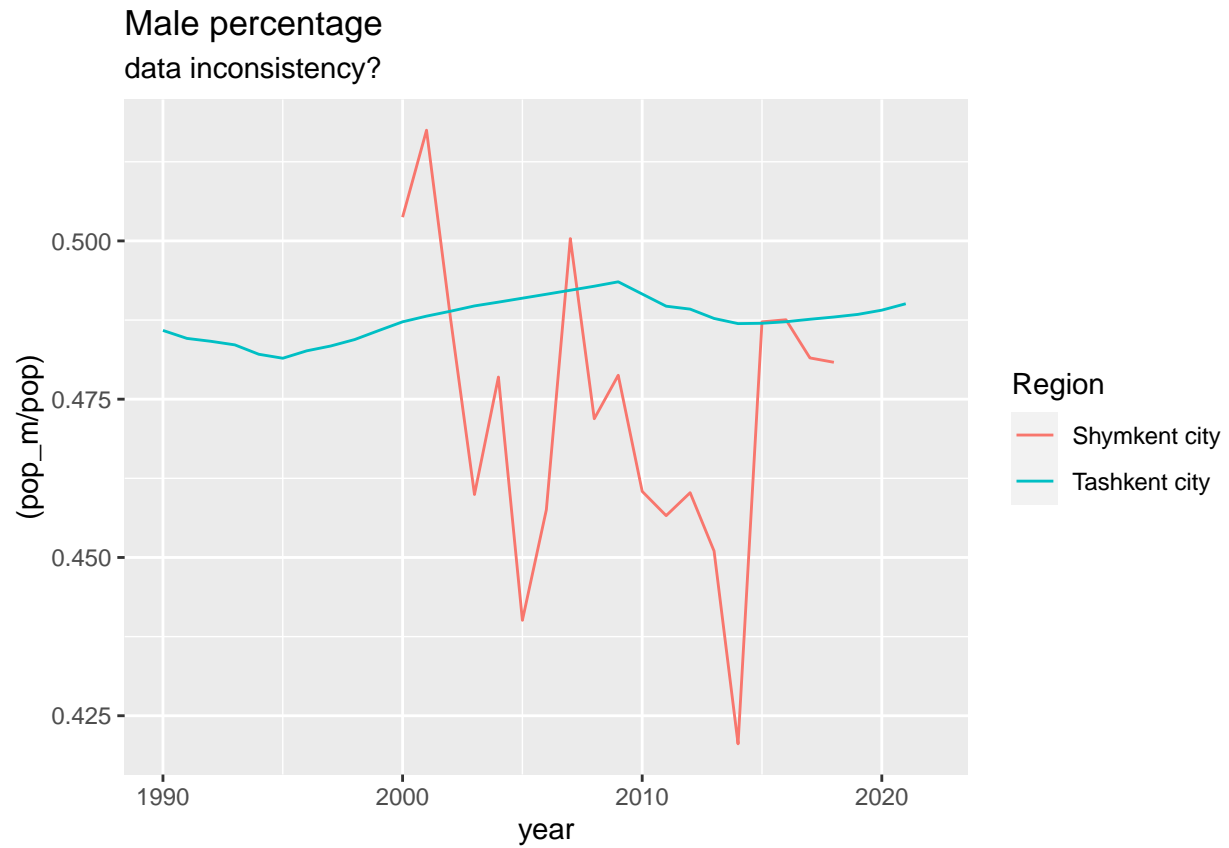
```
## Warning: Removed 1 row containing missing values (`geom_line()`).
```



```
ggplot(comparisons, aes(x = year, y = (pop_m / pop))) +
  geom_line(aes(color = Region)) +
  labs(title = "Male percentage",
       subtitle = "data inconsistency?")
```

```
## Warning: Removed 1 row containing missing values (`geom_line()`).
```

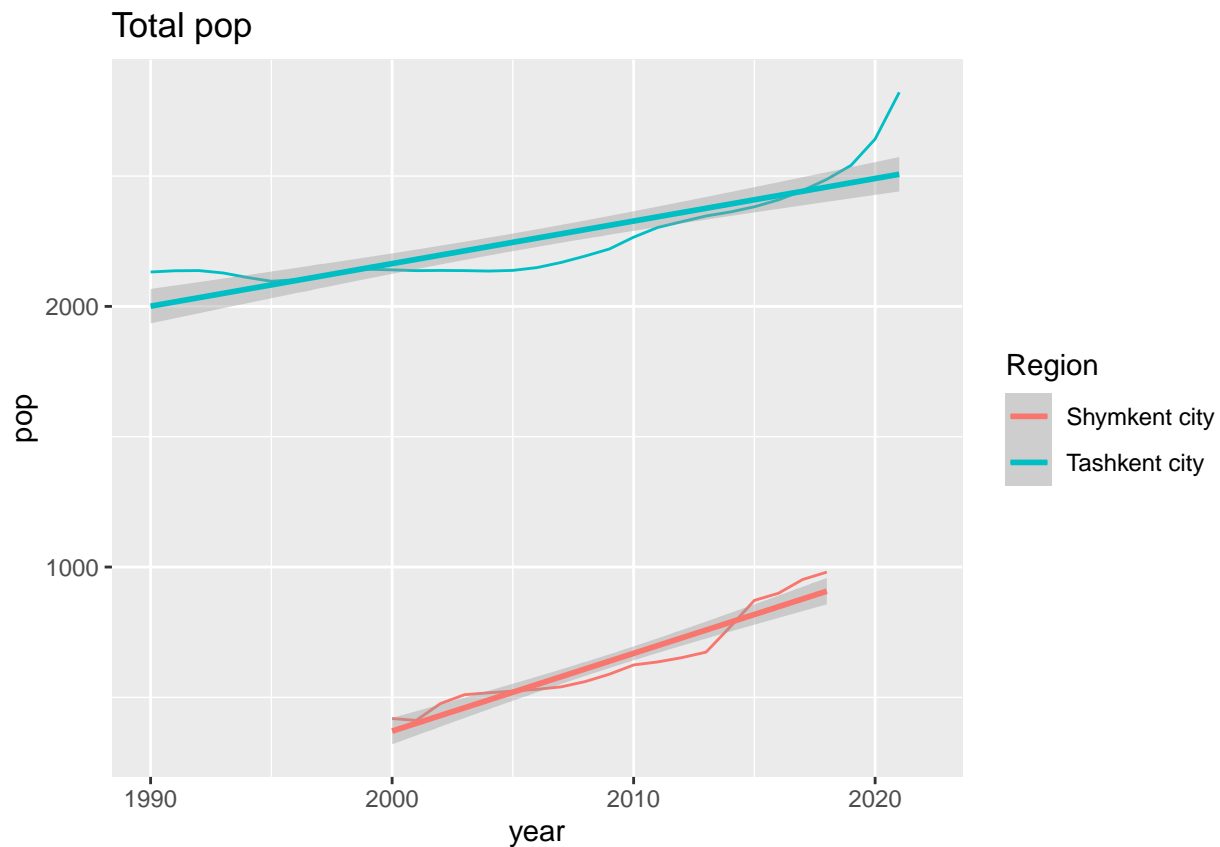## Male percentage
data inconsistency?


```
ggplot(comparisons, aes(x = year, y = pop)) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Total pop")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
## Removed 1 row containing missing values (`geom_line()`).
```
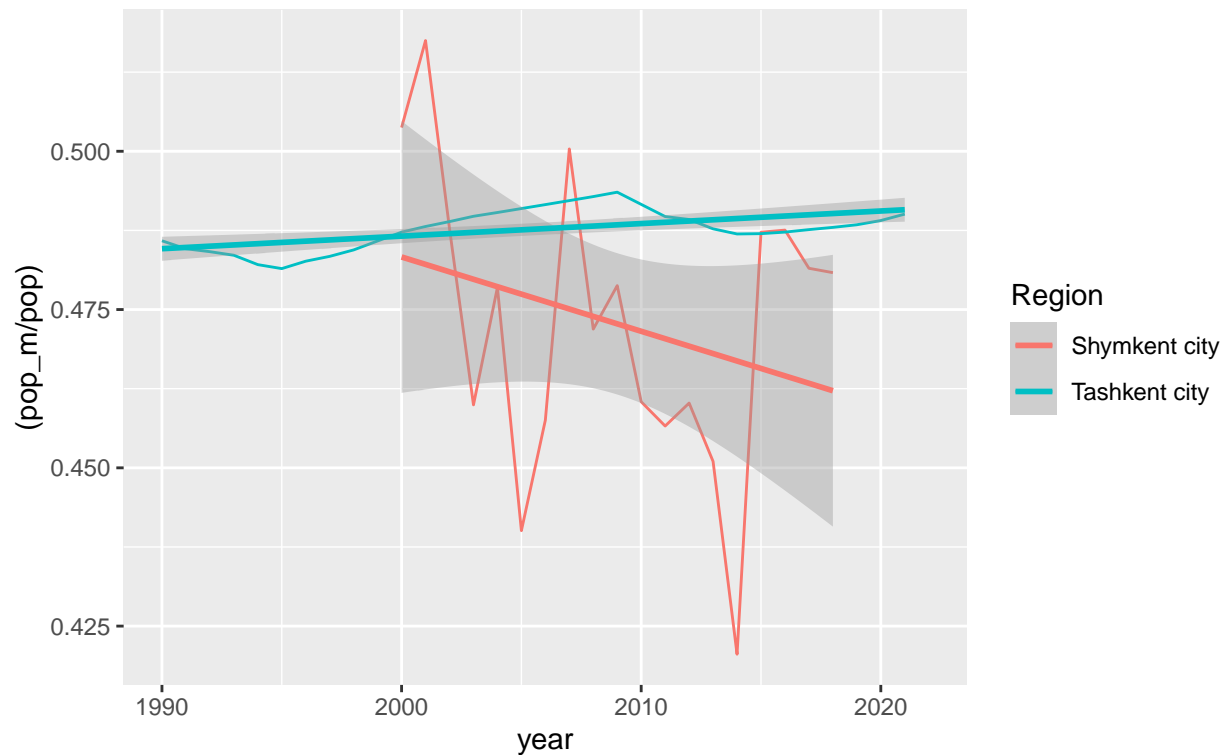
Total pop


```
ggplot(comparisons, aes(x = year, y = (pop_m / pop))) +
  geom_line(aes(color = Region)) +
  geom_smooth(method = "lm", aes(color = Region)) +
  labs(title = "Male percentage",
       subtitle = "data inconsistency?")
```

`geom_smooth()` using formula = 'y ~ x'

Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
Removed 1 row containing missing values (`geom_line()`).

Male percentage
data inconsistency?


```
popmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit(pop ~ year + Region, data = comparisons)
tidy(popmodel)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic  p.value
##   <chr>                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)        -36846.     3189.     -11.6 1.81e-15
## 2 year                  18.7      1.59      11.8 9.78e-16
## 3 RegionTashkent city 1680.      27.0       62.3 1.42e-47
```

```
glance(popmodel)$p.value < 0.01 #regional effects extremely significant
```

```
## value
##  TRUE
```

```
malepercmodel <- linear_reg() %>%
  set_engine("lm") %>%
  fit((pop_m / pop) ~ year + Region, data = comparisons)
tidy(malepercmodel)
```

```
## # A tibble: 3 x 5
##   term              estimate std.error statistic p.value
```

```
##    <chr>                     <dbl>      <dbl>      <dbl>    <dbl>
## 1 (Intercept)            0.550       0.508        1.08  0.284
## 2 year                  -0.0000387  0.000253    -0.153 0.879
## 3 RegionTashkent city    0.0148      0.00430      3.44  0.00120
```

```
glance(malepercmodel)$p.value < 0.01 #regional effects significant
```

```
## value
##  TRUE
```